Protocol A5481008

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

Statistical Analysis Plan (SAP)

Version: 3.1

Author: PPD

Date: January 30, 2017

TABLE OF CONTENTS

1. AMENDMENTS FROM PREVIOUS VERSION(S)

The SAP is being amended to reflect the changes in Protocol Amendments 1 (October 1, 2013), 2 (January 3, 2014), 3 (March 21, 2014), 4 (September 18, 2014), 5 (December 2 2014), 6 (April 17, 2015), and 7 (October 15, 2015).

1.1 Amendment 1

1.1.1 Sample size adjustment to address potential impact of antacid concomitant medication and fed condition (Protocol Amendment 3).

PD-0332991 was initially administered as a free base capsule formulation under minimal fasting conditions (fasting 1 hour before and two hours after PD-0332991 dosing). Data from clinical pharmacology studies that emerged after the initiation of Study A5481008 showed that administration of free base capsule formulations under fasting conditions resulted in high variability in PD-0332991 exposure due to low PD-0332991 exposure of in a subgroup of PK profiles (referred to as "low-liers"). Data from the Food Effect study in healthy subjects (Study 1021) demonstrated that low-liers were not observed when the PD-0332991 free base capsules were administered with food or in-between meals (moderate-fat meal 1 hour before and 2 hours after dosing). In addition, preliminary results from clinical pharmacology study A5481018 showed that proton pump inhibitors (PPI) administered concomitantly with PD-0332991 free base capsule under overnight fasting conditions substantially decreased the exposure of PD-0332991. The combination of these observations resulted in Protocol Amendment 2, to revise the dosing instructions from drug taken in a fasted state to drug taken with food. The protocol was also revised to prohibit the concomitant use of proton pump inhibitors.

Given the fact that patients who took PD-0332991 concomitantly with PPIs/other antacids or under fasted conditions prior to the implementation of Protocol Amendment 2 may have experienced lower PD-0332991 exposure the treatment effect in the ITT population could have been diluted, potentially resulting in a study with lower power than originally designed (90%). To mitigate this potential risk, an adjustment for the alternative hypothesis was made in Protocol Amendment 3 from HR=0.64 to HR=0.69. It resulted an increase of PFS events by 108 and additional 200 patients, which is approximately the number of patients enrolled prior to the initiation of Amendment 2. The study still maintains 90% power to detect a hazard ratio of 0.69 or better in favor of PD-0332991 plus letrozole combination in the larger ITT population.

In addition, considering that patients who may have been affected by the use of PPI or dosing under fasting conditions would have been enrolled in the early part of the study (prior to the Protocol Amendment 2), the interim analysis is being conducted after 226 events or 65% of total PFS events (instead of 55% in the original plan). As a result of the sample size revision, the final analysis will occur at 347 events).

1.1.2 Modification of efficacy boundary of primary endpoint PFS for interim analysis (Protocol Amendment 5).

A modification for interim analysis was proposed to, and agreed with FDA to increase the stringency of the efficacy stopping boundary in the interim analysis to ensure that the results are not only statistically significant but also clinically meaningful. Specifically, the efficacy stopping boundary is changed from O'Brien-Fleming to the Haybittle-Peto approach. A p-value of 0.000013 will be used as the efficacy boundary for interim analysis. The overall significance level for the efficacy analysis of PFS will be preserved at 0.025 for 1-sided test (section 3).

1.1.3 Gate-keeping procedure for primary PFS hypothesis testing in ITT population and a sub-population in the final analysis.

The treatment effect on PFS for all patients randomized in the study (ITT population) and a sub-group population who were not impacted by PPI, and/or other antacids concomitant medication will be assessed. A gate-keeping procedure is added in this SAP amendment as a prospectively planned analysis to test the treatment effect of PFS of PD-0332991 plus letrozole in this optimal sub-population (section 8.1.1).

1.1.4 Prospective ophthalmic assessments (Protocol Amendment 3).

Based on available limited pre-clinical and clinical data, it was not evident that a clinical risk exists for developing PD-0332991-associated cataracts, however it was viewed to be important to increase the level of clinical information in this area. The protocol was therefore amended to implement prospective ophthalmic assessments and lens grading in all newly enrolled patients at baseline and while on study treatment.

The results of the ophthalmic assessments will provide a better understanding of any potential risk of ocular adverse events. The analysis plan for the ocular assessment data is added in the amendment.

Additionally, the following changes were made in this SAP amendment:

1) A detailed sequence of hierarchical tests of PFS and OS to preserve the family wise error rate at the level of 0.025 is provided in section 8.1.2;
2) Sensitivity analyses in section 8.2 are modified to provide clarity to each analysis.
3) Clarification is added in the PRO data analysis section.

1.2 Amendment 1 Revision

After the SAP amendment, additional data with regard to the effect of acid reducing agents (ARAs) including PPI, H2RA and local antacid on PD-0332991 exposure under fed condition became available. The results showed that co-administration of ARAs with PD-0332991 under fed conditions did not have clinically relevant effect on PD-0332991 exposure. This led to the revision of SAP Amendment 1 and the subgroup definition is modified in section 8.1.1.

1.3 Amendment 2

Protocol Amendment 7 expanded data collection of disease progression date for the subsequent anti-cancer therapy in post treatment follow up phase. The SAP is modified to reflect the protocol changes. Corresponding analysis may be performed if appropriate.

Other editorial modifications include:

- Section 5.4.4 modified definition of biomarker analysis set.
- Section 8.2.4 a kappa test is added to the summary of concordance between investigator and BICR assessments.
- Section 10.6.2 clarified the calculation of Relative Dose Intensity.

Per FDA request, a formal interim analysis on data cutoff of November 24, 2016 is added for OS analysis. To preserve the type-1 error rate, a fraction of alpha (0.0001) will be spent for this analysis (section 3).

2. INTRODUCTION

This document describes the planned statistical analyses for original Protocol A5481008 dated on October 30, 2012. The changes in Protocol Amendment 1 (October 1, 2013), 2 (January 3, 2014), 3 (March 21, 2014), 4 (September 18, 2014), and 5 (December 2 2014) were also incorporated. 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 to PD-0332991. Preliminary data from Phase 1/2 Study A5481003 suggests that the combination of PD-0332991 inhibition of CDK 4/6 (blocking DNA synthesis by prohibiting progression of the cell cycle from G1 to S phase) with the anti-proliferative effects of letrozole provides greater antitumor activity and prolongs PFS (i.e. median 26.1 months vs. 7.5 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.

This randomized Phase 3 study (A5481008) provides the opportunity to confirm the clinical benefit of the combination of PD-0332991 with letrozole observed in Study A5481003. The study is designed to demonstrate that PD-0332991 in combination with letrozole provides superior clinical benefit compared to letrozole in combination with placebo in postmenopausal women with ER(+)/HER2(-) locoregionally recurrent or metastatic breast cancer who have not received any prior systemic anti-cancer therapies for their advanced disease.

2.1. Study Design

This is an international, multicenter, randomized, double-blind, placebo-controlled, parallel-group, Phase 3 clinical trial comparing the efficacy and safety of PD-0332991 in combination with letrozole versus placebo in combination with letrozole in postmenopausal women with ER(+)/HER2 (-) advanced breast cancer. 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. To further ensure that the study will enroll only those women for whom letrozole is the most appropriate treatment, patients with advanced, symptomatic, visceral spread, that are at risk of life threatening complications in the short term will be excluded from this study. 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 therapies for their advanced disease and will not be candidates for curative therapies. Patients must have measurable disease as per RECIST v.1.1 or bone disease as their only site of disease. Tumor tissue availability is required for patient participation.

Approximately 650 eligible patients will be randomized 2:1 to receive either PD-0332911 plus letrozole (Arm A ≈433 patients) or placebo plus letrozole (Arm B ≈217 patients).

Patients will be stratified 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 vs. no prior hormonal therapy).

Patients randomized to Arm A (experimental arm) will receive:

- PD-0332991, 125mg, orally once daily on Day 1 to Day 21 of every 28-day cycle followed by 7 days off treatment;

in combination with

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

Patients randomized to Arm B (control arm) will receive:

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

in combination with

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

Patients will continue to receive 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 progression disease (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. In this case, the patient would continue with routine safety assessments as per the Schedule of Activities for the active treatment period.

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 (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 (e.g., 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 progression. Frequently off schedule or incomplete disease assessments have the potential to weaken the conclusion of this clinical trial.

Patients who discontinue study treatment for reasons other than radiographically and/or clinically (for photographed or palpable lesions) documented PD as per RECIST v.1.1 will continue to have tumor assessments performed during the follow-up visits every 12 weeks (± 7 days) and bone scans (if applicable) every 24 weeks (± 7 days) until RECIST-defined disease progression, initiation of new anticancer therapy or discontinuation of patient from overall study participation (e.g., death, patient's request, lost to follow-up), whichever occurs first.

Efficacy analyses will be performed using the local radiologist’s/investigator’s tumor assessments as 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 as assessed by the investigator.

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.

Patients will undergo study-related safety, efficacy, and PK assessments as outlined in the relevant Schedule of Activities located in the Appendix section.

The study also includes:

- QTc monitoring to evaluate the effect of PD-0332991 on QT interval via triplicate ECGs time-matched with select serial PK draws (subset study in at least 60 patients enrolled at selected sites);

- Ocular safety assessments. Corresponding to protocol amendment 3, ophthalmic procedures are added to assess a potential risk of PD-0332991-associated crystalline lens changes.

- Quantification of trough PD-0332991 plasma concentration.

- A molecular profiling component aimed at assessing the relationship between breast tumor sensitivity and resistance to PD-0332991 and the alteration of cell cycle pathway-related genes and proteins in tumor tissues.

2.2. Study Objectives

Primary Objective:

- To demonstrate that the combination of PD-0332991 with letrozole is superior to placebo plus letrozole in prolonging PFS in postmenopausal women with ER(+)/HER2 (-) advanced breast cancer who have not received any prior systemic anti-cancer therapies for their advanced/metastatic disease.

Secondary Objectives:

- To compare measures of tumor control duration and overall survival between the treatment arms;
- To compare safety and tolerability between the treatment arms;
- To compare health related quality of life between the treatment arms;
- To characterize the effects of PD-0332991 at therapeutic doses in combination with letrozole on QTc interval in this patient population;
- To determine trough PD-0332991 plasma concentration in this patient population and explore the correlations between exposure and response and/or safety findings;
- To characterize alterations in genes, proteins, and RNAs relevant to the cell cycle (e.g., CCND1 amplification, CDKN2A deletion), drug targets (e.g. CDK 4/6), and tumor sensitivity and/or resistance (e.g., Ki67, pRb) in tumor tissues.

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

The study is designed to have one interim analysis and the final analysis at 347 events based on the primary PFS endpoint with the investigator assessment. The Haybittle-Peto efficacy boundary (Haybittle 1971; Peto et al. 1976) for rejecting the null hypothesis will be used at the time of the interim analysis. To protect the integrity of the study and to preserve the type-1 error rate, a fraction of alpha (0.000013) for efficacy will be spent at the interim analysis and accounted for in the overall type I error rate. The overall significance level for the efficacy analysis of PFS will be preserved at 0.025 (1-sided test).

The purposes of the interim analysis are to allow early stopping of the study for futility and efficacy, to assess the safety of the combination regimen, and to potentially adjust the sample size. The analysis will be performed after approximately 226 PFS events (documented progressive disease or death; approximately 65% of the total events expected).

- If the value of the test-statistic exceeds the efficacy boundary ($z \geq 4.2059$, $p \leq 0.000013$) the trial may be stopped for efficacy. It is assumed that a hazard ratio less than 1.0 is in favor of the PD-0332991 plus letrozole arm.
- If the hazard ratio between the PD-0332991 plus letrozole arm and the placebo plus letrozole arm equates 1.0 or greater, the trial will be stopped for futility.
- 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 clinical trial.
- The final analysis of PFS will be performed after approximately 347 PFS events have been observed if the Interim Analysis is not significant and the sample size is not adjusted after the interim analysis.

If the study is determined to be stopped due to efficacy, the interim analysis will become the final and all the subsequent efficacy and safety analyses specified in this SAP will be performed.

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

Analysis	Number of Event (%)	Z Scale	p-value
Interim	226 (65%)	4.2059	0.000013
Final	347 (100%)	1.96	0.025

Interim analyses of efficacy are also planned for the secondary OS endpoint and OS will be tested in a hierarchical approach. The analysis will be performed at the time of the interim or final PFS analyses if the primary analysis for PFS is positive. The nominal significance levels for the interim and final analyses of OS will also 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).

- The first possible time for OS interim analysis will be at the time of the PFS interim analysis. If improvement in PFS is significant at its interim, the OS interim analysis can be performed at this time. However, the number of deaths could be relatively low. The reasonable time for OS interim analysis will be at approximately 131 deaths (at the estimated time for planned PFS final analysis). If OS is not significant at the interim analysis, a final analysis will be performed after 390 deaths have been observed.
- If PFS is not significant at the time of the interim analysis of PFS, a final analysis for PFS will be performed. If PFS is significant at its final analysis, the OS interim analysis will also be performed at this time. If OS is not significant at its interim, the final analysis for OS will be performed when a total of 390 deaths have been recorded.
- If PFS is not significant at its final analysis or did not meet the futility stopping boundary, 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 either a manual or electronic process. Blinding codes should only be broken in emergency situations to preserve the patient safety. Blinding codes may also be broken after a patient discontinues treatment due to disease progression, as determined by the treating investigator using RECIST v.1.1 criteria, only if deemed essential to allow the investigator to select the patient's next treatment regimen and after discussion and agreement with the sponsor. Codes should not be broken in the absence of emergency situations or progressive disease as per RECIST v.1.1 (e.g., 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.

During sNDA reviewing, US FDA requested a formal OS analysis on data cutoff of November 24, 2016. This interim analysis does not have intention to change the course of planned final OS analysis at 390 deaths. To protect the integrity of the study and to preserve the type-1 error rate, a fraction of alpha (0.0001) will be spent for this interim analysis and accounted for in the overall type I error rate.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The primary purpose of this study is to demonstrate that the combination of PD-0332991 with letrozole is superior to placebo plus letrozole in prolonging PFS in 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 (i.e. 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 PD-0332991 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 PD-0332991 arm.

4.2. Sample Size Determination and Statistical Decision Rules

4.2.1. Sample Size for PFS

The sample size for this study is determined based on the assumptions that the median PFS for patients receiving placebo plus letrozole in the first-line treatment setting is 9 months and a risk reduction by 31% (a hazard ratio of 0.69) or an improvement by 44% to median PFS of 13 months in the PD-0332991 plus letrozole treatment is clinically significant. With one interim analysis performing at 65% of total PFS events, a total of 347 events are required in the two arms of the study based on a 2:1 randomization to have 90% power to detect a hazard ratio of 0.69 in favor of PD-0332991 plus letrozole arm using a 1-sided, unstratified log-rank test at a significance level of 0.025. Assuming a 15% drop-out rate on either treatment arm, a non-uniform accrual accomplished over a 15-month period and follow-up for about 10 months after the last patient is enrolled, a total sample size of approximately 650 patients (~433 in PD-0332991 plus letrozole arm and ~217 in 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 randomization stratification factors is significant at the significance level of 0.000013 at the time of the interim or 0.025 at the final analysis in favor of PD-0332991 plus letrozole combination.

4.2.2. Sample Size for OS

The sample size described above will also allow the assessment of differences in the secondary endpoint of OS. No crossover to experimental therapy is permitted in the control group within the current study design. The OS outcome of a reported phase 3 clinical trial with a similar patient population was 34 months for the arm receiving letrozole. Using this value as an assumption with a hypothesized 26% reduction risk (a hazard ratio of 0.74) or 35% improvement in median OS (from 34 months to 46 months) in patients randomized to receive PD-0332991 plus letrozole and follow-up period of approximately 68 months, evaluation of 390 events using a 1-sided, unstratified log-rank test is required for a significance level of 0.025 and power of 80% to detect the difference.

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

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

4.2.3. Sample Size for QTc

The QTc analysis will be based on a non-inferiority hypothesis testing framework. Approximately 40 PD-0332991 plus letrozole treated patients are needed to establish non-inferiority (no unacceptable QTc prolongation from the PD-0332991 plus letrozole combination) between post dose baseline and baseline (ΔQTc) at all 5 QTc sampling time points on Cycle 1 Day 14 with 90% power. The test is based on a 1-sided difference in means t-test for paired ΔQTc with significance level 0.05. The difference in means between ΔQTc under the alternative hypothesis is 10 mSec, assuming a non-inferiority margin of 20 mSec and the standard deviation of the paired differences equal 16 mSec based on Study A5481003). If the upper bounds of one-sided 95% confidence intervals of ΔQTc for all 5 QTc sampling time points are below 20 mSec, the post-baseline dose QTc interval is considered to be "non-inferior" to the baseline; the QTc effect of the PD-0332991 in combination with letrozole is concluded to be not unacceptable.

Since this is a double-blind study with a 2:1 randomization, approximately 60 patients will be needed for QTc evaluation to ensure approximately 40 patients from the PD-0332991 plus letrozole arm.

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. QTc Analysis Set

A subset of AT patients, who are enrolled at selected sites (considered as Group 1) and have their QTc monitored to evaluate the effect of PD-0332991 on QT interval via serial triplicate ECGs time-matched with PK draws, must have at least one pair of time-matched baseline (Day 0) and post-PD-0332991 dose (Day 14 of Cycle 1) ECG measurements.

5.4.2. Ocular Analysis Set

A subset of AT patients, whose ocular safety assessments are collected. 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. The analysis of ocular assessment data will be performed in ocular analysis set.

5.4.3. Pharmacokinetic Analysis Set

A subset of AT patients, who are treated with PD-0332991 and have at least one measured plasma concentration.

5.4.4. Biomarker Analysis Set

A subset of AT patients, who have baseline values for at least one biomarker.

5.4.5. Patient Reported Outcome (PRO) Analysis Set

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

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.

5.6. Protocol Deviations

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

5.6.1. Protocol Deviations Assessed Prior to Randomization

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

5.6.2. Protocol 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 analyses. Otherwise patients are not excluded from analyses due to post-randomization deviations.

6. ENDPOINTS AND COVARIATES

6.1. Efficacy Endpoints

6.1.1. Primary Endpoint

- **Progression Free Survival (PFS)** is defined as the time from the date of randomization to the date of the first documentation of objective 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 (e.g., 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 (e.g., 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 10.4 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 prior to the first planned assessment (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 10.3.). **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 10.3.).

- **Disease Control (DC)/Clinical Benefit Response (CBR)** is defined as the overall complete response (CR), partial response (PR) , or stable disease (SD) ≥24 weeks according to the Response Evaluation Criteria in Solid Tumors (RECIST version 1.1; Appendix 1). **Disease Control Rate (DCR)/ Clinical Benefit Response Rate (CBRR)** is defined as the proportion of patients with CR, PR, or SD ≥24 weeks relative to (1) all randomized patients and 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/CBRR.

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 10.3.).

- **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 (i.e. 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 prior to 28 days after discontinuation of treatment, 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 Data

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

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.

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

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

Tier-2 events: These are events that are not tier-1 but are "common". A MedDRA preferred term is defined as a tier-2 event if there are at least 10% for all a grades in any treatment group. For grade 3/4/5 analysis, the events should be reported in at least 5% patients in any treatment group.

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

6.2.1. Treatment Emergent Adverse Event

An adverse event is considered treatment emergent if:

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

6.2.2. 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 10.1.

Baseline evaluations for laboratory are those collected

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

6.2.4. Electrocardiogram (ECG)

ECG measurements will include PR interval, QT interval, RR interval, and QRS complex. Triplicate ECGs will be performed for all patients to determine the mean QTc interval for eligibility purpose.

Patients found to be eligible will be part of one of the two groups highlighted below depending on which site screened the patient. ECG frequency for each group is described below:

- Patients in Group 1 (approximately 60 patients) will be enrolled at selected sites and will have their QTc monitored to evaluate the effect of PD-0332991 on QT interval via serial triplicate ECGs time-matched with PK draws. ECGs will be obtained on the day preceding treatment initiation (Day 0) at time 0 (time of first ECG also referred to as ECG1) and then 2, 4, 6, and 8 Hrs after ECG1, and on Day 14 of Cycle 1 pre-dose (0 Hrs) and 2, 4, 6, and 8 Hrs following PD-0332991 administration. Timing of ECGs performed on Day 14 MUST be time-matched (clock time ±35 minutes) with ECG assessments performed on Day 0 (e.g, if the ECG1 on Day 0 was performed at 10:00 AM then the 0 hour triplicate ECGs on Day 14 must be performed within the 9:30 AM - 10:30 AM timeframe but as close as possible to 10:00 AM whenever feasible). On Day 14 of Cycle 1, study treatment should be administered immediately after the pre-dose PK draw has been collected. All ECGs should be obtained after a fast of at least 1 hour. ECGs should be performed immediately before PK blood draws at respective time points. Patients who cannot complete ECG measurements on Day 0 and/or PK-matched ECG measurements (both ECG and PK collected) on Day 14 will need to be replaced. Additionally, triplicate ECGs will be obtained for safety monitoring at 0 hour (pre-dose) on Day 1 of Cycle 1, Day 14 of Cycle 2, then on Day 1 of Cycles 4, 7, and 10. ECGs beyond Cycle 10 will be performed as clinically indicated.
- Patients in Group 2 (all other patients) will have triplicate ECGs performed for safety monitoring at 0 hour (pre-dose) on Day 1 of Cycle 1, Day 14 of Cycles 1 and 2, then on Day 1 of Cycles 4, 7, and 10. ECGs beyond Cycle 10 will be performed as clinically indicated. All ECGs should be obtained after a fast of at least 1 hour.

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. All triplicate ECG tracings will be sent electronically to a core ECG laboratory for blinded manual interval measurements. The blinded manual interval measurements from the core ECG laboratory will be used for primary statistical analysis of ECG data in Group 1 patients.

QT measurements corrected by heart rate (QTc) will be used for the data analysis and interpretation. In addition to commonly used techniques including Bazett's (QTcB) and Fridericia's (QTcF) methods, a study-specific correction method will be evaluated (QTcS) for the ECG Group 1. QTcF will be used for the primary analysis.

6.2.5. Ocular Safety Assessment

Corresponding to protocol amendment 3, ocular safety assessments will be applied to all the newly enrolled, lens grading evaluable, patients in the study. Ocular assessments include Snellen best corrected visual acuity, intraocular pressure measurements, slit-lamp biomicroscopy, funduscopy, and lens grading. The ocular assessments will be performed at baseline (screening), every 3 months in the first year, and every 12 months thereafter during study treatments.

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 and Pharmacodynamic Data

All efforts will be made to obtain the pharmacokinetic samples at the scheduled nominal time relative to dosing. However, samples obtained within 10 min or 10% of the nominal time and collected prior to administration of the investigational product on that day (with the exception of the planned post-dose samples that will be collected on Day 14 of Cycle 1) will be considered protocol compliant.

One 3 mL sample of venous blood will be collected in appropriately labeled K2 EDTA collection tube for assessment of PD-0332991 (including its active metabolites, if appropriate) 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 PD-0332991 on Day 14 of Cycle 1 and Cycle 2 before administration of investigational product on that 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.

For patients participating in the dedicated QTc portion of the trial (Group 1), all PK draws at Day 14 of Cycles 1 and 2 will be done immediately after triplicate ECGs have been performed so that the PK samples are collected at the nominal time.

As part of understanding the pharmacokinetics of the PD-0332991, plasma samples may be used for metabolite identification and/or evaluation of the bioanalytical method. These data will be used for internal exploratory purposes and will not be included in the Clinical Study Report.

6.3.2. Biomarkers

Retrospective confirmatory testing of tumor tissue samples for ER status will be performed in a central laboratory designated by the sponsor using a validated 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 CCND1 and CDKN2A gene copy number, cdk4 and cdk6 RNA expression, and Ki67, pRb and p16 protein expression.

Genomic and metabonomic variations 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 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 (e.g., 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 Appendix 10.7 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). 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 10.8 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 and Stratification Factors

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

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

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

7. HANDLING OF MISSING VALUES

7.1. Missing Dates

In compliance with Pfizer standards, if the day of the month is missing for any date used in a calculation, the 1st of the month will be used to replace the missing date unless the calculation results in a negative time duration (e.g., 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 objective status cannot be a CR even if all target disease has disappeared. Otherwise, missing non-target disease assessments do not necessarily affect response determination. Such cases will be reviewed carefully.

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

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

7.3. Missing Data in PFS Derivation

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

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

For PFS analysis, no values will be imputed for missing data. For time to event endpoints, non-event observations will be censored as defined in Section 6.

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 all triplicate measurements are missing at a time point for an ECG parameter, no values will be imputed for this time point and no analyses related to this time point will be performed. If the triplicate is not good because of an artifact, then if the triplicate is repeated within about ± 15 minutes can be used at that nominal time. Patients who have data on other days or unscheduled ECGs but not at the times of the formal statistical analysis will be included in the categorical tables but not the statistical analyses.

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:

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

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. Gate-keeping procedure for primary endpoint PFS

At the final analysis of the primary endpoint PFS, the treatment effect of letrozole plus PD 0332991 versus letrozole plus placebo will be assessed in two ways.

(1) The treatment effect on PFS for all patients randomized in the study (ITT population).
(2) The treatment effect on PFS in a sub-group patients who administered PD-332991 with food.

The sub-group is defined as all patients who were randomized after January 21, 2014 when both the guidance of administration of PD-332991 with food and prohibition of PPI had been communicated to study sites. This is the patient population who administered PD-332991 with food.

A Gate-keeping procedure will be used for hypotheses testing in a hierarchical approach to control the family-wise error rate for the analyses of primary endpoint PFS. The significance level α for the test will be adjusted for interim analyses. The testing begins with comparing letrozole plus PD 0332991 treatment to letrozole plus placebo treatment in all patients randomized. This testing serves as a gatekeeper in the sense that the null hypothesis of "no treatment effect" must be rejected prior to proceeding to the next level of comparisons which will be the hypotheses testing for the sub-group population.

If the null hypothesis in ITT population can be rejected, then the study is claimed positive and the same hypotheses will be tested for the sub-group population with the same significant level as used for ITT population.

8.1.2. 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: *In a special case when PFS is statistically significant at T_1 and OS is not 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.
   Interim Analysis of OS at T_1,Test OS at $\alpha_o(t_o(T_1))$.
   If $u_o(t_o(T_1))$ is crossed, H_{0o} is rejected.
   If $u_o(t_o(T_1))$ is not crossed, go to 2.
   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 $\alpha_o(t_o(T_2))$ adjusting for $\alpha_o(t_o(T_1))$†.
   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.

Note: †$\alpha_o(t_o(T_2))$ needs to be adjusted for $\alpha_o(t_o(T_1))$ even under the scenario when OS was not tested at T_1.

8.1.3. 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 baseline 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 (subject to the multiplicity adjustment at the final analysis for PFS and OS) will be estimated using Cox proportional hazards regression.

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

8.1.4. 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) test stratified by Site of disease (one of the baseline stratification factors as listed in Section 6.4.2) and an 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.

8.1.5. Analyses of Continuous Data

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

8.1.6. Analyses for Categorical Data

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

8.1.7. 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.

8.1.8. Analyses for QTc Data

8.1.8.1. Derived Analysis Variables

All ECGs will be recorded in triplicate i.e. three ECGs taken approximately 2 minutes apart. Only the scheduled ECGs will be used for the formal analysis. ECGs obtained only for safety will not be included in the formal statistical analysis.

The data will be centrally read by Biomedical Systems (BMS) and will include the following parameters:

ECG Parameter	Units	Abbreviation
QTc, Fridericia's correction	msec	QTcF
QTc, Bazett's correction	msec	QTcB
QT Interval	msec	QT
Heart Rate	bpm	HR
PR Interval	msec	PR
RR Interval	msec	RR
QRS Complex	msec	QRS

The following variables will be derived as follows:

ECG Parameter	Units	Abbrev.	Derivation
QTc, study specific correction	msec	QTcS	$= QT/(RR)^{S}$ (derivation of S is given in section)

Averaging of triplicate measurements:

After the above variables have been derived within each patient and scheduled time-point, each ECG parameter (including QTcF, QTcB, QTcS, QT, HR, PR, RR, QRS) should each be averaged as follows: (1^{st} measurement + 2^{nd} measurement + 3^{rd} measurement) / 3. All summary statistics, analyses and figures will be based on the triplicate averaged data.

8.1.8.2. Derivation of Study Specific QT Correction Factor

The study specific QT correction factor (ICH E14 Step 4, May 12, 2005) will be derived. The dataset will consist of Day 0 all three measurements of QT and RR observed at the 0, 2, 4, 6 and 8 hour time points for all Group 1 patients.

Prior to estimating the regression, QT and RR will be transformed by the natural logarithm to Ln(QT) and Ln(RR), respectively. Ln (RR) will be treated in the regression as the

explanatory variable with Ln (QT) as the response variable. The regression equation will be as follows:

$$Ln(QT) = Intercept + S \times Ln(RR) + Error$$

Where S = slope of the regression line.

Once (S) is estimated, it will be used to derive the study specific QT correction, QTcS.

8.1.8.3. Assessment of QT Correction Methods

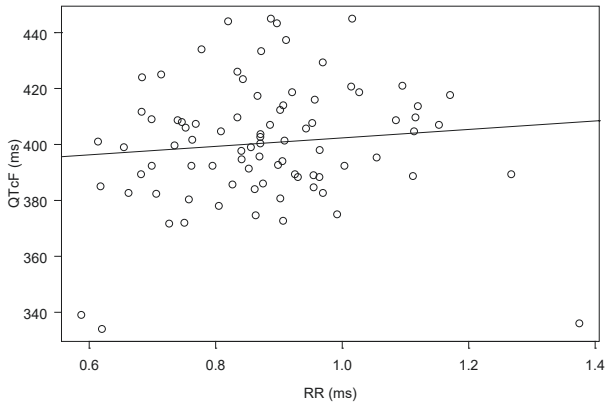
The relationship between QT/QTc and RR and the adequacy of each of the 3 QT correction methods will be assessed by the following scatterplots (see Figure 1):

- QT, QTcF, QTcB, and QTcS vs. RR

Interpretation of the scatterplots should give the following information:

- Variability of the QT/QTc relative to changing RR
- Slope or lack of slope and pattern to assess adequacy of QT correction method
- Apparent patterns in the data

Figure 1. A Sample scatterplot of relationship between QT/QTc and RR



8.1.8.4. Baseline Values for ECG Parameters

Time-matched baseline: Day 0 series of 5 QTc (QTcF, QTcB or QTcS as appropriate) measurements will be used as time-matched baseline values for the corresponding time-points (Hour 0 through Hour 8) on Day 14. This is matched within time-point and within patient.

8.1.8.5. Change from Baseline Definition

The change from baseline calculations for ECG measurements (abbreviated as “change”) is derived from the triplicate averaged measurements. Change from baseline is defined as a

patient's parameter value at a particular time-point minus the appropriately matched baseline value (value - baseline value). Change from baseline calculations should only use post-dose ECG measurements.

8.1.8.6. Change from Baseline Analysis

For Group 1, Cycle 1 Day 14 ECG assessment, a linear mixed effects modeling approach will be used to quantify the relationship between study drug plasma concentrations and ΔQTc (time-matched, baseline-adjusted). Based upon this relationship, the predicted population average ΔQTc and its corresponding upper 95% 1-sided confidence interval bound may be computed at appropriate concentrations, e.g., the mean maximum plasma concentrations under therapeutic doses. This analysis will be reported separately as a population modeling analysis report (PMAR).
As part of the CSR, for Groups 1 and 2, the following summaries by time-point will be presented by treatment group:

- the mean absolute QTc (QTcF, QTcB and QTcS) RR, PR and QRS, with two-sided 95% confidence intervals;
- the baseline-adjusted mean QTc (ΔQTcF, ΔQTcB and ΔQTcS) RR, PR and QRS, with two-sided 90% confidence intervals (Baseline is defined as the mean of triplicate measurements closest to the first palbociclib dose but prior to dosing. For group 1, baseline is time-matched, see protocol design for further details);
- Group 1 only: The point estimates of mean of change from baseline (time-matched) at all 5 time points and their 2-sided 90% confidence interval as error bars will be displayed graphically

8.1.8.7. Summary of Categorical Analysis

Shift summaries of QT/QTc values will be summarized and tabulated by the following CTCAE grade v.4.0 according to Day 0 or Day 14 with summary statistics as a supplement to the primary analysis.

Grade	1	2	3	4	5
Prolonged QTc interval	QTc 450 – 480 msec	QTc 481 – 500 msec	QTc ≥501 msec on at least two separate ECGs	QTc ≥501 or >60 msec change from baseline and Torsade de pointes or polymorphic ventricular tachycardia or signs/symptoms of serious arrhythmia	Death

The change from baseline will summarize occurrences of shift by ≥ 1 grade by CTC.
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.

8.1.9. Analyses for Ocular Assessment Data

8.1.9.1. Snellen Best Corrected Visual Acuity and Refraction

Snellen visual acuity charts are commonly used for safety evaluations. In the most familiar acuity test, a Snellen chart is placed at a standard distance, twenty feet in countries where that is the customary unit of measure. This line, designated 20/20, is the smallest line that a person with normal acuity can read at a distance of twenty feet. Three lines above, the letters have twice the dimensions of those on the 20/20 line. The chart is at a distance of twenty feet, but a person with normal acuity could be expected to read these letters at a distance of forty feet. This line is designated by the ratio 20/40. If this is the smallest line a person can read, the person's acuity is "20/40," meaning, in a very rough kind of way, that this person needs to approach to a distance of twenty feet to read letters that a person with normal acuity could read at forty feet.

In countries using the metric system, the standard chart distance is six meters, normal acuity is designated 6/6, and other acuities are expressed as ratios with a numerator of 6.

Besides Snellen fraction (such as 20/20, 20/40), visual acuity can also be specified with several other scales, for example the decimal visual acuity and the logarithm of the minimal angle of resolution (LogMar). Decimal visual acuity is obtained by dividing the numerator of the Snellen fraction by the denominator. The logarithm of the reciprocal of this decimal visual acuity approximates the logarithm of the minimal angle of resolution. The table below displays equivalent visual acuity measurements.

Equivalent Visual Acuity Measurements

Snellen Visual Acuities			Decimal Fraction	LogMar
4 Meters	6 Meters	20 Feet		

4/40	6/60	20/200	0.10	+1.0
4/32	6/48	20/160	0.125	+0.9
4/25	6/38	20/125	0.16	+0.8
4/20	6/30	20/100	0.20	+0.7
4/16	6/24	20/80	0.25	+0.6
4/12.6	6/20	20/63	0.32	+0.5
4/10	6/15	20/50	0.40	+0.4
4/8	6/12	20/40	0.50	+0.3
4/6.3	6/10	20/32	0.63	+0.2
4/5	6/7.5	20/25	0.80	+0.1
4/4	6/6	20/20	1.00	0.0
4/3.2	6/5	20/16	1.25	-0.1
4/2.5	6/3.75	20/12.5	1.60	-0.2
4/2	6/3	20/10	2.00	-0.3

In summary tables, proportion of subjects with visual acuity of improvement, no change, decrease by 1 line, 2 line and >2 lines will be displayed by treatment groups. If different visual acuity scales are used by different investigators, all visual acuity scores will be converted to LogMAR scale according the table above before data summarization. The change of 0.1 in LogMAR is equivalent to 1 line change Snellen chart.

$$\text{LogMAR} = - \log_{10}(\text{Decimal Fraction})$$

In the event of a decrease in visual acuity of 3 lines or more from baseline, refraction will be rechecked at all subsequent study visits. A change in refraction power (spherical or cylindrical) of +/– 1.25 diopters compared with the baseline examination will be reported in data listing by treatment groups.

8.1.9.2. Intraocular Pressure Measurement

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

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

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

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

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

8.1.10. 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).

8.2. Statistical Analyses

All efficacy analyses will be conducted on intent-to-treat (ITT) population and the sub-group population who were not impacted by antacids. Some efficacy analyses will also be performed on the MITT and AT populations, if appropriate. 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). No adjustments are planned for multiple testing/comparisons in the secondary and supportive hypothesis tests except OS.

8.2.1. Primary Efficacy Analysis

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

8.2.2. Sensitivity Analyses for the Primary PFS Endpoint

The primary efficacy analysis on the primary PFS endpoint is based on well-documented and verifiable progression events and deaths due to any cause. Other data are censored on the day following the date of the last tumor assessment documenting absence of progressive disease and death. In addition, several 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.

Influence of Disease Assessment Scheduling - A sensitivity analysis will be performed to investigate whether deviations in disease assessment scheduling influenced the outcome of the primary endpoint PFS. If disease progression is documented between 2 scheduled tumor assessments, then the date of progression will be assigned to the earlier scheduled tumor assessment. In the event of death, the date of the endpoint will not be adjusted. Handling of missed disease assessments will be similar to that in the primary analysis except that any missed assessment will result in censoring.

Influence of Deviations in Tumor Lesion Assessment - A sensitivity analysis will be performed to investigate whether deviations in tumor lesion assessment influenced the outcome of the primary endpoint PFS. If a lesion is classified as Unable to Evaluate (UE) at Time Point X, and is adequately evaluated as PD at the next Time Point (X+1), then PD will be assigned to the Time Point X or earlier (the first date of the consecutive UEs) instead of the date of the next Time Point (X+1) as the primary analysis.

Influence of Censoring for Patients Who Discontinued from Study due to Adverse Event or Systemic Deterioration – A sensitivity analysis will be performed to investigate whether the censoring for patients discontinued without PD due to adverse event or systemic deterioration of health. In this analysis, the patients who are censored in primary analysis will be counted as events.

Influence of Patients With Major Protocol Deviation – A sensitivity analysis to investigate whether the patients with major protocol deviation from inclusion and exclusion criteria affect the outcome of the primary PFS analysis by excluding those patients in the analysis.

Influence of Antacid Concomitant Medication (e.g. PPI) and /or Food Effect - Sensitivity analyses will be performed to investigate the impact of PPI, other antacids, and fasting condition on the treatment effect of the combination of PD-332991 with letrozole. PFS will be evaluated in the sub-populations excluding patients who took PPI, other antacids, and/or in fasting condition during the active treatment phase, if appropriate.

Influence of Bone-only Disease Patients - Sensitivity analyses will be performed to investigate whether addition of bone-only disease patients influenced the outcome of the primary endpoint PFS, if appropriate.

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

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

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

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

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

(4) Time to disease progression due to existing bone lesion progression, and/or new bone lesion.

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/country, and selected biomarkers on the primary PFS endpoint will be evaluated.

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.

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 (subject to the multiplicity adjustment at the final analysis) will be estimated.

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

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. CMH test stratified by Site of disease and exact test will be used to compare ORR between two treatment arms.

The number and proportion of patients achieving disease control/clinical benefit response (CR or PR and SD≥24 weeks) will be summarized in the ITT population and the ITT population with measurable disease at baseline along with the corresponding exact 2-sided 95% confidence interval calculated using a method based on Clopper-Pearson method. A stratified (by Site of disease) CMH test and exact test will be used to compare DCR/CBRR 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 with kappa statistics by treatment arms as well.

A supportive analysis will be performed by combining the OS data from this study and from the randomized phase 2 Study A5481003 with similar approaches described above. The Study as a stratification factor (A5481003 vs. A5481008) will also be included in the analysis. Since the median OS time for the studied patient population is relatively long and it is anticipated a small fraction of OS events would be available at the time of OS interim analysis, this analysis will certainly increase the power of detecting the OS difference between two treatment arms, given both randomized studies have similar patient populations and OS follow up processes.

8.2.5. Standard Analyses

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

- **Study Conduct and Patient Disposition** - an accounting of the study patients will be tabulated including randomized (per stratification factors), treated, accrual by study center, assessed for AEs, laboratory data, biomarkers, PK, and QTc, etc. Patients not meeting the eligibility criteria will be identified. Patients not completing the study will be listed along with the reason for their premature discontinuation. Reasons for premature discontinuation will be summarized. Randomization errors and stratification errors will be described.

- **Baseline Characteristics** - patient characteristics such as patient age, height, weight, race, ethnicity, ECOG performance status, primary diagnosis, 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 and percent of patients beginning 1, 2, 3, 4, 5+ cycles of either study drug
    - The number of cycles started (median, minimum, maximum) will be reported (overall and by study treatment).
    - Duration of treatment (weeks) (overall and by study treatment)

- Cumulative dose and relative dose intensity (see Appendix 10.6 for details) (overall and by cycle; by study treatment)

- **Treatment Delays and Dose Modifications**

  Dose 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 10.6 for details):

  - The number of patients with at least one PD 0332991 dose reduction and the number of patients with at least one PD 0332991 or letrozole dose omission at any time during drug administration will be reported.
  - The number of patients with at least one PD 0332991 dose reduction due to an adverse event will be reported.
  - The number of patients with at least one PD 0332991 dose delay (i.e. 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 (i.e. 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. Disease progression on follow up therapies will be collected according to protocol amendment 7. Disease progression on new anti-cancer therapies may be explored if sufficient data is collected and the analyses is appropriate.

8.2.6. Safety Analyses

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

8.2.6.1. Adverse Events

All patients treated with at least one dose of study treatment (i.e. PD-0332991/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).

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates this exploratory analysis.

Adverse events will be summarized by treatment and by the frequency of patients experiencing treatment emergent adverse events corresponding to body systems and MedDRA preferred term. Adverse events will be graded by worst NCI CTCAE v.4.0 grade. Adverse events will be summarized by cycle and by relatedness to trial treatment. Detailed information collected for each AE will include a description of the event, duration, whether the AE was serious, intensity, relationship to study drug, action taken, and clinical outcome. Emphasis in the analysis will be placed on AEs classified as treatment emergent.

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

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

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

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

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

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

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

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

- Disease progression will not be included

- There will be no adjustment for multiplicity

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

- Discontinuations Due to Adverse Events including causality: all cause, treatment related, including relationship to specific study treatment of letrozole, PD-0332991, and placebo

- Temporary Discontinuations or Dose Reductions Due to Adverse Events including causality and relationship to specific study treatment of letrozole, PD-0332991, and placebo

- Treatment-Emergent Adverse Events (All Causality, and Treatment Related) including the number of patients evaluable for adverse events, total number of adverse events (counting each unique preferred term across all patients), number of patients with serious adverse events, number of patients with Grades 3 and 4 adverse events, number of patients with Grade 5 adverse events, and number with dose reductions or temporary discontinuations due to adverse events

- Treatment-Emergent Adverse Events by MedDRA System Organ Class, Preferred Term and Maximum NCI CTCAE v.4.0 Grade (All Causality, and Treatment related)

- Treatment-Emergent Adverse Events by MedDRA Preferred Term sorted by Descending Order of AE Frequency (All Causality, and Treatment related)

- Treatment-Emergent Adverse Events by Preferred Term – Grade 3/4/5 events with number of patients experienced Grade 3-5 AEs and total number of Grade 3-5 AEs, sorted by Descending Order of AE Frequency (All Causality, and Treatment Related)

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

8.2.6.2. Laboratory abnormalities

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

8.2.7. 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.

For all patients in the safety analysis population, individual change in QTc (QTcF, QTcB, QTc S) will be calculated for each nominal post-baseline time point. These individual changes will be summarized using descriptive statistics.

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

Shift summaries of QT/QTc values will be summarized and tabulated by the following CTCAE grade v.4.0 according to 8.1.6.7.

1. The change from baseline will summarize occurrences of shift by ≥ 1 grade by CTC.

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

3. The number and percentage of patients with maximum post-dose QTcF/QTcB/QTcS (<450, 450-480, 481-500, and ≥501 ms), including all scheduled and unscheduled ECG's.

4. The number and percentage of patients with maximum increase from baseline in QTcF/QTcB/QTcS (<30, 30- 60, and > 60 ms), including all scheduled and unscheduled ECG's.

5. PR changes from baseline ≥ 50% if absolute baseline value was < 200 ms, and ≥ 25% if absolute baseline value was > 200 ms.

6. QRS changes from baseline ≥ 50% if absolute baseline value was < 100 ms, and ≥ 25% if absolute baseline value was > 100 ms.

The analyses described above for all patients will be repeated separately for the ECG sub-population in Group 1 for QTcF/QTcB/QTcS.

For safety analysis population, the following summaries by time-point will be presented by treatment group:

- the mean absolute QTc (QTcF, QTcB and QTcS) RR, PR and QRS, with two-sided 95% confidence intervals;
- the baseline-adjusted mean QTc (ΔQTcF, ΔQTcB and ΔQTcS) RR, PR and QRS, with two-sided 90% confidence intervals;
- Group 1 only: The point estimates of mean of change from baseline at all 5 time points and their 2-sided 90% confidence interval as error bars will be displayed graphically

8.2.8. Ocular Assessment Data Analyses

Snellen Visual Acuity data will be analyzed by using LogMAR value, reporting the maximum changes from baseline of LogMAR.

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.

8.2.9. Pharmacokinetic and Pharmacodynamic Analyses

Concentration data of PD-0332991 will be listed by patient and by actual collection time and day.

Average 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. PK parameters will be calculated for Group 1 patients on Day 14 of Cycle 1 using the predose PK value as the 24 hrs value (assuming steady state conditions). Summary statistics of the PK parameters will be provided. All patients treated with PD-0332991 and for whom drug plasma concentration results (from at least 1 visit) are available will be included in the analysis.

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

In addition, the relationship between exposure and PD, efficacy and safety endpoints will be explored, as necessary, based on emerging PD, efficacy/safety data.

The results of these modeling analyses may be reported separately from the clinical study report.

8.2.10. Biomarkers Analyses

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 PD-0332991 anti-tumor efficacy.

8.2.11. Patient Reported Outcomes Analyses

The analysis for PRO endpoints will be in PRO analyses set 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 questionnaire (FACT-B and EQ-5D) 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, at least one item within the questionnaire.

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 observed values as well as 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 mean FACT-G over time for each treatment group will also be provided.

8.2.11.1. 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.

8.2.11.2. 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

8.2.11.3. 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 for FACT-G. The MID for FACT-B is 7-8 points.

8.2.11.4. 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.

8.2.11.5. EQ-5D Index

Analysis of EQ-5D will follow the same methodology as for FACT-B. In addition, an EQ-5D health state profile, which consists of a display of the numbers and percentages of patients in each of the 3 response levels for each of the 5 dimensions, will also be provided.

8.2.11.6. EQ-VAS

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

8.2.11.7. 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 Analysis	Endpoint	Analysis Set	Statistical Method
Primary	PFS	ITT Investigator assessment (See 8.2.1)	Stratified log-rank test (stratified by disease site), K-M method (median and 95% CI) HR and 95% CI from the stratified Cox model
Secondary	PFS	ITT BICR	Stratified log-rank test (stratified by disease site), K-M method (median and 95% CI) HR and 95% CI from the stratified Cox model
Secondary *Sensitivity analysis1- 5*	PFS	ITT Investigator assessment and BICR (See 8.2.2)	Stratified log-rank test (stratified by disease site) K-M method (median and 95% CI) HR and 95% CI from the stratified Cox model
Secondary *Sensitivity analysis 6 - 7*	PFS	ITT Investigator assessment and BICR (See 8.2.2)	Unstratified log-rank test K-M method (median and 95% CI) HR and 95% CI from the unstratified Cox model
Secondary	PFS	ITT Investigator assessment And BICR	Unstratified log-rank test K-M method (median and 95% CI) HR and 95% CI from the unstratified Cox model
		MITT Investigator assessment and BICR	Stratified log-rank test (stratified by disease site) K-M method (median and 95% CI) HR and 95% CI from the stratified Cox model
		AT Investigator assessment and BICR (See 8.2.3)	Stratified log-rank test (stratified by disease site) K-M method (median and 95% CI) HR and 95% CI from the stratified Cox model
Secondary	OS	ITT (See 8.2.3)	Stratified log-rank test (1-sided, α=0.025), Cox model, K-M method (median and 95% CI) Wilcoxon test RPSFT method Survival probability at 1, 2, and 3 years
Secondary	OR	ITT Investigator assessment and BICR (See 8.2.3)	stratified exact tests (stratified by disease site,) Exact CI based on Clopper-Pearson method (95% CI),
Secondary	DC/CBR	ITT Investigator assessment and BICR (See 8.2.3)	stratified exact tests (stratified by disease site) Exact CI based on Clopper-Pearson method (95% CI),
Secondary	DR	ITT patients with a CR or PR Investigator assessments and BICR (See 8.2.3)	K-M method (median and 95% CI)

Abbreviations:
ITT: intent-to-treat; At: as-treated; MITT: modified intent-to-treat;
DCR: disease control rate; DR: duration of response; ORR: objective response rate; OS: overall survival; PFS: progression-free survival.

9. 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. Cui L, Hung HMJ, Wang S, et al. Modification of sample size in group-sequential clinical trials. Biometrics. 1999; 55: 853-7.
8. Dmitrienko A, Tamhane AC, Bretz F (2010). Multiple Testing Problems in Pharmaceutical Statistics. Chapman and Hall/CRC, Boca Raton.
9. Pampallona S and Tsiatis AA, Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. J. Statistical Planning and Inference, 1994; 42:19-35.
10. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. Brit. J. Radiology 1971; 44:793-797.
11. Peto, R., Pike, M. C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J., Smith, P.G. Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient: I. Introduction and Design. Brit. J. Cancer 1976; 34:585–612.
12. 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.
13. ICH Harmonised Tripartite Guideline. The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs. E 14. May 12, 2005.
14. National Cancer Institute (NCI). Common Terminology Criteria For Adverse Events (CTCAE). http://evs.nci.nih.gov/ftp1/CTCAE/About.html
15. 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.
16. 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.

17. 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.

10. APPENDICES

10.1. Schedule of Activities

The Schedule of Activities table provides an overview of the protocol visits and procedures. Refer to Study Procedures and Assessments for detailed information on each procedure and assessment required for compliance with the protocol.

A5481008 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		Cycles ≥3		
Study Day	Within 28 days prior to randomization unless specified otherwise	Day 1[b, v]	Day 14	Day 1[v]		
Time Window		±2d	±2d	±2d		±7d
Baseline Documentation						
Informed Consent Process[e]	X					
Medical / Oncological History[f]	X					
Baseline Signs / Symptoms		X[g]				
Retained Pharmacogenomic Blood Sample[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	X	X	
Blood Chemistry[l]	X	X[b]	X	X	X	
12-Lead ECG[m]	X[m]	X[b]	X[m]	X[n]	X	
Disease Assessment						

Protocol Activity	Screening	Active Treatment Phase[a] - One Cycle = 28 days			End of Treatment / Withdrawal[c]	Post-Treatment Follow-Up[d]
		Cycles 1 and 2		Cycles ≥3		
Study Day	**Within 28 days prior to randomization unless specified otherwise**	**Day 1[b, v]**	**Day 14**	**Day 1[v]**		
Time Window		**±2d**	**±2d**	**±2d**		**±7d**
Computed Tomograhy (CT)/ Magnetic Resonance Imaging (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[o]	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
EuroQol; EQ-5D[u]		X		X	X	
FACT - Breast Questionnaire[u]		X[u]		X	X	X[y]
Survival Follow-up						X
Study Treatment						
Randomization	X					
Letrozole (both treatment arms)		Once Daily ◄--►[w]				
PD-0332991 or Placebo		◄--►[x] Once Daily on Day 1 to Day 21 of each cycle followed by 7 days off				
Special Laboratory Studies						
Pharmacokinetics[m]			X			

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. For the purposes of this trial 1 cycle is 28 days. A cycle could be longer than 28 days if persistent toxicity delays the initiation of the subsequent cycle.

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). End of Treatment/Withdrawal visit will be performed as soon as possible but no later than 4 weeks (ie, 28 days) ±7days from last dose of study treatment and prior to the initiation of any new anticancer therapy.
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 (with the exception of certain imaging assessments if meeting the criteria defined in Section 6.1).
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. **Retained Pharmacogenomic Blood Sample:** 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 possible analysis of genetic associations with pharmacokinetics, drug response or adverse drug reactions. Examples of genes that may affect pharmacokinetics or drug response include, but may not be limited to, genes encoding drug metabolizing enzymes and transporters, and genes thought to be related to the mechanism of drug action.
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 core punches that will be used to generate a tissue microarray are needed. If FFPE tissue block cannot be submitted, at least 12 glass slides, each containing an unstained 5-micron FFPE tissue section, will be required for patient participation. Tissue sample from a metastatic or recurrent tumor lesion must be provided whenever possible. If such tissue sample is unavailable, a *de novo* fresh biopsy is recommended when, in the investigator's judgment, such biopsy is feasible and can be safely performed. A sample of the original diagnostic tissue (ie, archival) will also be collected when available and sent to the sponsor-designated central laboratories for assessment of biomarkers associated with sensitivity and/or resistance to PD-0332991 (eg, Ki67, CDKN2A (p16), pRb). Retrospective confirmation of ER status and prospective confirmation of HER2 status when required to be repeated for eligibility purpose will be performed using the most recent tumor sample. Original diagnostic tumor tissue will be used for confirmation of ER and HER2 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. Details on preparation of these samples including processing, storage, and shipment will be provided in the Study Manual.
j. **Physical Examination/Vital signs:** A full physical examination including an examination of all major body systems (including general appearance, head, ears, eyes, nose, mouth, throat, neck, thyroid, lungs, heart, breasts, abdomen, and musculoskeletal), 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:** Once Amendment #3 is IRB approved, all newly enrolled patients will undergo an ophthalmic examination at baseline, 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 protocol 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. Additionally, hemoglobin A1c will be measured during the active treatment phase in all patients every 3 months from the date of randomization (ie, C4D1, C7D1, C10D1, etc), and at the end of treatment visit.
m. **12-Lead ECG/Pharmacokinetics:** Refer to Pharmacokinetic and ECGs Schedule of Activities table for details and timing of procedures.
n. **12-Lead ECG:** To be performed on Day 1 of Cycles 4, 7, and 10. ECGs beyond Cycle 10 will be performed as clinically indicated. Refer to Pharmacokinetic and ECGs Schedule of Activities table for further details and timing of procedures.
o. **Disease Assessments:** Please refer to the tumor assessment requirement flowchart for details and timing of procedures.
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:** PD-0332991, placebo and letrozole bottle(s)/blisters 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. Following the active safety reporting period, other SAEs of which the investigator becomes aware should be reported to Pfizer, unless the SAE is attributed by the investigator to complications of either the underlying malignancy or any subsequent anti-cancer therapy or to the patient's participation in a subsequent clinical study. 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 PD-0332991/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. **Letrozole (both treatment arms):** To be taken orally, daily, and continuously.
x. **PD-0332991 or Placebo:**To be taken orally, daily from Day 1 to Day 21 (21 days) of every 28-day cycle followed by 7 days off treatment.
y. **FACT-B Assessments:** After patients discontinue from the active treatment phase, FACT-B questionnaire will continue to be collected during the follow-up period every 6 months (+/- 7 days) from the last dose of investigational product until patient permanent discontinuation from study or end of follow-up period whichever occurs first. During the follow-up period, all self-assessment questionnaires should preferably be completed by the patients during a scheduled clinic visit. However, if no clinic visits are being scheduled during the follow-up period interviewer administration via phone call may be used instead and documented accordingly in the patient source notes.

Pharmacokinetic and ECGs, Schedule of Activities															
Protocol Activity	**Screening**						**Active Treatment Phase**								**End of Treatment**
	Within 28 days prior to randomization unless specified otherwise						**Cycle 1**						**Cycle 2**	**Cycle ≥4**	
	Day -27 to Day 0	**Day 0 (day prior to Day 1)**					**Day 1**	**Day 14**					**Day 14**	**Day 1**	
	ECG_E*	**ECG1°**	**2 Hrs Post-ECG1**	**4 Hrs Post-ECG1**	**6 Hrs Post-ECG1**	**8 Hrs Post-ECG1**	**0 Hrs Pre-dose**	**0 Hrs Pre-dose**	**2 Hrs Post-dose**	**4 Hrs Post-dose**	**6 Hrs Post-dose**	**8 Hrs Post-dose**	**0 Hrs Pre-dose**	**0 Hrs Pre-Dose**	**Anytime**
Group 1 (approx. 60 Patients) at selected sites															
12 Lead ECG[a]	X	X	X	X	X	X	X[a]	X	X	X	X	X	X	X	X
Pharmacokinetics[b]								X	X	X	X	X	X		
Group 2 (All other patients)															
12 Lead ECG[a]	X						X[a]	X					X	X	X
Pharmacokinetics[b]								X					X		

***ECG_E** = Triplicate ECGs performed to determine patient's eligibility.
°ECG1 = First triplicate ECGs performed for patient in Group 1.

a. **12-lead ECG:** A 12-lead (with a 10-second rhythm strip) tracing will be used for all ECGs. ECGs will be performed in triplicate approximately 2 minutes apart but within 10 minutes for all 3 ECGs. It is preferable that the machine used has the capacity to calculate the standard intervals automatically. ECG interval readings by the ECG recorder's algorithm will be read and interpreted at the investigational site for eligibility determination and patient safety monitoring and documentation stored in the source documents. Blinded manual interval measurements at a core ECG laboratory will be used for primary statistical analysis of ECG data in group 1 patients (as described below). ECG measurements will include PR interval, QT interval, RR interval and QRS complex. Additional ECGs may be performed as clinically indicated.

Triplicate ECGs (also referred as ECG_E) will be performed during the screening period for all patients to determine the mean QTc interval for eligibility purpose.

Patients found to be eligible will be part of one of the two groups highlighted below depending on which site screened the patient. ECG frequency for each group is described below:

Group 1 (approximately 60 patients) at selected sites:

- On the day preceding treatment initiation (day 0), triplicate ECGs will be obtained at time 0 (first ECG also referred to as ECG1), and then 2, 4, 6, and 8 hours after ECG1.
- At Cycle 1 Day 14, triplicate ECGs will be obtained at 0 hour (pre-dose) and then, 2, 4, 6, and 8 hours following PD-0332991/placebo dosing. Timing of ECGs performed on Day 14 MUST be time-matched (clock time +/- 35 minutes) with ECG assessments performed on Day 0 (eg, if ECG1 on Day 0 was performed at 10:00 AM then the 0 hour triplicate ECGs on Day 14 must be performed within the 9:30AM-10:30AM timeframe but as close as possible to 10:00AM whenever feasible). On day 14 of Cycle 1, study treatment should be administered immediately after the pre-dose PK draw has been collected.

All ECGs should be obtained after a fast of at least 1 hour. ECGs should be performed immediately before PK blood draws at respective time points. Patients who cannot complete ECG measurements on Day 0 and/or PK-matched ECG measurements (both ECG and PK collected) on Day 14 will need to be replaced.

- Additionally, triplicate ECGs will be obtained for safety monitoring at 0 hour (pre-dose) on Day 1 of Cycle 1*, Day 14 of Cycle 2, then on Day 1 of Cycles 4, 7, and 10. ECGs beyond Cycle 10 will be performed as clinically indicated.

Group 2 (all other patients):

- Triplicate ECGs will be obtained for safety monitoring at 0 hour (pre-dose) on Day 1 of Cycle 1*, Day 14 of Cycles 1 and Cycle 2, then on Day 1 of Cycles 4, 7, and 10. ECGs beyond Cycle 10 will be performed as clinically indicated.

*NOTE: Triplicate ECGs do not need to be repeated on Day 1 of Cycle 1 if ECG_E was performed within 7 days of the date of randomization.

All triplicate ECG tracings will be sent electronically to a core ECG laboratory for blinded manual interval measurements.

If at any time the mean QTc is prolonged (≥501 msec on at least two separate ECGs, ie, CTCAE≥ Grade 3), then the ECGs should be re-evaluated by a qualified person at the site for confirmation as soon as the finding is made, including verification that the machine reading is accurate. If manual reading confirms a QTc of ≥501 msec, immediate search for reversible causes (including electrolyte abnormalities, hypoxia and concomitant medications for drugs with the potential to prolong the QTc interval) should be performed. In addition, repeat ECGs should be immediately performed hourly for at least 3 hours until the QTc interval falls below 501 msec.

- If QTc interval reverts to less than 501 msec, and in the judgment of investigator(s) in consultation with the sponsor the cause is determined to be due to cause(s) other than study drug, treatment may be continued with regular ECG monitoring under hospital supervision.
- If in that timeframe the QTc intervals remain above 501 msec the study drug will be held until the QTc interval decreases to <501 msec.

Prior to concluding that an episode of prolongation of the QTc interval is due to study drug, thorough consideration should be given to potential precipitating factors (eg, change in patient clinical condition, effect of concurrent medication, electrolyte disturbance) and possible evaluation by specialist. If investigational product causality cannot be ruled out, Investigational product dose adjustment and/or discontinuation should be performed according to protocol Section 5.3.4: Recommended Dose Modification. Additional triplicate ECGs may be performed as clinically indicated.

b. **Pharmacokinetics (collected right after ECG assessment):** For patients in the ECG Group 1, plasma PK samples for PD-0332991 (including its active metabolites, if appropriate) determination will be obtained at the times indicated for ECGs. One additional plasma PK sample will be collected pre-dose on Day 14 of Cycle 2.

   For all other patients (ECG Group 2), plasma PK samples for PD-0332991 (including its active metabolites, if appropriate) 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.

10.2. 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.

10.3. 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 examination 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 (e.g. '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 (e.g., 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

10.4. 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 (e.g., superficial lesions one day, scans another), the last date is used as the assessment date.

10.5. 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	For Group 1, the day preceding treatment initiation (Day 0) For Group 2, 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	ECG Group 1 (approximately 60 patients) ECG Group 2 (all other patients)
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.

10.6. Study Treatment Modification and Compliance

10.6.1. 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, PD-0332991/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.

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

The following types of summaries are proposed for administration of PD-0332991 (PD991) and letrozole.

- When PD991 is administered in combination with letrozole (orally once daily continuously), on an 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 PD991: by Cycle and Overall
- RDI for letrozole: by Cycle and Overall

Note: the denominator for tables summarizing "letrozole" will be all patients who took at least a dose of letrozole and for tables summarizing "PD991" will be all patients who took at least a dose of PD991

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 (2.5mg [once daily continuous] x 28 days for letrozole and 125mg [once a day for 21 days followed by 7 days break in a 28 days treatment cycle] x 21 days for PD991) 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 PD991
    - 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. (e.g. 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 per CRF.

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) Note: Calculation of RD is optional † = Calculated based on prescribed dose at the beginning of the study	• PD991 is to be dosed at 125 mg QD on a 3/1 Schedule • Actual PD991 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 of overall actual total dose) / (Sum of overall actual treatment duration) *Intended Overall Dose Intensity* = (Sum of overall Intended Total Dose during the study) /(Sum of overall Intended Treatment Duration)	• PD991 is to be dosed at 125 mg QD on a 3/1 Schedule • Actual PD991 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) /7 * (4 + 5 + 3) = 83.3 mg/day Intended Overall Dose Intensity=125 * (3 + 3 + 3)*7 /[4 + 4 + 3)*7] = 102.3 mg/day RDI = (83.3 / 102.3) *100% = 81.5%

Table 2

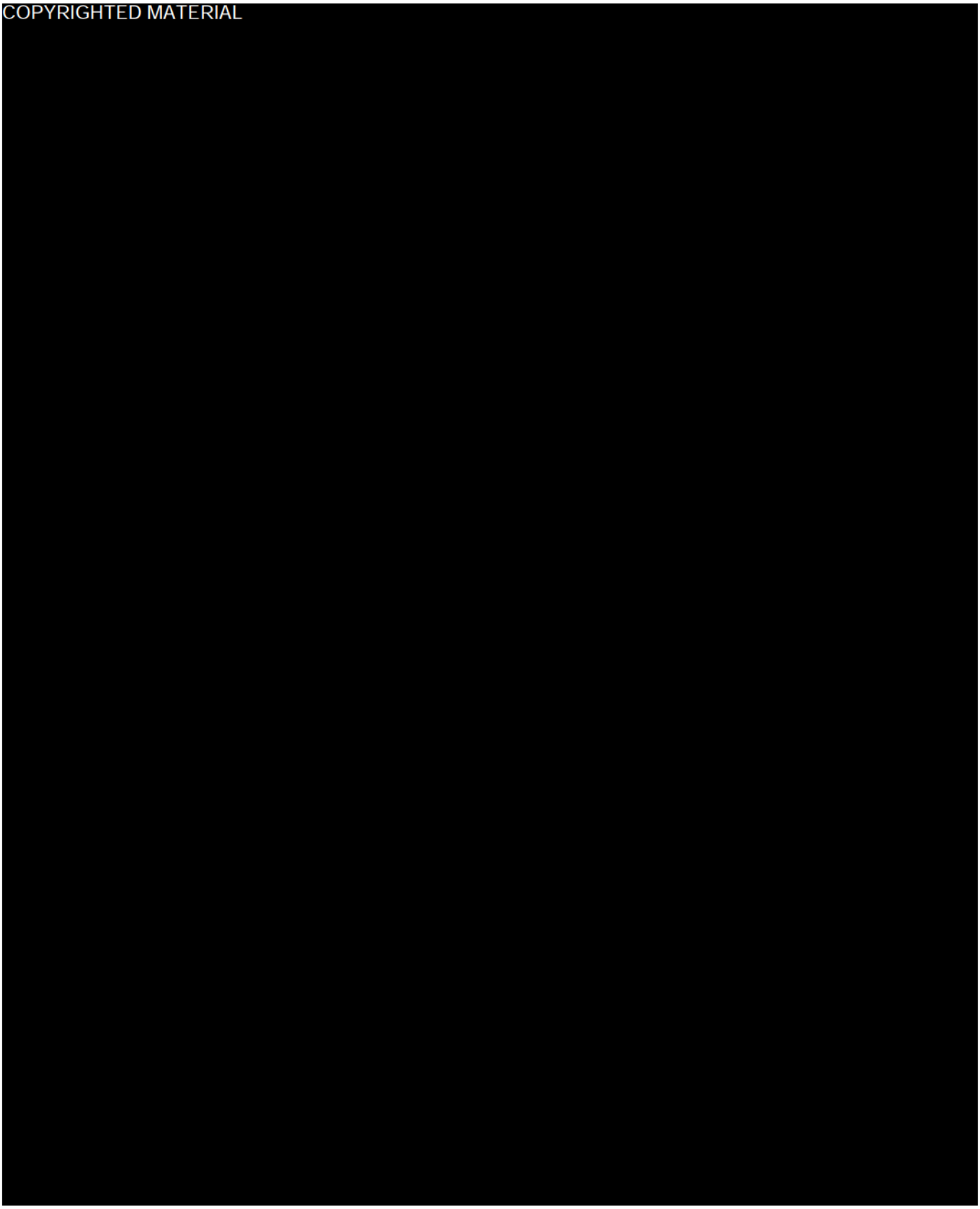
Treatment / Summary Type	Calculation of RD/RDI	Example
Cyclical PD991 / By Cycle	$RDI = \frac{Actual\ Dose\ Intensity}{Intended\ Dose\ Intensity} * 100\%$ *Actual Dose Intensity (per day)* = (Actual Total Dose per cycle) / (Actual treatment duration for the cycle) For a regular cycle *Actual treatment duration for the Cycle* = (Start date of next cycle – Start date of current cycle). For 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 for the 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]	• PD991 is to be dosed at 125 mg QD on a 3/1 Schedule • Actual PD991 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 <u>Cycle 1:</u> Actual Dose Intensity = (125*7*3) / (4*7) = 93.75 mg/day RDI = (93.75/93.75) * 100% = 100% <u>Cycle 2:</u> Actual Dose Intensity = (125*7*3) / (5*7) = 75 mg/day RDI = (75/93.75) * 100% = 80% <u>Cycle 3 (Last Cycle):</u> Actual Dose Intensity = (125*7*2) / (3*7) = 83.3 mg/day RDI = (83.3/125) * 100% = 66.7%

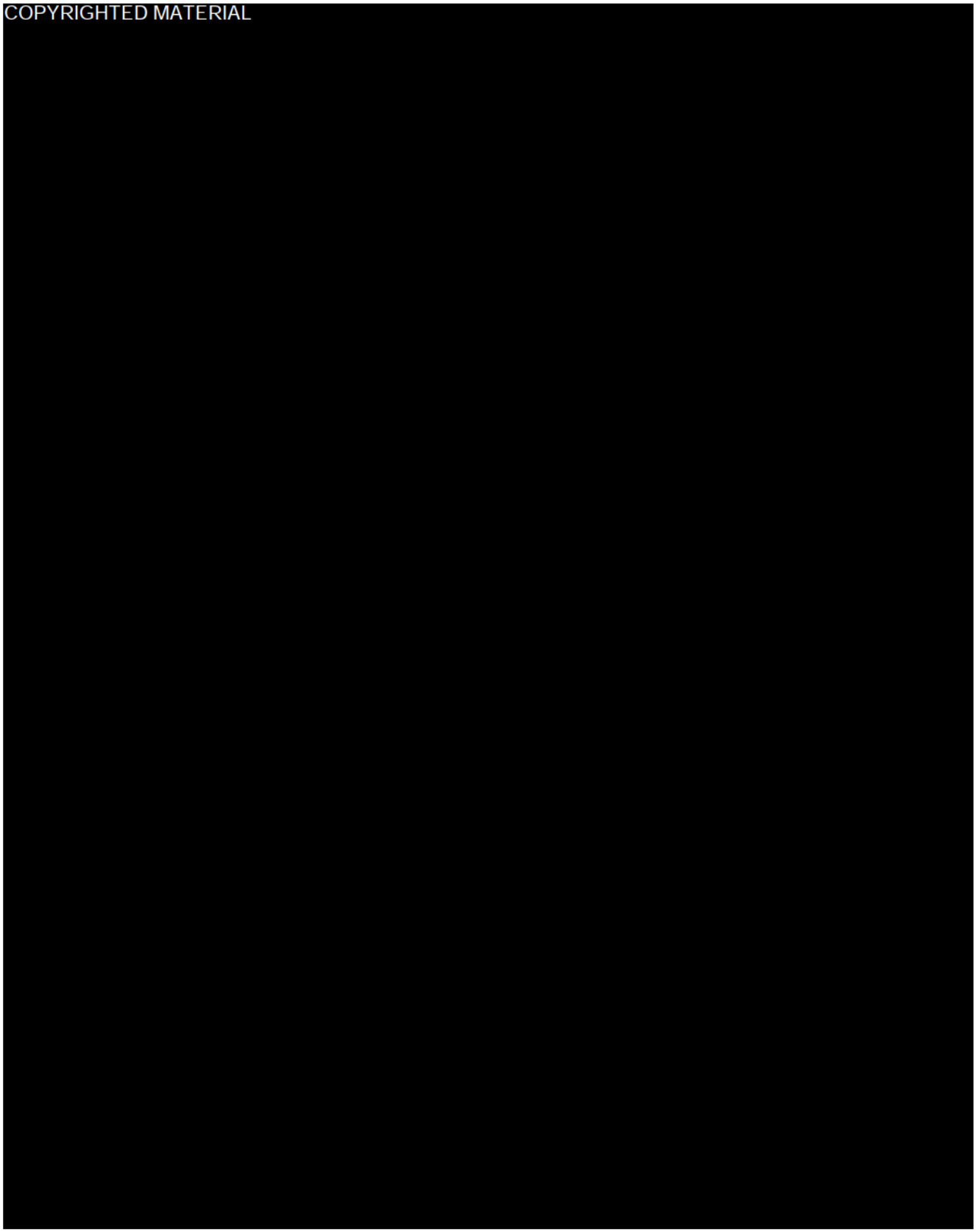
Table 3

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 day)* = (Actual Total Dose per cycle) / (Actual treatment duration for the cycle) *Actual Number of Days in Cycle* = (Start date of next cycle – Start date of current cycle). *Intended Dose Intensity (per day)* = (Intended Total Dose per cycle) / (Intended treatment duration for the 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 per cycle"*) / (Sum over all cycles of the "*Actual treatment duration for the cycle"*) *Intended Overall Dose Intensity = Intended Dose Intensity (per day)*	• 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 (i.e. 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 70mg in Cycle 1 and 52.5 mg in Cycle 2 <u>Cycle 1:</u> 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% <u>Cycle 2:</u> 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% <u>Overall:</u> 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%

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

COPYRIGHTED MATERIAL

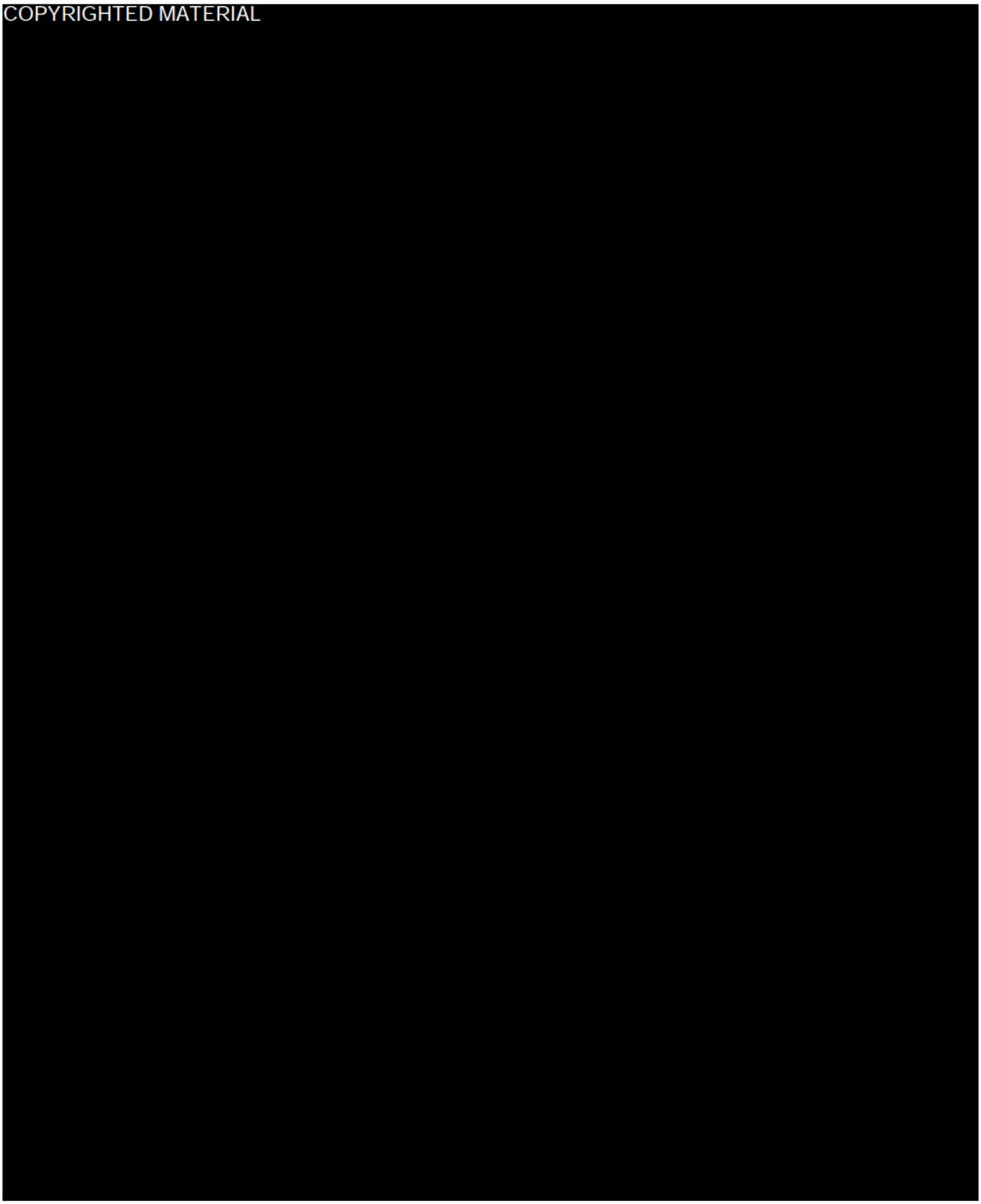


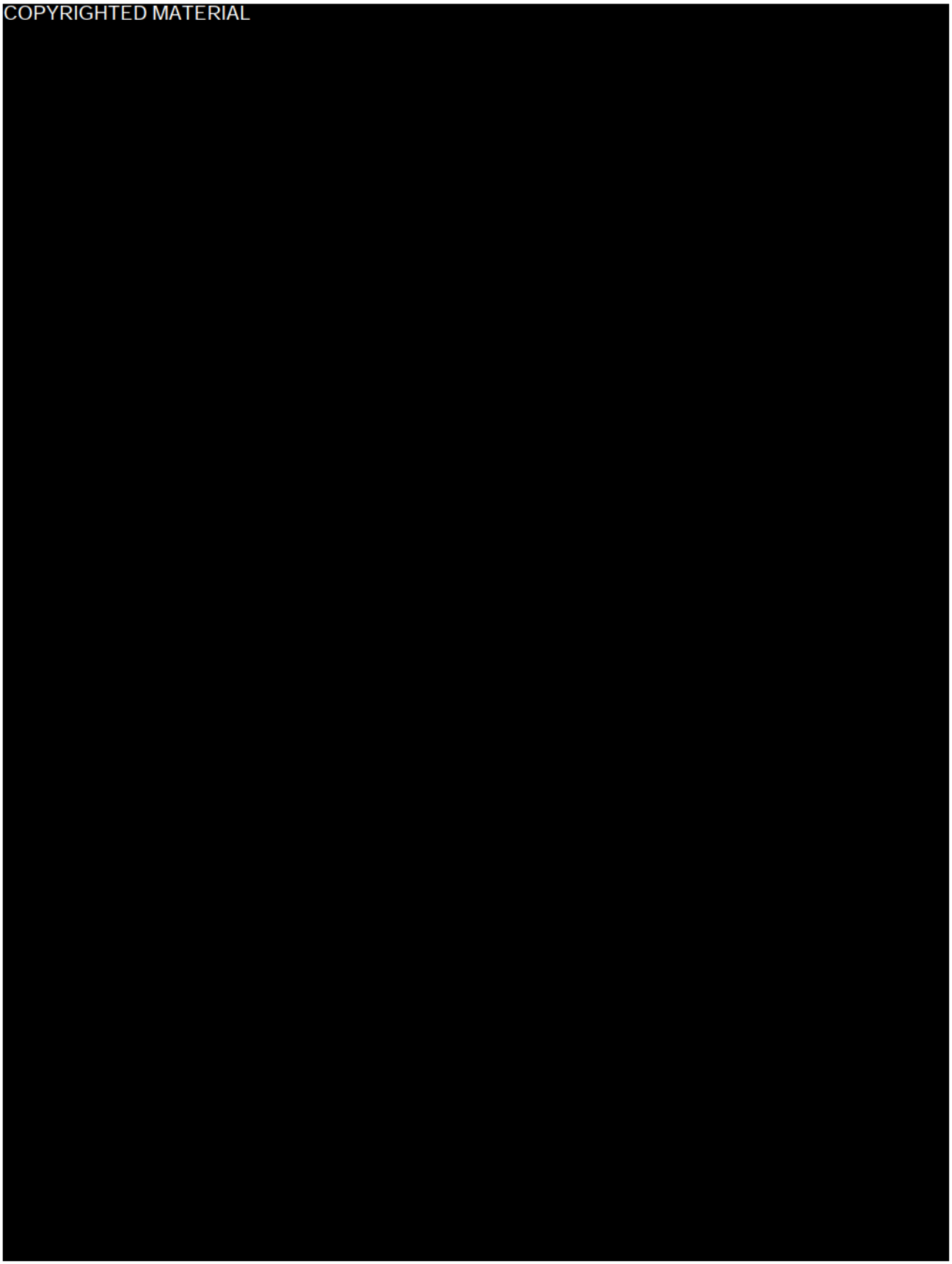



COPYRIGHTED MATERIAL

10.8. EuroQol Health Utilities Index EQ-5D

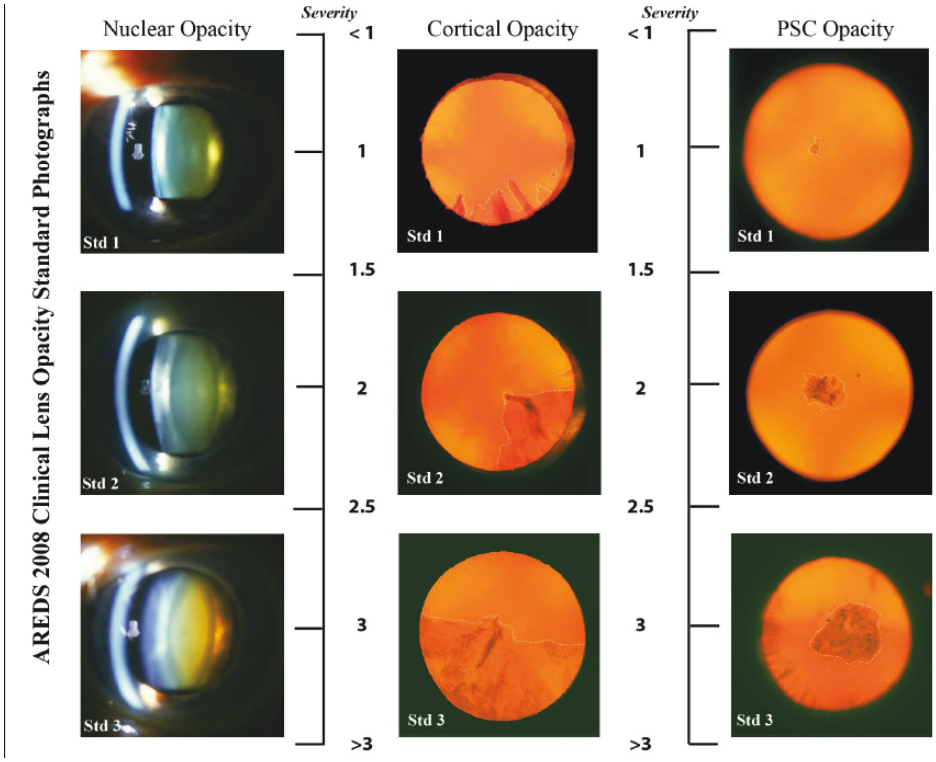
COPYRIGHTED MATERIAL





10.9. Wisconsin Age-Related Eye Disease Study (AREDS) 2008 Clinical Lens Opacity Grading Procedure

- Dilate pupils to at least 5 mm diameter
- Use slit lamp with ~10X magnification
- Use brightest beam intensity
- Nuclear opacity
    - Orient beam at 45° to viewing axis
    - Adjust slit beam to standard parameters: 8 mm height and 0.3 mm width
    - Compare opalescence of nucleus with that in standard photos
- Cortical and PSC opacities
    - Select wide slit beam setting optimum for retro-illumination of lens
    - Visualize lens opacities against red fundus reflex background
    - Count only opacities definitely visible against red reflex
    - Mentally combine all cortical opacities into one contiguous area
    - Compare total opacity area with that in standard photos
- Classify each opacity with scale defined by 3 standard photos
- Select nearest half-step
    - Similar to standard or between two standards
    - Obviously less than mildest standard or greater than most severe



10.10. List of Abbreviation

ABC	Advanced Breast Cancer
ASCO	American Society of Clinical Oncology
AE	Adverse Event
ALT	Alanine Aminotransferases
AI	Aromatase Inhibitor
ANC	Absolute Neutrophil Count
AREDS	Age-Related Eye Disease Study
AST	Aspartate Aminotransferases
AT	As Treated
AUC	Area Under the Curve
BC	Breast Cancer
BUN	Blood Urea Nitrogen
CAP	Chest, Abdomen, Pelvis OR College of American Pathologists depending on context.
CBR	**Clinical Benefit Response**
CBRR	**Clinical Benefit Response Rate**
CCND1	Cyclin D1
CDK	Cyclin-Dependent Kinase
CDKN2A, $p16^{Ink4A}$	Cyclin-Dependent Kinase Inhibitor 2A
CI	Confidence Interval
CISH	Chromogenic In Situ Hybridization
CLIA	Clinical Laboratory Improvement Amendments
C_{max}	Maximum Plasma Concentration
CMH	Cochran Mantel Haenszel
CNS	Central Nervous System
CR	Complete Response
CRF	Case Report Form
CSA	Clinical Study Agreement
CSF	Colony-Stimulating Factors
CT	Computed Tomography
CTA	Clinical Trial Application
CTCAE	Common Terminology Criteria for Adverse Events
CYP	Cytochrome P-450
DC	Disease Control
DCR	Disease Control Rate
DFI	Disease Free Interval
DLT	Dose Limiting Toxicity
DNA	Deoxyribonucleic Acid
DR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
E-DMC	External Data Monitoring Committee

EDR	The early discrepancy rate
EDTA	Ethylenediaminetetraacetic acid
EIU	Exposure In Utero
EQ-5D	Dimension Health State EuroQoL Score
ER	Estrogen Receptor
FACT	Functional Assessment of Cancer Therapy
FACT-B	Functional Assessment of Cancer Therapy - Breast
FACT-G	Functional Assessment of Cancer Therapy - General
FDA	US Food and Drug Administration
FFPE	Formalin Fixed Paraffin Embedded
FIH	First in Human
FISH	Fluorescent In Situ hybridization
GCP	Good Clinical Practice
G-CSF	Granulocyte Colony Stimulating Factor
GM-CSF	Granulocyte Macrophage Colony Stimulating Factor
Hb	Hemoglobin
HDPE	High Density Polyethylene
HER	Human Epidermal Growth Factor Receptor
hERG	Human Ether-à-Go-Go
HR	Heart Rate
IB	Investigator's Brochure
IC_{50}	Concentration of 50% Inhibition
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IEC	Independent Ethics Committee
IHC	Immunohistochemistry
INR	International Normalized Ratio
IOBU-SDMC	Internal Oncology Business Unit – Safety Data Monitoring Committee
IOP	Intraocular pressure
IMPAKT	Improving Care and Knowledge through Translational Research
IRB	Institutional Review Board
IRT	Interactive Randomization Technology
ITT	Intent-to-treat
LDR	The late discrepancy rate
LFT	Liver Function Test
LPD	Local Product Document
LSLV	Last Subject Last Visit
MedDRA	Medical Dictionary for Regulatory Activities
MITT	Modified Intent-to-Treat
MMRM	Mixed Model Repeated Measures
MRI	Magnetic Resonance Imaging
MTD	Maximum Tolerated Dose
NCI	National Cancer Institute

OBU	Oncology Business Unit
OR	Objective Response
ORR	Objective Response Rate
OS	Overall Survival
PCD	Primary Outcome Completion Date
PD	Progressive Disease
PET	Positron Emission Tomography
PFS	Progression Free Survival
PK	Pharmacokinetic
PR	Partial Response or Progesterone Receptor (depending on context)
PR	The PR interval is measured from the beginning of the P wave to the beginning of the QRS complex.
PS	Performance Status
PRO	Patient Reported Outcome
PT	Prothrombin Time
QD	Quaque Die (once daily)
QRS	The QRS complex is a name for the combination of three of the graphical deflections seen on a typical electrocardiogram. The QRS complex reflects the rapid depolarization of the right and left ventricles.
QT	Time between the start of the Q wave and the end of the T wave in the heart's electrical cycle
QT_c	QT interval corrected for heart rate
QTcB	QT interval corrected for heart rate using Bazett's fomula
QTcF	QT interval corrected for heart rate using Fridericia's fomula
QTcS	Study specific QTc: QT interval corrected for heart rate using the study
RANKL	Receptor Activator of Nuclear Factor Kappa B Ligand
RB/Rb	Retinoblastoma
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic Acid
RP2D	Recommended Phase 2 Dose
RPSFTM	Rank-Preserving Structural Failure Time Model
RR	The interval between an R wave and the next R wave
R_{ac}	Accumulation Ratio
SAE	Serious Adverse Event
SD	Stable Disease or Standard Deviation (depending on context)
SPC	Summary of Product Characteristic
$t_{½}$	Terminal Elimination Half-life
TdP	Torsade de Pointes
T_{max}	Time for C_{max}
ULN	Upper Limit of Normal
USPI	United States Package Insert

V_z/F	Apparent Volume of Distribution
WBC	White Blood Cell