Protocol A5481027

A MULTICENTER, RANDOMIZED, DOUBLE-BLIND PHASE 3 STUDY OF PALBOCICLIB (ORAL CDK 4/6 INHIBITOR) PLUS LETROZOLE VERSUS PLACEBO PLUS LETROZOLE FOR THE TREATMENT OF PREVIOUSLY UNTREATED ASIAN POSTMENOPAUSAL WOMEN WITH ER(+), HER2(-) ADVANCED BREAST CANCER

Statistical Analysis Plan
(SAP)

Version: v 2.0

Author: PPD, Statistics-Shanghai, China

Date: 14-Oct-2020

Version	Effective Date	Change Type (New, Revise, Admin)	Summary of Revisions
2.0	14-Oct-2020	Revise	Updated final analysis in Section 3. Removed 3-tier approach in Section 6.2. Update the definition of Treatment Emergent Adverse Event (TEAE) in Section 6.2.1. Updated ECG reading used for analysis in Section 6.2.4. Removed Snellen best corrected visual acuity in Section 6.2.5. Added Baseline Variables in Section 6.4.3. Updated the method of analysis of QTc Data in Section 8.1.7. Removed Section 8.1.8.1 Snellen Best Corrected Visual Acuity and Refraction. Added the adjusted log-rank test and repeated confidence interval in Section 8.2.1 Primary Efficacy Analysis. Removed the sensitivity analysis by influence of disease assessment scheduling, influence of deviations in tumor lesion assessment and estimation of treatment effect in Asian Region in Section 8.2.2. Updated the summary of Treatment Administration and Compliance in section 8.2.5. Removed the unnecessary analysis in Section 8.2.6.1.

			Updated QTc analyses in Section 8.2.6.3. Snellen Visual Acuity data analysis will be removed in Section 8.2.7. Removed FACT-B individual items in Section 8.2.10.6.(the Section 8.2.10.6 of the 1st version SAP) and unnecessary analysis in the Section 8.2.10.6 (corresponding to Section 8.2.10.7 in 1st version SAP). Revised the definition of deterioration in Section 8.2.10.8 (corresponding to Section 8.2.10.9 in 1st version SAP). Added consideration on Covid-19 impact in Section 9. Updated dose modification in Section 11.6.2 (corresponding to Section 10.6.2 in 1st version SAP). Added details of adjusted log-rank test in Section 11.9.
1.0	15-Sep-2014	New	New required form.

TABLE OF CONTENTS

LIST OF TABLES

1. AMENDMENTS FROM PREVIOUS VERSION(S)

This is the first version of the Statistical Analysis Plan.

2. INTRODUCTION

Note: in this document any text taken directly from the protocol is ***italicised***.

This document describes the planned statistical analyses for Protocol A5481027 Amendment 1 dated 21 July 2014. This analysis plan is meant to supplement the study protocol. Any deviations from this analysis plan will be described in the Clinical Study Report.

In the first-line setting, letrozole is among the preferred anti-hormonal therapies for postmenopausal women with ER(+)/HER2(-) advanced breast cancer (ABC). It is approved and commercially available globally with a well-known and manageable safety profile. However, median progression-free survival (PFS) in this patient population remains less than 1 year and median overall survival (OS) is approximately 3 years. Furthermore, aromatase inhibitor failure has been linked to increased proliferative index and cell cycle dysregulation, providing a strong rationale for combining letrozole with Palbociclib. Final data from Phase 1/2 Study A5481003 suggests that the combination of Palbociclib inhibition of CDK 4/6 (blocking DNA synthesis by prohibiting progression of the cell cycle from G1 to S phase) with the antiproliferative effects of letrozole provides greater antitumor activity and prolongs PFS (ie, median 20.2 months vs 10.2 months) when compared to single agent letrozole. Additionally, the study showed that the combination is generally well tolerated with uncomplicated neutropenia as the most frequent adverse event.

Based on the encouraging results from this Phase 2 study in ABC, a global randomized Phase 3 study (A5481008) of palbociclib in combination with letrozole in postmenopausal patients with ER(+), HER2(-) ABC is currently ongoing.

This randomized Phase 3 study (A5481027) provides the opportunity to confirm the clinical benefit of the combination of palbociclib with letrozole observed in the randomized Phase 2 study in Asian patients. This study is designed to demonstrate that the combination of palbociclib with letrozole provides superior clinical benefit compared to letrozole in combination with placebo in Asian postmenopausal women with ER(+)/HER2(-) locoregionally recurrent or metastatic advanced breast cancer who have not received any prior systemic anti-cancer therapies for advanced stage disease.

2.1. Study Design

This is a multicenter, randomized (1:1), double blind, placebo-controlled, parallel-group Phase 3 trial comparing the efficacy and safety of palbociclib in combination with letrozole versus placebo plus letrozole in postmenopausal Asian women with ER(+)/HER2(-) ABC. Eligible patients will have histologically or cytologically proven diagnosis of adenocarcinoma of the breast with evidence of locoregionally recurrent or metastatic disease and will be candidates to receive letrozole as first-line treatment for their advanced disease. In order to avoid inclusion of patients who are refractory or resistant to non-steroidal aromatase inhibitors, patients who received anastrozole or letrozole as a component of their (neo)adjuvant regimen may only enter the study if their disease did not

progress while on or within 12 months from completion of their anastrozole/letrozole-containing (neo)adjuvant therapy. Patients will not have received any prior systemic anti-cancer therapy for their advanced disease and will not be candidates for curative therapies. Patients must have measurable disease as per RECIST v1.1 or bone disease as their only site of disease. Tumor tissue availability is required for patient participation.

At least approximately 330 patients will be randomized 1:1 between the investigational arm (ARM A: at least approximately 165 patients treated with palbociclib plus letrozole) and the comparator arm (Arm B: at least approximately 165 patients treated with placebo plus letrozole). Among these approximately 330 patients, at least 264 patients will be from China and the rest of the patients will be from other Asia countries.

Patients will be stratified at randomization by region (China vs Other), by site of disease (visceral vs non-visceral), by disease-free interval since completion of prior (neo)adjuvant therapy (de novo metastatic; ≤12 months; >12 months) and by the nature of prior (neo)adjuvant anticancer treatment received (prior hormonal therapy; no prior hormonal therapy).

<u>Patients randomized to Arm A (investigational arm) will receive</u>:

- *Palbociclib, 125 mg, orally once daily on Day 1 to Day 21 of every 28-day cycle followed by 7 days off treatment;*

in combination with

- *Letrozole, 2.5 mg, orally once daily (continuously).*

<u>Patients randomized to Arm B (comparator arm) will receive:</u>

- *Placebo orally once daily on Day 1 to Day 21 of every 28-day cycle followed by 7 days off treatment;*

in combination with

- *Letrozole, 2.5 mg, orally once daily (continuously).*

Patients will continue to receive their assigned treatment until objective disease progression, symptomatic deterioration, unacceptable toxicity, death, or withdrawal of consent, whichever occurs first. However, patients may continue treatment as assigned at randomization beyond the time of RECIST-defined disease progression 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.

The importance of timely and complete disease assessments in this study cannot be overstated. Disease assessments will be performed every 12 weeks (±7 days) from the date of randomization. Patients with bone lesions identified at baseline will also have repeat bone scans performed every 24 weeks (±7 days) from the date of randomization. Each assessment will be performed as scheduled according to the calendar regardless of any dosing delay to

prevent the introduction of bias into the assessment of efficacy. Failure to perform any of the required disease assessments will result in the inability to determine disease status for that time point. Tumor assessments will be performed until radiographically and/or clinically (ie, for photographed or palpable lesions) documented Progressive Disease (PD) as per RECIST v.1.1, 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. A series of incomplete disease assessments will result in censoring of the primary endpoint of PFS back to the time of the last full assessment that did not show disease progression. Off schedule or incomplete disease assessments have the potential to weaken the conclusion of this clinical trial and must be avoided wherever possible.

Patients who discontinue study treatment for reasons other than radiographically and/or clinically (ie, for photographed or palpable lesions) documented PD as per RECIST v.1.1 will continue to have tumor assessment performed during the follow-up visits every 12 weeks (±7 days) and bone scans (as applicable) every 24 weeks (±7 days) until 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.

Patients discontinuing the active treatment phase will enter a follow-up period during which survival and new anti-cancer therapy information will be collected every 6 months from the last dose of investigational product. The follow-up period will conclude at the time of the final OS analysis. Crossover will not be allowed in the trial.

Efficacy analyses will be performed using the local radiologist's/investigator's tumor assessments as the primary data source. However, a blinded independent third-party core imaging laboratory will complete a retrospective review of radiographic images and clinical information collected on-study to verify the protocol-defined endpoints of disease response and progression determinations as assessed by the investigator.

Patients will undergo study-related safety and efficacy assessments as outlined in the relevant schedule of activities located in the Appendix section.

The study also includes molecular profiling components aimed at characterizing alterations in genes, proteins, and RNAs relevant to the cell cycle, drug targets, and tumor sensitivity and/or resistance in tumor tissues, and assessing the relationship between germline polymorphism in CDK6 gene and palbociclib treatment related neutropenia. In addition, ophthalmic procedures are added to assess a potential risk of palbociclib-associated crystalline lens changes. Therefore, the study will also include ocular safety assessments.

2.2. Study Objectives

Primary Objective:

- *To compare the combination of Palbociclib plus letrozole with placebo plus letrozole in terms of progression-free survival (PFS) in Asian postmenopausal women with ER(+)/HER2(-)Advanced Breast Cancer (ABC) who have not received any prior systemic anti-cancer therapy for advanced disease.*

Secondary Objectives:

- *To compare Objective Response (OR), Duration of Response (DR), Disease Control (DC), and Overall Survival (OS) between the treatment arms;*
- *To evaluate the safety and tolerability between the treatment arms;*
- *To determine trough Palbociclib plasma concentration in this patient population and explore the correlations between exposure and response and/or safety findings;*
- *To compare health-related quality of life between the treatment arms;*
- *To characterize alterations in genes, proteins, and RNAs relevant to the cell cycle, drug targets, and tumor sensitivity and/or resistance in tumor tissues;*
- *To explore the relationship between germline polymorphism in CDK6 gene and palbociclib treatment related neutropenia.*

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

The study is designed to have one interim analysis and the final analysis based on the primary PFS endpoint. The purposes of the interim analysis are to allow early stopping of the study for efficacy, and to assess safety of the combination regimen. The analysis will be performed after approximately 139 patients have documented progressive disease or die (approximately 65% of the total events expected).

A formal efficacy boundary for rejecting the null hypothesis is constructed by using the spending function methodology of the power family design with Δ=-0.5. To protect the integrity of the study and to preserve the type-1 error rate, a fraction of alpha for efficacy will be spent at the interim analysis and accounted for in the overall type I error rate. For logistical and administrative reasons the actual number of events at the interim and final analyses might differ slightly from those that have been pre-specified above. In that case appropriate adjustments will be made to the efficacy boundaries based on the power family spending functions. The overall significance level for the efficacy analysis of PFS will be preserved at 0.025 (1-sided test).

- If the value of the test-statistic at the interim analysis exceeds the efficacy boundary ($z \geq 3.0415$, $p \leq 0.0012$) the trial may be stopped for efficacy.

- Alternatively, as appropriate, the sample size of the study may be adjusted using the method outlined by Cui et al. The detailed information about the sample size re-estimation will be described in a separate document for confidentiality reasons.

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

- The final analysis of PFS will be performed after approximately 213 PFS events have been observed if the interim analysis is not significant.

The efficacy boundary is listed below. Please note that the boundary depends on the number of events. The actual boundary used for the interim analysis will be re-calculated from the specified spending function based on the actual number of events achieved at the time of interim or final analysis.

Table 1. PFS Efficacy Stopping Boundary for Rejecting Null Hypothesis Expressed as Hazard Ratio, Z Scales and p-values

	Number of events	Hazard Ratio	Z-score	P-value (1-sided)
Interim Analysis	139	0.5967	3.0415	0.0012
Final Analysis	213	0.7644	1.9625	0.0249

An interim analysis of efficacy is also planned for the secondary endpoint OS. The analysis will be performed at the same time of the final analysis of PFS. Even if the improvement in PFS is significant at its interim and the study is stopped due to the overwhelming results, the interim OS analysis will still be performed at the approximately planned PFS final analysis time. The nominal significance levels for the interim analysis of OS will be determined by using the Lan-DeMets procedure with an O'Brien-Fleming stopping rule. The overall significance level for the efficacy analysis of OS will be preserved at 0.025 (one-sided test).

Table 2. OS Efficacy Stopping Boundary for Rejecting Null Hypothesis Expressed as Hazard Ratio, Z Scales and p-values

	Number of deaths	Hazard Ratio	Z-score	P-value (1-sided)
Interim Analysis	95	0.4946	3.4305	0.0003
Final Analysis	247	0.7791	1.9617	0.0249

In order to control the overall type I error rate, OS will be hierarchically tested for significance at the time of final PFS analysis, provided the primary endpoint PFS is statistically significant at the interim or final PFS analysis. If OS does not yield a significant result at this analysis, OS will be tested at the final OS analysis. If PFS is not significant at the final PFS analyses, OS will not be statistically evaluated.

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 recommendations made by the E-DMC to alter the conduct of the study will be forwarded to the Sponsor for final decision. The Sponsor will forward such decisions, which may include summaries of aggregate analyses of endpoint and safety data which are not endpoints, to regulatory authorities, as appropriate. 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.

At the initiation of the trial, the trial site will be instructed on the method for breaking the blind. The method will be by contacting the interactive response technology (IRT) provider. Blinding codes should only be broken in emergency situations for reason of 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, but 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.

The final analysis for the clinical study report will be done after the maturity of the primary endpoint PFS, and after all patients' data have been submitted and cleaned, unless other arrangements are agreed upon by the Sponsor. The analysis will be performed at the time of official data base release.

The planned interim analysis for PFS was conducted in 2017 with 139 PFS events observed as of 18MAY2017. The DMC meeting was conducted on 02SEP2017 and DMC recommended the final analysis of PFS to be performed after achieving 255 PFS events. The recommendation was accepted.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The primary purpose of this study is to compare the combination of palbociclib with letrozole versus placebo plus letrozole in prolonging PFS in Asian post-menopausal women with ER (+), HER2 (-) ABC who have not received any prior systemic anti-cancer treatment for advanced disease. All primary and secondary endpoints based on radiological (and photographical when applicable) assessments of tumor burden (ie, PFS, OR, DR, and DC) will be derived using the local radiologist's/investigator's assessment. Tumor assessments will also be performed by a blinded independent third-party core imaging laboratory and the data will be used for secondary supportive analyses. The study is designed to test the null hypothesis that the true PFS distributions for both letrozole plus palbociclib and letrozole plus placebo arms are the same with a median PFS 9 months versus the alternative hypothesis that the true PFS distribution has a median that is longer than 9 months for the letrozole plus palbociclib arm.

4.2. Statistical Decision Rules

The sample size for this study is determined based on the assumptions that the median PFS for ABC patients receiving placebo plus letrozole in the first-line treatment setting is 9 months and it is desired to detect a risk reduction of 36% (hazard ratio 0.64), equivalent to an improvement in median PFS to 14 months in subjects receiving palbociclib plus letrozole treatment. A total of approximately 213 events are required in the 2 arms of the study based on a 1:1 randomization to have 90% power to detect a difference assuming a true hazard ratio of 0.64 in favor of the palbociclib plus letrozole arm using a one-sided, log-rank test at a significance level of 0.025. Assuming a 10% drop-out rate on either treatment arm, an accrual accomplished over a 19-month period and follow-up for about 12 months after the last patient is enrolled, a total sample size of approximately 330 patients (approximately 165 in the palbociclib plus letrozole arm and approximately 165 in the placebo plus letrozole arm) is required.

This study will be considered a positive trial if the 1-sided, stratified log-rank test for PFS based on stratification factors specified in Section 6.4.2 is significant at the level determined at the time of the interim or the final analysis using the power family design stopping rule in favor of palbociclib plus letrozole combination.

The sample size described above will also allow the assessment of differences in the secondary endpoint of OS. The OS outcomes of a Phase 3 clinical trial in a similar patient population demonstrated a median OS of 34 months for the arm receiving letrozole. Using this value as an assumption with a hypothesized 30% risk reduction (a hazard ratio of 0.7) or 43% improvement in median OS (from 34 months to 48.6 months) in patients randomized to receive palbociclib plus letrozole and a follow-up period of approximately 61 months, 330 patients will provide approximately 247 events for 80% power to detect such a difference using a 1-sided, log-rank test at a significance level of 0.025.

OS will be hierarchically tested for significance at the time of final PFS analysis as specified in Section 3.

Other secondary and supportive analyses will be tested at an overall significance level of 0.025 (1-sided test). No adjustments are planned for multiple testing/comparisons in those secondary and supportive hypothesis tests.

5. ANALYSIS SETS

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

The intent-to-treat (ITT) population 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. Modified Intent-to-Treat (MITT) Population

The MITT population will include all patients who are randomized, with the Sponsor-designated central laboratories confirmed ER (+) status, 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. This will be the secondary population for evaluating all efficacy endpoints as well as patient characteristics.

5.3. 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.

5.4. Other Analysis Sets

5.4.1. Biomarker Analysis Set

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

5.4.2. Patient Reported Outcome (PRO) Analysis Set

A subset of ITT patients, who have both baseline and at least one follow-up PRO assessments.

5.4.3. Ocular Analysis Set

A subset of AT patients, who have both baseline and at least one follow-up values for at least one ocular assessment. Patients with ophthalmic conditions (eg, anophthalmus, phthisis, aphakia, pseudophakia) that would prevent grading of the lens in both eyes will not be considered evaluable for this ophthalmic assessment as they do not undergo these ophthalmic procedures.

5.4.4. Pharmacokinetic Analysis Set

A subset of AT patients, who are treated with palbociclib and have at least one measured plasma concentration.

5.5. 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 for all safety analyses.

- If patients were treated but not randomized, then by definition they will be excluded from the efficacy analyses since randomized treatment is missing, but will be reported under the treatment they actually received for all safety analyses.

5.6. Protocol Deviations

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

5.6.1. Deviations Assessed Prior to Randomization

Deviations prior to randomization are typically not allowed. Major deviations that do occur will be tabulated.

5.6.2. Deviations Assessed Post-Randomization

Major deviation is defined as having been treated according to the other treatment arm. Patients not treated with one of the protocol treatments are excluded from safety, biomarker, and PK analyses. Otherwise patients are not excluded from analyses due to post-randomization deviations.

6. ENDPOINTS AND COVARIATES

6.1. Efficacy Endpoint(s)

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 tumor progression as per RECIST v.1.1 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. PFS (in months) will be calculated as (first event date − randomization date +1)/30.4.

  Tumor assessments will be performed every 12 weeks (±7 days) and bone scans (as applicable) every 24 weeks (±7 days) from randomization until radiographically and/or clinically (for photographed or palpable lesions) documented PD as per RECIST v.1.1, 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). 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. Patients who discontinue study treatment for reasons other than radiographically and/or clinically (for photographed or palpable lesions) documented disease progression as per RECIST definitions will continue to have tumor assessment performed during the follow-up visits every 12 weeks (±7 days) and bone scans (as applicable) every 24 weeks (±7 days) until documented 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. Every effort should be made to perform a last tumor assessment before starting a new anticancer therapy. Additional unscheduled tumor assessments may be performed as clinically indicated at any time.

Patients last known to be 1) alive and 2) progression-free, are censored at the date of the last objective disease assessment that verified lack of disease progression (see Appendix 11.4 for determining the date in details). In addition,

- Patients with inadequate baseline disease assessment are censored at the randomization date.
- Patients with no on-study disease assessments are censored at the randomization date unless death occurred within acceptable interval (in which case the death is an event).
- If a new anti-cancer treatment is started prior to progression and death, then censorship is at the date of the last objective disease assessment that verified lack of disease 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 (>2 consecutive assessments) since the last tumor assessment will be censored at the time of last objective assessment documenting no progression.

6.1.2. Secondary Endpoints

- **Overall Survival (OS)** is defined as the time from the date of randomization to the date of death due to any cause. OS (in months) is calculated as (date of death − randomization date +1)/30.4. 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.

Following the End of Treatment visit, survival status will be collected in all patients (telephone contact is acceptable) every 6 months (±7 days) from the last dose of study treatment. Information on subsequent anticancer therapy will also be collected.

- **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; Appendix 11.3). **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 SD requires the criteria to be met at least 12 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.

  Tumor response will be determined from tumor assessment data (where data meet the criteria for CR or PR as described in Appendix 11.3).

- **Disease Control (DC)** 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; Appendix 1). **Disease Control Rate (DCR)** 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 achieving reaching a CR or PR and a best response of SD ≥24 weeks will be counted as non-responders in the assessment of DCR.

  Tumor response will be determined from tumor assessment data (where data meet the criteria for CR or PR and best response of SD as described in Appendix 11.3).

- **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.

Patients last known to be 1) alive and 2) progression-free, are censored at the date of the last objective disease assessment that verified lack of disease progression. In addition,

- If a new anti-cancer treatment is started prior to progression and death, then censorship is at the date of the last objective disease assessment that verified lack of disease 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 (>2 consecutive assessments) since the last tumor assessment will be censored at the time of last objective assessment documenting no progression.

6.2. Safety Endpoints

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.

Adverse events (AEs), hematology, blood chemistry 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, 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 death or 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 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.

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.

6.2.1. Treatment Emergent Adverse Event

An adverse event is considered treatment emergent (TEAE) relative to a given treatment if the event occurs during the on-treatment period.

On-treatment is defined as the time from the first dose of study treatment through 28 days after last dose or start day of new anti-cancer drug therapy, whichever occurs first [minimum (28 days + last dose of study treatment, start day of new anti-cancer drug therapy – 1 day)]. Adverse events occurring on the same day as the first dose of study treatment will be considered to have occurred during the on-treatment period. All other assessments which occur on the same day as the first dose of study treatment will be considered baseline assessments.

6.2.2. 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.

6.2.3. 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 11.1.

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.

6.2.4. Electrocardiogram (ECG)

ECG measurements will include PR interval, QT interval, RR interval, and QRS complex, as well as HR. Triplicate ECGs will be performed for all patients to determine the mean QTc interval. 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 for details) to determine the mean QTc interval.

QT measurements corrected for heart rate (QTc) using Bazett's (QTcB) and Fridericia's (QTcF) method will be used for the data analysis and interpretation.

As triplicate readings are collected, the average of the readings collected at each assessment time will be calculated for each ECG parameter.

ECG readings from central laboratories review will be used for analysis.

6.2.5. Ocular Safety Assessment

Ocular assessments include intraocular pressure measurements, slit-lamp biomicroscopy, funduscopy, and lens grading. The ocular assessments will be performed examination at screening, and on study treatment after 3 months (Cycle 4 Day 1), 6 months (Cycle 7 Day 1), 12 months (Cycle 13 Day 1), every 12 months (Day 1 of Cycles 25, 37 etc.) thereafter, and at the End of Treatment visit (see Protocol A5481027 Section 7.2.3 for details).

6.2.6. Other Safety Assessment

A full physical examination including an examination of all major body systems, 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 Cycle 1 and Cycle 2.

Symptom directed physical examinations, blood pressure and pulse rate will be performed at subsequent visits. Performance Status: The Eastern Cooperative Oncology Group (ECOG) performance status scale will be used.

6.3. Other Endpoints

6.3.1. Pharmacokinetic Endpoints

- Trough plasma concentration of Palbociclib.

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 product on that day will be considered protocol compliant.

One 3 mL sample of venous blood will be collected in appropriately labeled K2 EDTA collection tubes for assessment of Palbociclib (PD-0332991) levels at the protocol-specified times. Samples will be analyzed using validated analytical methods in compliance with Pfizer standard operating procedures.

Blood samples will be collected from all participating patients for PK assessments of Palbociclib on Day 14 of Cycles 1 and Cycle 2 before administration of investigational product on that day (24 hours after the dose on previous day). In the event a pre-dose sample cannot be/is not collected on Day 14 of Cycle 1 or Cycle 2 as scheduled, every effort should be made to collect a makeup pre-dose sample between Day 15 and Day 21 of the same cycle or between Day 14 and Day 21 of any subsequent cycles beyond Cycle 2 following the same rules described above.

Additional blood samples may be requested from patients experiencing unexpected or serious adverse events, or adverse events that lead to discontinuation.

CCI [redacted]

6.3.2. Biomarkers Endpoints

- Tumor tissue biomarkers, including genes, proteins, and RNA expression (eg, ER, Rb, Ki67).

Retrospective confirmatory testing of tumor tissue samples for ER status will be performed in a central laboratory designated by the sponsor using an FDA-cleared test. Results from this testing will be used for sensitivity analyses and will not be made available to the sites. In addition, tumor tissue biomarkers, including DNA, RNA and protein analytes, will be analyzed to investigate possible associations with resistance/sensitivity to treatment with study drugs. 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 not limited to Ki67, Rb, and ER protein expression.

Genomic and metabonomic variation may help to explain some of the variability in response seen with some drugs among different individuals. This is referred to as pharmacogenomics. Comparing the DNA, RNA, protein, and metabolite variation patterns of patients who respond well and those who respond poorly to treatment, or those who have adverse event such as neutropenia may help to better define the most appropriate group of patients in which to target a given treatment. Collecting samples for pharmacogenomic analyses and retaining them in the Pfizer BioBank makes it possible to seek explanations for differences in, for example, exposure, efficacy, tolerability, or safety not anticipated prior to the beginning of the study.

6.3.3. Patient Reported Outcome Endpoints

Patient reported outcomes of health-related quality of life and health status will be assessed using the Functional Assessment of Cancer Therapy-Breast (FACT-B) and EuroQol-5D (EQ-5D) instruments.

Patients will complete each instrument pre-dose on Day 1 of Cycle 1 through 3, then on Day 1 of every other subsequent cycles starting with Cycle 5 (eg, cycles 5, 7, 9, etc), and at the end of treatment.

6.3.3.1. Functional Assessment of Cancer Therapy-Breast (FACT-B) [Version 4]

The Functional Assessment of Cancer Therapy (FACT) is a modular approach to assess patient health-related quality of life using a ‘core’ set of questions (FACT-G) as well as a cancer site-specific module (see Protocol A5481027 Appendix 6 for details).

The FACT-G is a 27-item compilation of general questions divided into 4 domains: Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being.

The FACT-B consists of the FACT-G (27-item) and a breast-specific module: a 10-item instrument designed to assess patient concerns relating to breast cancer.

For all questions, patients are asked to respond to a five-level scale where 0=not at all, 1=a little bit, 2=somewhat, 3=quite a bit, and 4=very much.

6.3.3.2. EuroQol Health Utilities Index EQ-5D

The EuroQol EQ-5D is a brief self-administered health status instrument consisting of two parts. In the first part patients are asked to describe their health state on 5 dimensions (mobility, self-care, usual activities, pain or discomfort, and anxiety or depression) with each dimension having 3 levels of function (1=no problem, 2=some problem, and 3=extreme problem). The scores on the 5 dimensions are summarized to create a single summary score. Because the questions may be answered differently in different countries/regions due to different local customs and social perspectives, published weights from the EuroQol group are used in determining the country appropriate summary scores (EQ-5D User's Guide). The summary score is called the summary index or the health utility value (Shaw et al. 2005).[10] This study will use the UK summary score which ranges from -0.594 to 1 with lower scores corresponding to higher levels of dysfunction (see Appendix 11.8 for details).

The second part of EuroQoL EQ-5D is a visual analogue scale (VAS) in which the patients rate their overall health status using values from 0 (worst imaginable) to 100 (best imaginable).

6.4. Covariates

6.4.1. 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.

6.4.2. Stratification Factors

As described in Section 2.1, randomization in this study will be stratified by the following factors:

- Region (China vs. Other);
- Site of disease (visceral[1] vs. non-visceral[2]);
- Disease free interval since completion of prior (neo)adjuvant therapy[3] (de novo metastatic; ≤12 months; >12 months);
- The nature of prior (neo)adjuvant anticancer treatment received (prior hormonal therapy[4] vs. no prior hormonal therapy[5])

[1] "Visceral" refers to lung and/or liver involvement.

[2] "Non-visceral" refers to absence lung and/or liver involvement.

[3] "Disease free interval" is defined from the stop date of previous systemic therapy in the CRF page to the date of recurrence/metastatic disease from the primary diagnosis CRF page. If the disease is newly diagnosed or the patient didn't use any prior (neo) adjuvant therapy, then he/she should be classified as 'de novo metastatic'.

[4] The hormonal therapy includes tamoxifen and aromatase inhibitor (anastrozole, exemestane and letrozole).

[5] If the patient didn't use any prior (neo) adjuvant therapy, then he/she should be classified as 'no prior hormonal therapy'.

However, only the below factor will be used in the stratified analyses for primary PFS and OS analysis:

- Site of disease (visceral[1] vs. non-visceral[2]).

The effect of other stratification factors will be explored as secondary/supportive analyses.

6.4.3. Baseline Variables

The date of first dose (start date) of study treatment is the earliest date of non-zero dosing of the study drug. The date of last dose of study treatment is the latest date of non-zero dosing of the study drug.

No windowing will be applied when defining baseline, except as noted in Section 11.3. Adequate Baseline Tumor Assessment. Any deviations from the protocol specified window will be documented as protocol deviations.

For efficacy analyses and baseline characteristics associated with tumor assessments the last assessment prior to or on randomization will serve as the baseline assessment.

For safety (including Eastern Cooperative Oncology Group (ECOG) performance status) the last assessment performed on or prior to date of the first dose of study treatment (or prior to randomization for patients randomized but not dosed) will serve as the baseline assessment. If there are no observations meeting these criteria, then baseline is considered missing.

For PRO the last measurement prior to or on the first dose of study treatment will be used as the baseline measurement. If there are no observations meeting these criteria, then baseline is considered missing.

Triplicate ECGs are collected; therefore the baseline for each ECG measurement is the average of the pre-dose measurements on Cycle 1 Day 1 or, if not available, the average of the screening assessment. Unscheduled assessments will not be included in the calculation of the average. Most of the ECG parameters will not be derived as they are provided by sites on the CRF. QTcF (Fridericia's correction) and QTcB (Bazett's correction) are the only parameters that will be derived based on RR and QT. The average of the replicate measurements will be determined after the derivation of the individual parameters at each timepoint.

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 is missing for any date used in a calculation, January 1 will be used to replace the missing date.

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

- For the start date, if the day of the month is missing, the 1st day of the month will be used to replace the missing date. If both day and month are missing, January 1 of the non-missing year will be used to replace the missing date. If the first dose date is later than this imputed date, then impute the start date again to the first dose date.

- For the stop date, if the day of the month is missing, the last day of the month will be used to replace the missing date. If both day and month are missing, December 31 of the non-missing year will be used to replace the missing date.

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.

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 a patient who qualifies for an objective status of CR based on target disease will be classified as PR. 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 within acceptable gap.

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.

7.4. Missing QTc Data

For QTc analysis, no values will be imputed for missing data except for averaging of triplicate measurements. If one or two of the triplicate measurements for an ECG parameter are missed, the average of the remaining two measurements or the single measurement can be used in the analyses. If the triplicate is not good because of an artifact, then the triplicate repeated within about ±15 minutes can be used at that nominal time.

7.5. Missing Patient Reported Outcome Data

For the FACT-B and EQ-5D an ambiguous answer to a question will be assigned the worst score of the answers circled. For the individual domains of FACT-B, a prorated subscale score will be calculated as long as more than 50% of the items are answered. For the total scores FACT-B, FACT-G, TOI, a prorated score will be calculated if more than 80% of the overall items are answered (eg, for FACT-G at least 22 of the 27 items) and all the component subscales have valid scores. Prorated scores may be obtained using the formula below:

$$\text{Prorated subscale score} = [\text{Sum of item scores}] \times [\text{N of items in subscale}] \div [\text{N of items answered}]$$

For the EQ-5D, since each dimension has a single item, responses to all 5 items are needed to calculate an index-based summary score.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Statistical Methods

8.1.1. Hierarchical Group Sequential Testing for PFS and OS

To protect the family-wise error rate (FWER) at level of 0.025, the hierarchical group sequential testing with separate error spending functions at level 0.025 for PFS and OS hypotheses is proposed in this study. To preserve the FWER at 0.025, it is only necessary that the secondary hypothesis for OS is tested whenever the primary null hypothesis for PFS is rejected. That is, if the primary null hypothesis has been rejected (positive PFS) at an interim analysis, the secondary hypothesis can be tested on partial data at the interim analyses as well and again at the final analysis, if it was not significant before.

- 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: Interim analysis (driven by PFS events),

  T_2: Planned final analysis of PFS* when targeted number of PFS events is expected to be observed,

  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 $t_o(T_1)$, $t_o(T_2)$, $t_o(T_3)$ represent information fractions for OS at time points T_1, T_2^*, T_3, respectively.

- $u_p(t)$ and $u_o(t)$ are the efficacy stopping boundaries for PFS and OS, respectively, at information fraction t.

Note: *OS won't be tested at the interim analysis of PFS. In the 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, the overall type-I error rate can be controlled in a strong sense.

1. Interim Analysis of PFS at T_1, Test PFS at $\alpha_p(t_p(T_1))$.

   If $u_p(t_p(T_1))$ is crossed, H_{0p} is rejected.

   No interim Analysis of OS at T_1, go to 2 for interim analysis of OS at T_2 only.

   If $u_p(t_p(T_1))$ is not crossed, go to 2.

2. Final Analysis of PFS at T_2, Test PFS at $\alpha_p(t_p(T_2))$ adjusting for $\alpha_p(t_p(T_1))$.

   If $u_p(t_p(T_2))$ is crossed, H_{0p} is rejected.

   Interim Analysis of OS at T_2, Test OS at $\alpha_o(t_o(T_2))$.

   If $u_o(t_o(T_2))$ is crossed, H_{0o} is rejected.

   If $u_o(t_o(T_2))$ is not crossed, go to 3.

   If $u_p(t_p(T_2))$ is not crossed, PFS not met, study failed and no testing for OS.

3. Final Analysis of OS at T_3, Test OS at $\alpha_o(t_o(T_3))$ adjusting for $\alpha_o(t_o(T_2))$.

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

   If $u_o(t_o(T_3))$ is not crossed, OS not met.

8.1.2. 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 Site of disease (one of the randomization stratification factors as listed in Section 6.4.2) 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 will be estimated using Cox proportional hazards regression.

Cox proportional hazard models will also be used to explore the potential influences of the randomization stratification factors (as listed in Section 6.4.2) 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 will be applied to the overall survival analysis.

8.1.3. Analyses for Binary Data

The rates of binary endpoints for the two treatments will be tested with a 1-sided significance level of 0.025 using a Cochran Mantel Haenszel (CMH)/exact test stratified by Site of disease (one of the randomization stratification factors as listed in Section 6.4.2) and an unstratified Pearson's Chi-Squared/exact test. The odds ratio and its 95% confidence interval will be calculated. 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, while the point estimate of the difference of the rates between treatment arm will be provided along with corresponding approximate 2-sided 95% confidence intervals based on normal distribution.

8.1.4. Analyses of Continuous Data

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

8.1.5. Analyses for Categorical Data

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

8.1.6. Evaluation of the Discordance Rate between Investigator and Independent Core Imaging Laboratory on Assessing PFS data

Potential evaluation bias between the investigator and independent core imaging laboratory assessments with respect to either the progression status of the patient or the timing at which progression occurs will be evaluated using two measures, the early discrepancy rate and late discrepancy rate. The agreement between the investigator and independent core imaging laboratory within a treatment arm is represented in a tabular form below:

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

Note: In practice an investigator PD occurring later than an independent core imaging laboratory PD (a2) would be observed rarely.
a1: number of agreements on timing and occurrence of PD.
a2: number of times investigator declares PD later than independent core imaging laboratory.
a3: number of times investigator declares PD earlier than independent core imaging laboratory.

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 independent core imaging laboratory 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 independent core imaging laboratory 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 can be calculated for each treatment arm and the differential discordance around each measure can be defined 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.

8.1.7. Analyses of QTc Data

QT/QTc values will be summarized by maximum on-treatment values using the following categories:

- ≤450 msec
- >450 msec but ≤480 msec;
- >480 msec but ≤500 msec;
- >500 msec.

The change from baseline will be summarized by shift table of maximum on-treatment values versus baseline values.

Individual QT and QTc values ≥501 msec from each ECG within a triplicate will be flagged in data listings.

In addition, the maximum QTc value for each patient can be categorized and summarized in the following cut-offs. All post dose QTc interval data should be used in determining the maximum for a patient, including all scheduled and unscheduled ECG's.

Absolute QTc interval prolongation
QTc <450 msec
450 msec ≤ QTc ≤480 msec
481 msec ≤ QTc ≤500 msec
QTc ≥501 msec

The maximum increase from baseline QTc value for each patient by treatment will be categorized and summarized as well. For reporting the maximum increase QTc value the following categories: <30 msec, 30-59 msec and ≥60 msec will be used.

8.1.8. Analyses for Ocular Assessment Data

8.1.8.1. 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 by treatment groups.

8.1.8.2. Slit-lamp Biomicroscopy, Funduscopy (Ophthalmoscopy), and Lens Grading

Slit-lamp biomicroscopy results will be graded according to Intraocular Inflammation Grading Scale for Biomicroscopy criteria.

	Grade				
	0	1	2	3	4
Grading of aqueous flare[a]	Completely Absent	Barely Detectable	Moderate (iris and lens details clear)	Marked (iris and lens details hazy)	Intense (formed fibrin in aqueous)
Grading of cells in the aqueous[a,b]	No cells	1 to 5 cells	6 to 10 cells	11 to 20 cells	>20 cells

[a] Evaluation of Anterior Chamber Inflammation:
1. Examination of the anterior chamber for cells must be performed before either dilation or applanation tonometry.
2. The light intensity of the slit lamp is turned to the maximum tolerated by the patient.
3. High magnification and 1 x 2 mm slit are used.
4. The ray of light as directed at an angle of approximately 45o to the plane of the iris.

[b] Modified from Hogan et al. 1959.[76]

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.

During the study, any new findings or deterioration from baseline findings will be reported as an adverse event.

For lens grading, the Wisconsin AREDS 2008 Clinical Lens Opacity Grading Procedure will be used.

8.1.9. Analyses for Patient Reported Outcomes Data

The FACT-B produces five subscale scores: physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), functional well-being (FWB), and a breast cancer subscale (BCS). These subscale scores are used to derive three assessment outcomes: FACT-B total score, FACT-G score, and Trial Outcome Index (TOI), which are calculated as follows:

FACT-B total score = PWB+SWB+EWB+FWB+BCS

FACT-G total score = PWB+SWB+EWB+FWB

TOI score = PWB+FWB+BCS

The EQ-5D index is derived by combining one level from each of the 5 dimensions and converting it to a single summary index or health utility value (Shaw, et al, 2005).[10]

8.2. Statistical Analyses

All efficacy analyses will be based on intent-to-treat (ITT) population. Some efficacy analyses will also be performed on the MITT and AT populations. All analyses will be repeated for China subgroup. All analyses will be performed by using SAS® Version 9.1.3 or higher.

The primary and secondary analyses of endpoints dependent on disease assessments (PFS, OR, DR, and DC) will be based on investigator assessments of disease response and progression. However, a blinded independent third-party core imaging laboratory will complete a retrospective review of radiographic images and clinical information collected on-study to verify the protocol defined endpoints of disease response and progression determinations as assessed by the investigator. Analyses based on the independent third-party core imaging laboratory assessments are considered as secondary and supportive. Evaluation of the Discordance Rate between Investigator and Independent Core Imaging Laboratory on Assessing PFS data will be provided. Summary of concordance between Investigator and Independent Core Imaging Laboratory Assessments of tumor response will be provided as well.

All primary, secondary, and supportive analyses will be tested at a significance level of 0.025 (1-sided test).

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 will be estimated. A log-rank test (1-sided, α=0.025) stratified by Site of disease will be used to compare PFS between the two treatment arms. Repeated confidence interval will also be provided, to account for alpha spending, when appropriate.

Due to the sample size re-estimation based on the interim analysis, the log-rank test stratified by Site of disease or unstratified will be adjusted based on a generalization of the CHW statistic (Cui, Hung, and Wang 1999)[15] to time-to-event analysis. (See details in Appendix 11.9).

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 on the day following the date of the last tumor assessment documenting absence of progressive disease and death. In addition, the following sensitivity analyses on the primary PFS endpoint will be performed in determining whether the primary PFS analysis is robust. The stratified log-rank test (1-sided, α=0.025) will be used to evaluate the primary efficacy endpoint, PFS, in the ITT population.

Influence of additional anti-cancer therapy prior to disease progression or death -A sensitivity analysis will be performed by following patients until PD after discontinuation of the study treatment regardless the initiation of additional anti-cancer therapies. PFS data will be censored on the day of the last tumor assessment documenting absence of progressive disease or death for patients who:

- Are removed from the study prior to documentation of objective tumor progression; or
- Remaining on study without PD or death at the time of the analysis.

The same censoring rule in Section 6.1.1 will be applied to PFS except that patient will not be censored due to start of new anti-cancer therapy.

Influence of Censoring for Patients Who Discontinued from Study Treatment due to Adverse Event or Global Deterioration – A sensitivity analysis will be performed to investigate whether the censoring for patients discontinued without PD due to adverse event or global deterioration of health status influenced the outcome of the primary endpoint PFS. In this analysis, the above patients who discontinued without PD due to adverse event or global deterioration of health status and are censored in primary analysis will be counted as events.

Influence of Patients With Major Protocol Deviation – If there're substantial number of patients with major protocol deviation, a sensitivity analysis to investigate whether these patients affect the outcome of the primary PFS analysis by excluding those patients with major protocol deviation in the analysis.

Influence of Bone-only Disease Patients - A sensitivity analysis will be performed to investigate whether addition of bone-only disease patients influenced the outcome of the primary endpoint PFS by excluding those bone-only disease patients from the analysis.

8.2.3. Secondary Analyses

An unstratified log-rank test (1-sided, α=0.025) and Cox regression model will be used on the primary PFS endpoint as supportive analyses. In addition, the potential influences of the stratification factors (as listed in Section 6.4.2) and baseline patient characteristics such as age, ethnic origin, ECOG performance status, geographical region, and selected biomarkers on the primary PFS endpoint will be evaluated as specified in Section 8.1.2.

The following subgroup analyses will be conducted, the median PFS (and other quartiles) and a 2-sided 95% CI will be provided for each subgroup. The unstratified log-rank test and cox regression model will be used for treatment comparison within each subgroup.

- Subgroup defined by randomization stratification factors;
- Subgroup for interested groups (ie, defined by age group, baseline ECOG, etc.)

The stratified (by Site of disease) log-rank test (1-sided, α=0.025) may be used to evaluate the primary efficacy endpoint, PFS, in the AT and MITT Populations.

A Kaplan-Meier plot for PFS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the PFS censoring and event indicators.

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 Site of disease) log-rank test will be used to compare OS between two treatment arms and the hazard ratio and its 95% confidence interval will be estimated. An unstratified log-rank test (1-sided, α=0.025) and Cox regression model will be used on OS as supportive analyses. The same subgroup analyses conducted for PFS will also be repeated for OS. In addition, subgroup analysis for patients without any subsequent anti-cancer therapy will be performed.

Evaluation of survival follow-up since discontinuation from study treatment will be summarized. Median, mean, maximum and minimum and standard deviation will be provided by treatment arm. Number of patients lost to follow-up will be provided by every 6 months since discontinuation from study treatment and treatment arm.

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

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. A CMH/exact test stratified by Site of disease and an unstratified Pearson's Chi-squared/exact test and will be used to compare ORR between two treatment arms.

The number and proportion of patients achieving disease control (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. A stratified (by Site of disease) CMH/exact test and an unstratified Pearson's Chi-Squared/exact test and will be used to compare DCR between two treatment arms.

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

Supportive analyses for the time-to event endpoints such as PFS and DR will be performed in the ITT population based on the independent third-party core imaging laboratory assessment as described as 8.2.1. and 8.2.3. In addition, analyses will be performed for OR and DC in the ITT population based on the independent third-party core imaging laboratory assessment as described as 8.2.3. 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.

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 baseline characteristics will be conducted in the ITT population, while treatment administration/compliance will be conducted in the as-treated population.

- **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, ER 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 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 11.6 for details) (overall; by study treatment).

- **Treatment Delays and Dose Modifications**

  Dose reductions are not allowed for letrozole. Treatment delays and dose modifications of study treatments will be summarized as follows including number and percent (see Appendix 11.6 for details):

  - The number of patients with at least one palbociclib dose reduction and the number of patients with at least one palbociclib or letrozole 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.

- **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.
  - The time from last dose of the study treatment to the 1st follow-up cancer therapy will be summarized with descriptive statistics (ie, mean, median, and range).

8.2.6. Safety Analyses

Listings of AE, SAE, death, lab data, vital signs, ECG, and physical examinations will be provided according to reporting standard. The safety analyses will be conduct on safety analysis set.

8.2.6.1. Adverse Events

All patients treated with at least one dose of study treatment (ie, palbociclib/placebo or letrozole) 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).

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. 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 death or 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 following summaries of treatment emergent adverse events will also be provided by arm:

- Discontinuations Due to Adverse Events including causality: all cause, treatment related.

- Temporary Discontinuations or Dose Reductions Due to Adverse Events including causality.

- 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 (including Clusters of Preferred Terms) and Maximum NCI CTCAE v.4.0 Grade sorted by Descending Order of AE Frequency (All Causality, and Treatment related).

- Treatment-Emergent Adverse Events by Preferred Term (including Clusters of Preferred Terms) and Maximum NCI CTCAE v.4.0 Grade – 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.

Vital signs categorical summary will be reported by treatment arm.

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.

8.2.6.3. QTc Analyses

All ECGs obtained during the study will be evaluated for safety. The triplicate data will be averaged and all summary statistics and data presentations will use the triplicate averaged data. Any data obtained from ECGs repeated for safety reasons after the nominal time-points will not be averaged along with the preceding triplicates.

Individual ECG data will be listed by treatment and time.

For all patients in the safety analysis population, categorical analysis of the QTcF/QTcB data will be conducted and summarized as follows:

1. QT/QTc values will be summarized and tabulated by the following categories according to 8.1.7.
2. The change from baseline will summarized by shift tables of maximum on-treatment values versus baseline values.
3. Individual QT and QTc values ≥501 msec from each ECG within a triplicate will be flagged in data listings.
4. The number of and percentage patients with maximum post-dose QTcF/QTcB (<450, 450-480, 481-500, and ≥501 ms), including all scheduled and unscheduled ECG's.
5. The number and percentage of patients with maximum increase from baseline in QTcF/QTcB (<30, 30- 59, and ≥60 ms), including all scheduled and unscheduled ECG's.
6. PR changes from baseline ≥50% if absolute baseline value was <200 ms, and ≥25% if absolute baseline value was >200 ms.
7. QRS changes from baseline ≥50% if absolute baseline value was <100 ms, and ≥25% if absolute baseline value was >100 ms.

8.2.7. Ocular Assessment Data Analyses

Ocular assessment will be conducted on Ocular analysis set by treatment arm.

Any IOP increase of greater than 10 mmHg above baseline or any IOP that increases above 25 mm Hg will be reported by treatment groups in data listing.

Any new finding or deterioration from baseline findings in Slit-lamp biomicroscopy, Lens Grading, and Funduscopy will be reported by treatment groups in data listing.

8.2.8. Pharmacokinetic and Pharmacodynamic Analyses

Individual and Average Palbociclib trough concentrations will be listed by patient. Summary statistics will be provided for trough concentrations by study cycle and for average trough concentrations by patient. The relationship between trough concentration and potential covariates may be evaluated. All patients treated with Palbociclib and for whom drug plasma concentration results (from at least 1 visit) are available will be included in the analysis.

Exposure/Response Analysis: In addition, the relationship between exposure and efficacy/safety endpoints will be explored, as necessary, based on emerging biomarker, efficacy and safety data. The results of these modeling analyses may be reported separately from the the clinical study report.

8.2.9. Biomarkers Analyses

The biomarker analyses will be conducted on biomarker analysis set.

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

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

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

8.2.10. Patient Reported Outcomes Analyses

The primary analysis set for the PRO endpoints will be the ITT population. Some analyses will be performed on the PRO evaluable population as appropriate.

The PRO endpoints will be FACT-G total score, FACT-G subscales, breast cancer subscale (BCS), FACT-B total score, trial outcome index (TOI), EQ-5D, EQ-VAS, and time to deterioration.

For each endpoint a completion status table will be provided showing the numbers and percentages of patients at each visit and the numbers and percentages of patients at that visit who completed none, at least one, or all of the items for that endpoint.

8.2.10.1. FACT-G

FACT-G is the sum of the scores from the 27 questions from the FACT-G domains. FACT-G will be summarized using means, medians, standard deviations, and 95% confidence intervals at each assessment point, based on the changes from baseline both within group and between groups. Comparisons between groups will be based on a repeated measures analysis using a mixed effects model. The variables in the model will be treatment, time, treatment-by-time, with baseline as covariate. The minimally important difference (MID) for

FACT-G is 5-6 points. In fitting the mixed model, time will be used as a continuous variable and the method of restricted maximum likelihood will be used assuming an unstructured covariance matrix. No adjustments for multiple comparisons will be made. The dropout rates of the two treatment groups will be compared and, if warranted, a pattern mixture analysis will be performed. This will augment the other analyses and may be available after the completion of the study report.

In addition to the analysis described above, a graphical display of FACT-G change from baseline over time for each treatment group will also be provided.

8.2.10.2. FACT-G Subscales

The 4 FACT-G subscales, called domains, are Physical, Social/Family, Emotional, and Functional well-being (PWB, SWB, EWB, FWB, respectively). Analysis of the FACT-G subscales will follow the same methodology as for FACT-G. The MID for the FACT-G subscales is 2-3 points.

8.2.10.3. Breast Cancer Subscale (BCS)

This subscale consists 10 items associated with breast cancer. Analysis of BCS will follow the same methodology as for FACT-G. The MID for BCS is 2-3 points.

8.2.10.4. FACT-B

This is the overall FACT-B questionnaire consisting of the 27 items from FACT-G and the 10 items from BCS making it 37 items altogether. Analysis of FACT-B will follow the same methodology as as for FACT-G. The MID for FACT-B is 7-8 points.

8.2.10.5. Trial Outcome Index (TOI)

The trial outcome index is defined to be the sum of (PWB+FWB+BCS) making it 24 items altogether. Analysis of TOI will follow the same methodology as for FACT-G. The MID for TOI is 5-6 points.

8.2.10.6. EQ-5D Index

Analysis of EQ-5D will follow the same methodology as for FACT-G.

8.2.10.7. EQ-VAS

Analysis of EQ-VAS will follow the same methodology as for FACT-G.

8.2.10.8. Time to Deterioration

In addition to the above analyses, an examination of the time to deterioration (TTD) will be carried out using survival analysis methods including Kaplan Meier plots and log-rank tests to compare the two treatment groups. Deterioration will be defined as a decrease of a pre-specified number of points based on MID. This analysis will be carried out for the variables BCS, FACT-B, FACT-G, and TOI, and the pre-specified decrease are 2, 7, 5, and 5, respectively.

8.3. Summary of Key Efficacy Analyses

Type of Analyses	Endpoint	Analysis Set	Statistical Method	Missing Data
Primary Analysis	PFS	ITT Investigator assessment (See 8.2.1)	• Stratified log-rank test (stratified by site of disease per randomization, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model	Censor patients on the day following the date of the last tumor assessment documenting in the situations described in 6.1.1
Supportive Analysis	PFS	ITT BICR	• Stratified log-rank test (stratified by site of disease per randomization, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model	Same as primary analysis
Sensitivity analysis	PFS	ITT Investigator assessment and BICR (See 8.2.2)	• Stratified log-rank test (stratified by site of disease per randomization, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model	See 8.2.2
Secondary analysis	PFS	ITT Investigator assessment and BICR (See 8.2.3)	• Unstratified log-rank test (1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the unstratified Cox model	Same as the primary analysis
Secondary analysis	PFS	ITT Investigator assessment and BICR	• Stratified multivariate Cox model with disease, baseline, and demographic characteristics as covariates (stratified by site of disease, 2-sided p-value) • HR and 95% CIs from the stratified cox model	Same as the primary analysis

Secondary analysis	PFS	MITT Investigator assessment and BICR (See 8.2.3)	• Stratified log-rank test (stratified by site of disease per randomization, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model	Same as the primary analysis
Secondary analysis	PFS	AT Investigator assessment and BICR (See 8.2.3)	• Stratified log-rank test (stratified by site of disease per randomization, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model	Same as the primary analysis
Secondary Analysis	OS	ITT (See 8.2.3)	• Stratified log-rank test (stratified by site of disease per randomization, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model	Censor patients who did not die or who lack any data beyond randomization
			Wilcoxon test	
			RPSFT method	
			• Unstratified log-rank test (1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the unstratified Cox model	
Secondary analysis	1-, 2- and 3-year Survival	ITT (See 8.2.3)	K-M method (95% CI)	Censor patients who did not die or who lack any data beyond randomization
Secondary analysis	OR	ITT Investigator assessment and BICR (See 8.2.3)	• Stratified CMH/exact test (stratified by site of disease, 1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CIs) • Stratified odds ratio and 95% CIs.	Treat patients without on-study tumor assessment or who die, progress or drop out for any reason, or receive anti-tumor treatment prior to reaching a CR or PR as nonresponders
			• Unstratified Pearson's Chi-Squared/exact test (1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CIs) • Unstratified odds ratio and 95% CI.	

Secondary analysis	DC	ITT Investigator assessment and BICR (See 8.2.3)	• Stratified CMH/exact test (stratified by site of disease, 1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CIs) • Stratified odds ratio and 95% CIs.	Treat patients without on-study tumor assessment or who die, progress or drop out for any reason, or receive anti-tumor treatment prior to reaching a CR or PR, a SD≥24 weeks as nonresponders
			• Unstratified Pearson's Chi-Squared/exact test (1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CIs) • Unstratified odds ratio and 95% CI.	
Secondary analysis	DR	ITT patients with a CR or PR Investigator assessments and BICR (See 8.2.3)	K-M method (median and 95% CI)	Censor patients on the day following the date of the last tumor assessment documenting (6.1.1)

Abbreviations:
ITT: intent-to-treat; AT: as-treated; MITT: modified intent-to-treat; BICR=Blinded Independent Central Review;
DCR: disease control rate; DR: duration of response; ORR: objective response rate; OS: overall survival; PFS: progression-free survival.

9. CONSIDERATION ON COVID-19 IMPACT

During the study, COVID-19 emergency started and impacted study conduct.

As of the date of finalization of the current SAP, no COVID-related deaths have been reported, most of the scheulded visits occurred as per plan, no patients missed consecutive tumor assessment and permanently discontinued from treatment due to COVID-19. Therefore the following analysis will be conducted to assess the impact due to COVID-19.

Protocol deviations due to COVID-19 will be summarized.

A listing of patients impacted by COVID-19 will be provided.

If there are deaths due to COVID-19, sensitivity analysis for PFS and OS may be considered.

10. REFERENCES

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

2. Cox DR. Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, Series B 1972; 34:187-220.

3. DeMets D, Lan KKG. Interim analysis: the Alpha Spending function approach. Statistics in Medicine 1994; 13:1341-1352.

4. 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.

5. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53:457-481.

6. 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.

7. 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.

8. ICH Harmonised Tripartite Guideline. The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs. E 14. May 12, 2005.

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

10. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Medical Care 2005; 43:203-220.

11. The EuroQol Group. EQ-5D: A measure of health-related quality of life developed by the EuroQol Group: User Guide. V 2.0. 2009. The EuroQol Group, Rotterdam.

12. Sprangers MAG, Groenvold M, Arraras JL, et. Al. The European Organization for Research and Treatment of Cancer Breast Cancer-Specific Quality-of-Life Questionnaire Module: First Results From a Three-Country Field Study. J Clin Oncol 1996; 14(10): 2756-2768.

13. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81:515-26.

14. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 Bio-Algorithms and Med-Systems 2013; 9(4):183-9.

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

11. APPENDICES

11.1. 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 subject.

A5481027 Schedule of Activities

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		Cycle 1	Cycle 2	Cycles ≥3		
Study Day	Within 28 days prior to randomization unless specified otherwise	Day 1[b, v]	Day 14	Day 21	Day 21	Day 1[v]		
Time Window		±2d	±2d	±2d	±2d	±2d		±7d
Baseline Documentation								
Informed Consent Process[e]	X							
Medical / Oncological History[f]	X							
Baseline Signs / Symptoms		X[g]						
Banked Biospecimen[h]		X						
Tumor Tissue for Biomarker[i]	X						X[i]	
Physical Examination/Vital signs[j]	X	X[b]				X	X	
Ophthalmic Examination[k]	X					X	X	
ECOG Performance Status	X	X				X	X	
Laboratory Studies								
Hematology[l]	X	X[b]	X (Only for Cycle 1)	X	X	X	X	

A5481027 Schedule of Activities

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		Cycle 1	Cycle 2	Cycles ≥3		
Study Day	Within 28 days prior to randomization unless specified otherwise	Day 1[b, v]	Day 14	Day 21	Day 21	Day 1[v]		
Time Window		±2d	±2d	±2d	±2d	±2d		±7d
Blood Chemistry[l]	X	X[b]	X			X	X	
12-Lead ECG[m]	X	X[b]	X			X	X	
Fasting Glucose[n]		X (Cycle1Pre-dose)		X			X	
Fasting Insulin[n]		X (Cycle1Pre-dose)		X			X	
HgbA1C[n]		X (Cycle1Pre-dose)				C4D1 and every 3 months thereafter	X	
Fasting Lipid Panel (Total Cholesterol, HDL, LDL, triglycerides)[n]		X (Cycle1Pre-dose)					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[o]	X	◄--►[p o] Performed every 12 weeks (± 7 days) from the date of randomization					X	X[o]
Radionuclide Bone Scan, Whole Body[p]	X	◄--►[q o] Performed every 24 weeks (± 7 days) from the date of randomization					X	X[o]
Other Clinical Assessments								
Drug Compliance[r]		◄--►						
Averse Event Reporting[s]	X	X	X			X	X	X
Review Concomitant Medications/Treatments[t]	X	X	X			X	X	X

A5481027 Schedule of Activities								
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		**Cycle 1**	**Cycle 2**	**Cycles ≥3**		
Study Day	**Within 28 days prior to randomization unless specified otherwise**	**Day 1[b, v]**	**Day 14**	**Day 21**	**Day 21**	**Day 1[v]**		
Time Window		**±2d**	**±2d**	**±2d**	**±2d**	**±2d**		**±7d**
EuroQol; EQ-5D[u]		X				X	X	
FACT - Breast Questionnaire[u]		X[u]				X	X	
Survival Follow-up[w]								X
Study Treatment								
Randomization	X							
Letrozole (both treatment arms)		Once Daily ◄--►[x]						
Palbociclib or Placebo		◄--►[y] Once Daily on Day 1 to Day 21of each cycle followed by 7 days off						
Special Laboratory Studies								
Pharmacokinetics[z]			X(Pre-dose)					

a. **Active Treatment Phase:** All assessments should be performed prior to dosing with study medications on the visit day unless otherwise indicated. Acceptable time windows for performing each assessment are described in the column headers.
b. **Cycle 1/Day 1:** Blood chemistry, hematology, 12-lead ECG and physical examination not required if acceptable screening assessment is performed within 7 days prior to randomization.
c. **End of Treatment/Withdrawal:** Obtain these assessments if not completed during the previous 4 weeks on study (or within the previous 8 weeks for disease assessments).
d. **Post Treatment Follow-up:** After discontinuation of study treatment, post-treatment follow-up (including survival status and post-study anticancer therapy evaluation) will be collected every 6 months (±7 days) from the last dose of study treatment. Telephone contact is acceptable.
e. **Informed Consent:** Informed consent may be obtained greater than 28 days from randomization; however, must be obtained prior to any protocol required assessments being performed.
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. **Banked Biospecimen:** A single 4 mL blood sample (Prep D1; K2 EDTA whole blood collection optimized for DNA analysis) will be collected pre-dose at the Cycle 1 Day 1 visit from all patients, unless prohibited by local regulations, to be retained for potential pharmacogenomic analyses related to drug response or adverse drug reactions.

i. **Mandatory Tumor Tissue For Confirmatory Testing and for Biomarker Assessments:** Tumor tissue is required for patient participation. Submission of formalin-fixed paraffin embedded (FFPE) tumor samples (blocks) of adequate size to allow for three 0.6 mm diameter x 5 mm deep cores that will be used to generate a tissue microarray are needed. If FFPE tissue block cannot be provided, then 12 glass slides each containing an unstained 5-micron FFPE tissue section, will be required for patient participation (highly recommend to submit tissue block or 12 unstained slides, but if there would be technical difficulties or other issues refraining from obtaining tissue block or 12 unstained slides, however, it is still mandatory to collect a minimum 7 unstained slides). Archived FFPE specimen from the original diagnostic tumor tissue will be collected and sent to the sponsor-designated central laboratories for assessment of biomarkers associated with sensitivity and/or resistance to Palbociclib (eg, Ki67, ER). Tissue sample from a metastatic or recurrent tumor lesion, if available, will also be collected for restrospective confirmation of ER status by the central laboratory. If a tissue sample from a recurrent tumor or distant metastasis is unavailable, then a de novo fresh biopsy is recommended when, in the investigator's judgment, such biopsy is feasible and can be safely performed. Original diagnostic tumor tissue will be used for confirmation of ER status in the event that a recurrent/metastatic tissue sample is not available and a fresh biopsy of the recurrent/metastatic lesion is not feasible. An optional fresh tumor biopsy will be collected at the end of treatment visit,only for patients who discontinue treatment due to disease progression. The tumor tissue will be used to determine possible mechanisms of resistance. Tissue samples from all patients will be used for additional biomarker analyses. Detailed information about biomarker sample collection, preparation, storage, labeling, and shipment is indicated in the Study Reference Manual.

j. **Physical Examination/Vital signs:** A full physical examination including an examination of all major body systems, 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 Cycles 1 and 2. Symptom-directed physical examinations, blood pressure and pulse rate will be performed at subsequent visits.

k. **Ophthalmic Examinations:** All enrolled patients will undergo an ophthalmic examination at screening, and on study treatment after 3 months (Cycle 4 Day 1), 6 months (Cycle 7 Day 1), 12 months (Cycle 13 Day 1), every 12 months (Day 1 of Cycles 25, 37 etc.) thereafter, and at the End of Treatment visit. Additional ophthalmic examinations may be performed during the study as clinically indicated (including for patients randomized prior to Amendment 3 approval). The ophthalmic examinations will include: best corrected distant visual acuity (Snellen), refractive error associated with best corrected distant visual acuity, intraocular pressure (IOP – one reading), slit lamp biomicroscopy of the anterior segment including cell count and flare grading, crystalline lens grading using the Wisconsin Age-Related Eye Disease Study (AREDS) 2008 Clinical Lens Opacity Grading procedure, and fundoscopy. All ophthalmic examinations will be performed by an ophthalmologist. Refer to Section 7.2.3. Ocular Safety Assessments for further details on these procedures.

l. **Hematology, and Blood Chemistry Panel:** Hematology includes hemoglobin, WBC, absolute neutrophils, platelet count. Blood chemistry includes AST/ALT, alkaline phosphatase, sodium, potassium, magnesium, total calcium, total bilirubin, BUN (or urea), serum creatinine, and albumin. Additional hematology/chemistries panels may be performed as clinically indicated.

m. **12-Lead ECG:** At each time-point, three consecutive 12-lead ECGs will be performed approximately 2 minutes apart to determine the mean QTc interval. If the mean QTc interval is prolonged (>500 msec), then the ECGs should be re-read by a cardiologist or other qualified person at the site for confirmation. Additional triplicate ECGs may be performed as clinically indicated.

n. **Fasting Glucose, Fasting Insulin, HgbA1c and Fasting Lipid Panel (Total Cholesterol, HDL, LDL, triglycerides):** On Day 1, all the baseline assessments should be performed prior to dosing(pre-dose). Fasting glucose and fasting insulin will be assessed on Cycle 1 Day 21 and at the End of Treatment visit; HgbA1C will be assessed on Cycle 4 Day1, every 3 months (Day 1 of Cycles 7, 10 etc.) thereafter, and at the End of Treatment visit; fasting lipid panel will be assessed at the End of Treatment visit.

o. **Disease Assessments:** Please refer to the tumor assessment requirement flowchart for details and timing of procedures. A blinded independent third-party core imaging laboratory will complete a retrospective review of radiographic images and clinical information collected on-study to verify the protocol-defined endpoints of disease response and progression determinations as assessed by the investigator.
p. **CT/MRI Scans of Chest, Abdomen, Pelvis, any clinically indicated sites of disease, and of bone lesions; clinical evaluation of superficial disease:** Please refer to the tumor assessment requirement flowchart for details and timing of procedures.
q. **Radionuclide Bone Scan, Whole Body:** Please refer to the tumor assessment requirement flowchart for details and timing of procedures.
r. **Drug Compliance:** Palbociclib, placebo and letrozole bottles including any unused capsules/tablets will be returned to the clinic for drug accountability. Drug accountability will be performed on Day 1 of every cycle prior to dispensing drug supply for the next cycle.
s. **Adverse Events:** For SAEs, the active reporting period begins from the time that the patient provides informed consent through and including 28 calendar days after the last administration of the investigational product. AEs (serious and non serious) should be recorded on the CRF from the time the patient has taken at least one dose of study treatment through last patient visit.
t. **Concomitant Medications/Treatments:** Concomitant medications and treatments will be recorded from 28 days prior to the start of study treatment and up to 28 days after the last dose of study treatment.
u. **EQ-5D, FACT-B Assessments:** Patients will complete questionnaires prior to any study or medical procedure on Day 1 of Cycles 1, 2 and 3 and then Day 1 of every other cycle thereafter starting with Cycle 5 (ie, Cycle 5, 7, 9, etc), and at the end of treatment visit. 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.
v. **Cycle X, Day 1:** In the event that the start of a new cycle is delayed due to treatment related toxicity, procedures required on Day 1 of the given cycle will be performed when Palbociclib/placebo is resumed. New cycle Day 1 procedures (ie, physical examination, ECOG performance status, ECG, Quality of Life questionnaires, blood chemistry, hematology) that were performed prior to knowing the need to delay the start of the cycle do not need to be repeated (1) if not required to determine whether study drug may be resumed and (2) if performed within 7 days prior to study drug resumption.
w. **Survival Follow-Up:** After discontinuation of study treatment, post-study survival status (including post-study anticancer therapies) will be collected every 6 months (±7 days) from the last dose of study treatment. Telephone contact is acceptable.
x. **Letrozole (both treatment arms):** To be taken orally once daily continuously.
y. **Palbociclib or Placebo:** To be taken orally once daily from Day 1 to Day 21 (21 days) of every 28-day cycle followed by 7 days off treatment.
z. **Pharmacokinetics:** For all patients, plasma PK samples for Palbociclib determination will be collected prior to dosing (pre-dose) on Day 14 of Cycle 1 and Cycle 2.
Additional blood samples may be requested from patients experiencing unexpected or serious adverse events, or adverse events that lead to discontinuation.

11.2. Tumor Assessment Requirements Flowchart

	Screening[a]	Treatment Period[b]	End of Treatment Visit[c]
CT [d] or MRI of chest, abdomen, and pelvis (CAP)	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 correlative bone imaging	Required [g,h]	Required for sites of disease identified at screening or if clinically indicated [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 during the treatment period, every 12 weeks (±7 days) and bone scans (as applicable) every 24 weeks (±7 days) from randomization until radiographically and/or clinically (ie,for photographed or palpable lesions) documented PD as per RECIST v.1.1, 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 12 weeks (± 7 days) and bone scans (as applicable) every 24 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) 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. Brain scans performed before the signing of informed consent as routine procedures (but within 6 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) the same technique/modality can be used to follow identified lesions throughout the trial for a given patient (3) appropriate documentation indicating that these radiographic tumor assessments were performed as standard of care is available in the patient's source notes. 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) the same technique/modality can be used to follow identified lesions throughout the trial for a given patient (3) 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, every 12 weeks ±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. If bone lesions were identified at baseline bone scans will be repeated during the active treatment phase every 24 week (±7 days) from the date of randomization and at the time of confirmation of CR. If no bone lesions were identified at baseline, bone scans will only be repeated during the active treatment phase when 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 confirmation of CR. 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 case report form (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.

11.3. 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 treatment 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

11.4. 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.

11.5. 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.
Baseline triplicate ECGs	Cycle 1 Day 1 dose or from date closest to, but prior to, start of study treatment if C1D1 is not available.
Tumor assessment baseline values	From date closest but prior to first dose.
Design subsets	Chinese subset
Measurable disease	Defined by RECIST
Adequate baseline tumor assessment	Within 35 (28 + 7) days prior to first dose. 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.

11.6. Study Treatment Modification and Compliance

11.6.1. Dose Modification

No dose adjustment for letrozole is permitted but dosing interruptions are allowed. Treatment interruption for letrozole-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. Since letrozole is administered daily continuously, a treatment delay is not applied to letrozole.

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.

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

The following types of summaries are proposed for administration of palbociclib and letrozole.

When palbociclib is administered in combination with letrozole (orally once daily continuously), on a orally once a day for 21 days of every 28-day cycle followed by 7 days off treatment (cyclical dosing), the following summaries can be presented:

- RDI for palbociclib: Overall;
- RDI for letrozole: Overall.

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

- Regular Cycle or Complete Cycle: There is another cycle after the current one.
- Last Cycle: The treatment is permanently discontinued after the current cycle.

- Intended Total Dose Per Cycle is the same (2.5 mg [once daily continuous] x 28 days for letrozole and 125 mg [once a day for 21 days followed by 7 days break in a 28 days treatment cycle] x 21 days for palbociclib) for all regular cycles. The daily dose is fixed at the start of treatment rather than start of a cycle.

- Intended Dosed Days Per Cycle for palbociclib.

  - 21 days for a regular cycle, or

  - Minimum of (21 days, actual treatment duration) for the last cycle.

- Intended Treatment Duration is the same for the entire dosing period, except for the last cycle which is the actual duration of treatment up to 4 weeks. (eg, for a 3/1 dosing schedule, all cycles have an intended duration of 4 weeks).

  - 28 days for a regular cycle, or

  - Minimum of (28 days, actual treatment duration) for the last cycle.

- Actual Total Dose Per Cycle is the total dose a patient actually took in a cycle.

- Actual Treatment Duration is the treatment duration for a cycle.

  - Start date of next cycle – Start date of current cycle for a regular cycle,

  - Last dose date – Start date of the cycle+1 for the last cycle.

Table 1

Treatment / Summary Type	Calculation of RD/RDI	Example
Cyclical PD991 / 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) <u>Note:</u> 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 *7+ 5*7 + 4*7) = 76.92 mg/day Intended Overall Dose Intensity=656.25 /7=93.75mg/day RDI = (76.93 / 76.92) *100% = 82.1%

Table 2

Treatment / Summary Type	Calculation of RD/RDI	Example
Cyclical PD991 / By Cycle	$RDI = \frac{Actual\ Dose\ Intensity}{Intended\ Dose\ Intensity} * 100\%$ *Actual Dose Intensity (per day)* = (Actual Total Dose per cycle) / (Actual Treatment Duration for the cycle) For a regular cycle *Actual Treatment Duration for the Cycle* = Start date of next cycle – Start date of current cycle. For the last cycle *Actual Treatment Duration for the Cycle* = Last dose date – Start date of the cycle+1. *Intended Dose Intensity (per day)* = (Intended Total Dose per cycle) / (Intended Treatment Duration in cycle) For a regular cycle Intended Total Dose per cycle is always 125 * 21 = 2625 mg; Intended Treatment Duration is always 28 days For the last cycle Intended Total Dose in last cycle = 125 * [Min(21, actual treatment duration)] Intended Treatment Duration in last cycle = Min[(28, actual treatment duration)]	• 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 in cycle 1, 2 = (125*7*3) / (4*7) = 93.75 mg/day Intended Dose Intensity in cycle 3 (last cycle) = (125*7*3) / (3*7) = 125 mg/day Cycle 1: Actual Dose Intensity = (125*7*3) / (4*7) = 93.75 mg/day RDI = (93.75/93.75) * 100% = 100% Cycle 2: Actual Dose Intensity = (125*7*3) / (5*7) = 75 mg/day RDI = (75/93.75) * 100% = 80% Cycle 3 (Last Cycle): Actual Dose Intensity = (125*7*2) / (3*7) = 83.3 mg/day RDI = (83.3/125) * 100% = 66.7%

Table 3

Treatment / Summary Type	**Calculation of RD/RDI**	**Example**
letrozole By Cycle & Overall	$RDI = \frac{Actual\ Dose\ Intensity}{Intended\ Dose\ Intensity} * 100\%$ *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) $RDI = \frac{Actual\ Overall\ Dose\ Intensity}{Intended\ Overall\ Dose\ Intensity} * 100\%$ *Actual Overall Dose Intensity* = (Sum over all cycles of the *"Actual Total Dose Level"*) / (Sum over all cycles of the "*Actual number of weeks in cycle"*) *Intended Overall Dose Intensity = Intended Dose Intensity (per week)*	• letrozole is to be dosed at 2.5 mg daily continuously • Actual letrozole dosing: D1 to D28 on Cycle 1; D1 to D14 and D22 to D28 in Cycle 2 (ie, 7 days interruption) • Actual Dose is 2.5*28 = 70mg in Cycle 1 • Actual Dose is 2.5* 21 = 52.5 mg in Cycle 2 • Actual Total Dose level is 70 mg in Cycle 1 and 52.5 mg in Cycle 2 Cycle 1: Actual Dose Intensity=70/(4*7)=2.5 mg/day Intended Dose Intensity = 70/(4*7) = 2.5 mg/day RDI= (2.5 / 2.5) * 100% =100% Cycle 2: Actual Dose Intensity = 52.5/(4*7) = 1.875 mg/day Intended Dose Intensity = 70/(4*7) = 2.5 mg/day RDI= (1.875 / 2.5) * 100% =75% Overall: Actual Overall Dose Intensity = (70 + 52.5) / [(4 + 4)*7] = 2.1875 mg/day Intended Overall Dose Intensity = (70 + 70) / [(4 + 4)*7] = 2.5 mg/day RDI= (2.1875/2.5)*100% =87.5%

11.7. Functional Assessment of Cancer Therapy-Breast (FACT-B) (Version 4)

Below is a list of statements that other people with your illness have said are important**.**
Please circle or mark one number per line to indicate your response as it applies to the past 7 days.

	PHYSICAL WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GP1	I have a lack of energy	0	1	2	3	4
GP2	I have nausea	0	1	2	3	4
GP3	Because of my physical condition, I have trouble meeting the needs of my family	0	1	2	3	4
GP4	I have pain	0	1	2	3	4
GP5	I am bothered by side effects of treatment	0	1	2	3	4
GP6	I feel ill	0	1	2	3	4
GP7	I am forced to spend time in bed	0	1	2	3	4

	SOCIAL/FAMILY WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GS1	I feel close to my friends	0	1	2	3	4
GS2	I get emotional support from my family	0	1	2	3	4
GS3	I get support from my friends	0	1	2	3	4
GS4	My family has accepted my illness	0	1	2	3	4
GS5	I am satisfied with family communication about my illness	0	1	2	3	4
GS6	I feel close to my partner (or the person who is my main support)	0	1	2	3	4
Q1	*Regardless of your current level of sexual activity, please answer the following question. If you prefer not to answer it, please mark this box* ☐ *and go to the next section.*					
GS7	I am satisfied with my sex life	0	1	2	3	4

Please circle or mark one number per line to indicate your response as it applies to the <u>past 7 days</u>.

	EMOTIONAL WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GE1	I feel sad	0	1	2	3	4
GE2	I am satisfied with how I am coping with my illness	0	1	2	3	4
GE3	I am losing hope in the fight against my illness	0	1	2	3	4
GE4	I feel nervous	0	1	2	3	4
GE5	I worry about dying	0	1	2	3	4
GE6	I worry that my condition will get worse	0	1	2	3	4

	FUNCTIONAL WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GF1	I am able to work (include work at home)	0	1	2	3	4
GF2	My work (include work at home) is fulfilling	0	1	2	3	4
GF3	I am able to enjoy life	0	1	2	3	4
GF4	I have accepted my illness	0	1	2	3	4
GF5	I am sleeping well	0	1	2	3	4
GF6	I am enjoying the things I usually do for fun	0	1	2	3	4
GF7	I am content with the quality of my life right now	0	1	2	3	4

Please circle or mark one number per line to indicate your response as it applies to the <u>past 7 days</u>.

	ADDITIONAL CONCERNS	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
B1	I have been short of breath	0	1	2	3	4
B2	I am self-conscious about the way I dress	0	1	2	3	4
B3	One or both of my arms are swollen or tender	0	1	2	3	4
B4	I feel sexually attractive	0	1	2	3	4
B5	I am bothered by hair loss	0	1	2	3	4
B6	I worry that other members of my family might someday get the same illness I have	0	1	2	3	4
B7	I worry about the effect of stress on my illness	0	1	2	3	4
B8	I am bothered by a change in weight	0	1	2	3	4
B9	I am able to feel like a woman	0	1	2	3	4
P2	I have certain parts of my body where I experience pain	0	1	2	3	4

11.8. EuroQol Health Utilities Index EQ-5D

By placing a tick in one box in each group below, please indicate which statement best describe your own health state today.

Mobility

I have no problems in walking about ❑

I have some problems in walking about ❑

I am confined to bed ❑

Self-Care

I have no problems with self-care ❑

I have some problems washing or dressing myself ❑

I am unable to wash or dress myself ❑

Usual Activities (*e.g. work, study, housework, family or leisure activities)*

I have no problems with performing my usual activities ❑

I have some problems with performing my usual activities ❑

I am unable to perform my usual activities ❑

Pain/Discomfort

I have no pain or discomfort ❑

I have moderate pain or discomfort ❑

I have extreme pain or discomfort ❑

Anxiety/Depression

I am not anxious or depressed ❑

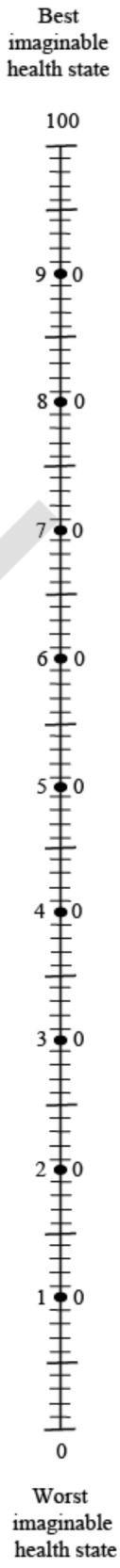
I am moderately anxious or depressed ❑

I am extremely anxious or depressed ❑

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today.

Your own health state today



11.9. Adjusted Log-rank Test

During the planned interim analysis, the sample size re-estimation was conducted based on CHW (Cui, Huang, and Wang 1999)[15] method. Due to the sample size re-estimation, the log-rank test will be adjusted based on a generalization of CHW statistics as specified below.

Assuming HR is the hazard ratio for the treatment arm relative to the control arm . We wish to test the null hypothesis H0: HR>=1, Let $D_{\max}$ be the maximum number of events required to achieve $1-\beta$ power at the alternative H1:HR<1, with a one-sided test of α level significance. Let D1 denote the number of event at the time of the interim analysis. Let Z_1 be the log-rank test statistic at the interim look. Let Z_2 be the log-rank statistic we would use at the final analysis if the maximum number of events remains unchanged and equals $D_{\max}$. In addition, let Z_2^* be the log-rank test statistic at the final analysis if the maximum number of events is adjusted from $D_{\max}$ to $D_{\max}^*$ based on the observed $Z_1=z_1$. In order to preserve the type-1 error of the test for H0, we must use the weighted statistic

$$T_2 = \sqrt{\frac{D_1}{D_{\max}}}Z_1 + \sqrt{\frac{D_{\max}-D_1}{D_{\max}}}\left(\frac{\sqrt{\frac{D_{\max}^*}{D_{\max}}}Z_2^* - \sqrt{\frac{D_1}{D_{\max}}}Z_1}{\sqrt{\frac{D_{\max}^*-D_1}{D_{\max}}}}\right) \quad (1)$$

instead of the unweighted statistic Z_2^* at the final analysis. The statistic T_2 utilizes the independent increment structure of the sequentially computed log-rank test and is a generalization of the CHW statistic (Cui, Hung, and Wang 1999;[15] East 6 User Manual) to time-to-event analysis. The type-1 error is protected by using T_2 at the final analysis. Since T2 follows asymptotic standard normal distribution under null hypothesis. The one-sided P-value can be calculated by P=1-$\Phi(T_2)$.

SAS Programming details:

Stratified analysis

```
PROC LIFETEST DATA=mydata timelist=(12);
TIME EEVALUEM*EVENTN(0);
Strata STRAT2F/Group=RANDTEXT trend;
RUN;
```

Unstratified analysis

```
PROC LIFETEST DATA=cut1 timelist=(12);
TIME EEVALUEM*EVENTN(0);
Strata RANDTEXT/trend;
RUN;
```

P= 1-probnorm(T_2)