

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*

**207103Orig1s000**

**STATISTICAL REVIEW(S)**



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation Research  
Office of Translational Science  
Office of Biostatistics

## STATISTICAL REVIEW AND EVALUATION CLINICAL STUDIES- ADDENDUM MEMO

**NDA /Serial Number:** 207103

**Drug Name:** Ibrance (Palbociclib)

**Applicant:** Pfizer

**Indication(s):** ER+, HER2- first-line metastatic breast cancer in combination with letrozole

**Date(s):** Submission Date: August 13<sup>th</sup>, 2014  
Original Review Date: January 15, 2015  
Date of Addendum: January 26, 2015

**Review Priority:** Priority

**Biometrics Division:** Division of Biometrics V

**Statistical Reviewer:** Erik Bloomquist, Ph.D.

**Concurring Reviewer:** Shenghui Tang, Ph.D., Team Leader

**Medical Division:** Oncology Drug Products 1

**Clinical Team:** Julia Beaver, M.D., Patricia Cortazar, M.D.

**Project Manager:** Amy Tiley

**Keywords:** Metastatic breast cancer, Accelerated approval, Unplanned interim analysis

In sections 3.4.1 and 3.4.5 of the original statistical review dated January 15, 2015, it was stated that the clinical review team did not review all patients in enrolled in the study. This statement is not correct. The FDA clinical review team reviewed all patients enrolled in the study.

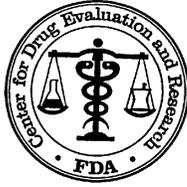
The sensitivity analysis mentioned in sections 3.4.1 and 3.4.5 only included the 96 patients where a discrepancy occurred between the blinded independent review committee and the investigator.

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

ERIK W BLOOMQUIST  
01/27/2015

SHENGHUI TANG  
01/27/2015



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Translational Sciences  
Office of Biostatistics

# STATISTICAL REVIEW AND EVALUATION

## CLINICAL STUDIES

**NDA Serial Number:** 207103  
**Drug Name:** Ibrance (Palbociclib)  
**Indication(s):** ER+, HER2- first-line metastatic breast cancer in combination with letrozole  
**Applicant:** Pfizer  
**Date(s):** August 13th, 2014

**Review Priority:** Priority

**Biometrics Division:** DB5  
**Statistical Reviewer:** Erik Bloomquist, PhD  
**Concurring Reviewers:** Shenghui Tang, PhD  
Rajeshwari Sridhara, PhD

**Medical Division:** OHOP/DOP1  
**Clinical Team:** Julia Beaver, MD  
Patricia Cortazar, MD

**Project Manager:** Amy Tiley

**Keywords:** Metastatic breast cancer, Accelerated approval, Unplanned interim analysis

# Table of Contents

<b>LIST OF TABLES.....</b>	<b>3</b>
<b>LIST OF FIGURES.....</b>	<b>4</b>
<b>1. EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>2. INTRODUCTION .....</b>	<b>5</b>
2.1 OVERVIEW.....	6
2.2 STUDY DESIGN AND CHANGES .....	6
2.3 DATA SOURCES .....	7
<b>3. STATISTICAL EVALUATION .....</b>	<b>7</b>
3.1 DATA AND ANALYSIS QUALITY .....	7
3.2 STUDY DESIGN .....	7
3.2.1 Enrollment, Treatment, and Randomization .....	7
3.2.2 Endpoints and Analysis Populations .....	8
3.2.3 Sample Size .....	9
3.2.4 Statistical Methodologies.....	9
3.3 PATIENT DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS .....	10
3.3.1 Patient Disposition, Demographic and Baseline Characteristics .....	10
3.3.2 Stratification Levels .....	11
3.4 PRIMARY ENDPOINT: PROGRESSION FREE SURVIVAL .....	12
3.4.1 Issues with Investigator-assessed PFS.....	14
3.4.2 Blinded Independent Review Committee .....	14
3.4.3 Differences between Investigator and BICR Assessment.....	16
3.4.4 Early and Late Discrepancy Rate.....	16
3.4.5 FDA Sensitivity Analysis.....	17
3.4.6 Additional Sensitivity Analyses .....	19
3.4.7 Overall Assessment of PFS.....	19
3.5 SECONDARY ENDPOINTS .....	19
3.5.1 Overall Survival.....	19
3.5.2 Objective Response Rate.....	21
<b>4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>21</b>
4.1 DEMOGRAPHIC AND BASELINE DISEASE CHARACTERISTICS .....	21
4.2 PROTOCOL VIOLATIONS AND FINANCIAL CONFLICTS OF INTEREST .....	22
4.3 BIOMARKER SELECTION .....	22
<b>5. SUMMARY AND CONCLUSIONS .....</b>	<b>23</b>
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE .....	23
5.2 CONCLUSIONS AND RECOMMENDATIONS .....	23

## LIST OF TABLES

Table 1: List of all studies included in analysis .....	6
Table 2: SAP Amendments .....	7
Table 3: Demographics.....	10
Table 4: Baseline-Disease Characteristics .....	10
Table 5: Stratification Factor Levels .....	11
Table 6: Investigator Assessed PFS.....	12
Table 7: BICR Assessed PFS .....	15
Table 8: Early and Late Discordance Rates.....	17
Table 9: FDA Sensitivity Analysis for PFS.....	18
Table 10: Overall Survival.....	20

## LIST OF FIGURES

Figure 1: PALOMA-1 Study Overview.....	8
Figure 2: Investigator Assessed PFS .....	13
Figure 3: Investigator Assessed PFS by Study Part.....	14
Figure 4: BICR Assessed PFS .....	15
Figure 5: FDA Sensitivity Analysis for PFS .....	18
Figure 6: Overall Survival .....	20
Figure 7: Investigator Assessed PFS Subgroup Analysis .....	21

## **1. EXECUTIVE SUMMARY**

Ibrance (Palbocicib) is new molecular entity for the treatment of metastatic breast cancer. Pfizer has submitted this NDA under the accelerated approval program based upon results seen in a phase 2 trial. Palbociclib was granted breakthrough therapy designation by the FDA in April 2013.

The sponsor is using the results from PALOMA-1 as the primary basis for accelerated approval. PALOMA-1 was an open-label, Phase 1/2, randomized study study of Palbocicib in combination with letrozole versus letrozole alone for the treatment of ER+, HER2-, metastatic breast cancer patients who had received no prior systemic therapy for their advanced disease.

At the outset, PALOMA-1 was not designed to be a registrational trial, and had several flaws of execution. The study also had several data-driven amendments to the statistical analysis plan. These included splitting the study into two parts, one an all-comers population and the other a biomarker selected population, and then recombining the two study parts back into one analysis population.

The primary analysis in PALOMA-1 was investigator assessed progression free survival (PFS). The results demonstrated a hazard ratio of 0.488 with a 20.2 month median PFS time in the Palbociclib + letrozole arm and 10.2 month median PFS time in the letrozole arm. A blinded independent committee review (BICR) of PFS found a hazard ratio of 0.621 with a 25.7 month median PFS time in the Palbociclib + letrozole arm and 14.8 month median PFS time in the letrozole arm.

The BICR PFS analysis was not completely consistent with the results for investigator assessment of PFS. Also, the BICR PFS analysis demonstrated some evidence of investigator bias towards the treatment arm in PALOMA-1. Sensitivity analyses, including a FDA review of case report forms and narratives, revealed that this bias did not completely confound the treatment effect of Palbociclib + letrozole.

There was positive trend in overall survival towards the Palbociclib + letrozole arm, but the data are immature at this time to make any positive conclusions.

Based upon the evidence in PALOMA-1, it appears that Palbociclib + letrozole treatment has a longer median PFS time than letrozole treatment. Due to the number of issues with the study, however, this magnitude of benefit is unclear and uncertain at this time. The final decision on the benefit-risk evaluation of palbociclib plus letrozole treatment for the patient population studied is deferred to the clinical review team.

## **2. INTRODUCTION**

Ibrance (Palbocicib) is new molecular entity for the treatment of metastatic breast cancer. Pfizer has submitted this NDA under the accelerated approval program based upon results seen in a

phase 2 trial. Palbociclib was granted breakthrough therapy designation by the FDA in April 2013.

## 2.1 Overview

Palbociclib is a cyclin-dependent kinases (CDK) inhibitor. Palbociclib stops cellular proliferation by prohibiting progression of the cell cycle from G1 into the S phase. Palbociclib is taken in capsule form orally with food. A recommended starting dose of Palbociclib is a 125mg capsule once daily with food for 21-days followed by 7 days off treatment.

The sponsor is using the results from study A5481003 (PALOMA-1) as the primary basis for accelerated approval. PALOMA-1 was a Phase 1/2 study used to assess the safety and efficacy of Palbociclib in combination with letrozole for the treatment of ER+, HER2-, metastatic breast cancer patients who had received no prior systemic therapy for their advanced disease. The Phase 1 portion of PALOMA-1 assessed safety and is not the primary basis of this review. Hereinafter, when referring to PALOMA-1, this review means solely the Phase 2 portion, unless otherwise mentioned.

**Table 1: List of all studies included in analysis**

Study	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
A5481003	Phase 1/2	Until disease progression or toxicity	OS until study end	84 P+L, 81 L	ER+, HER2- first-line mBC

P+L = Palbociclib + Letrozole; L = Letrozole

The Phase 2 portion of PALOMA-1 was an open-label, multicenter, randomized study comparing Palbociclib in combination with letrozole (P+L) versus letrozole (L) alone. There was no placebo given in the letrozole arm. The primary endpoint of the study was investigator-assessed progression free survival (PFS) with overall survival (OS), and overall response rate (ORR) being key secondary endpoints.

At the outset, PALOMA-1 was not designed to be a registrational trial, and had several flaws of execution. The study also had several data-driven amendments to the statistical analysis plan. Because these changes significantly affect the conclusions we can make from PALOMA-1, they are described in the next section.

## 2.2 Study Design and Changes

PALOMA-1 was originally designed to enroll 150 patients randomized 1:1 between a Palbociclib + letrozole (P+L) arm and a letrozole only (L) arm. While enrollment was underway, the sponsor altered the original protocol a total of 7 times. Major changes to the statistical analysis plan (SAP) were included in protocol amendments 3, 5, 6, and 7. Table 2 describes these amendments. Additional details on the study design follow in Section 3. The original protocol was finalized on March 27, 2008.

**Table 2: SAP Amendments**

SAP Amend#	Protocol Amend#	Date of Amend	Description of Changes
1	3	July 1, 2010	Phase 2 portion of study split into two parts: Part 1 (all-comers) and Part 2 (biomarker selected) based upon preclinical models. The biomarker selected population included patients with CCND1 amplification (CCND1 > 1.5) and/or CDKN2A loss (P16 < 0.8). Part 1 population enrollment halted at 66 and considered exploratory analysis population. Part 2 population set to be 150 patients and considered primary analysis population.
2	5	June 21, 2012	Primary analysis population becomes Part 1 combined with Part 2. Change due to interim look at Part 1 efficacy data.
3 <sup>a</sup>	6	Nov 8, 2012	Added BICR analysis for all patients enrolled in study.
3 <sup>a</sup>	7	July 11, 2013	Changed the final analysis time of PFS from 114 PFS events to 95 PFS events based upon slower than expected event rate.

a) SAP amendment #3 incorporated both protocol amendments #6 and #7.

### 2.3 Data Sources

The sponsor submitted data for PALOMA-1 electronically using the SDTM format. The sponsor also submitted their SAS analysis programs. The submission can be accessed at the following location: \\CDSesub1\evsprod\NDA207103\207103.enx.

## 3. STATISTICAL EVALUATION

This section focuses the efficacy results on PALOMA-1.

### 3.1 Data and Analysis Quality

The data submitted for PALOMA-1 was of good quality and sufficient for review purposes.

### 3.2 Study Design

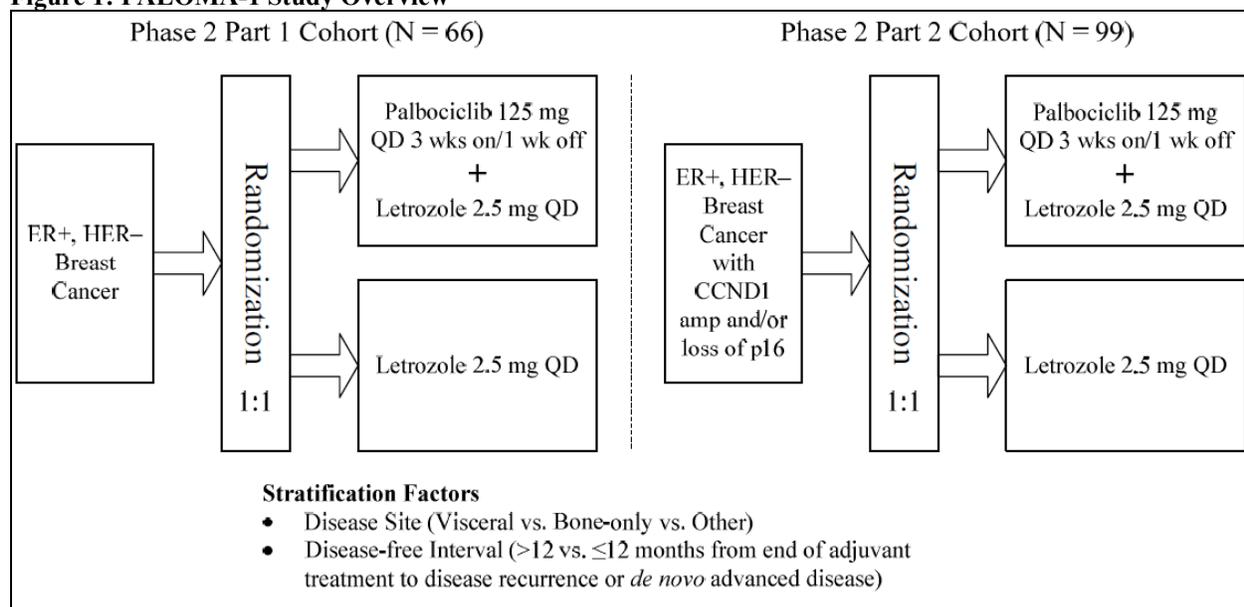
#### 3.2.1 Enrollment, Treatment, and Randomization

The Phase 2 portion of PALOMA-1 originally planned to enroll 150 patients, randomized 1:1 between a P+L arm and L only arm. The sponsor planned to use investigator-assessed PFS as the primary endpoint. Throughout the entire study, eligible patients for PALOMA-1 must have had an adenocarcinoma of the breast with either 1) locally recurrent disease not amendable for resection or 2) metastatic disease. Patients must also have been postmenopausal, have an ER+, HER2- tumor, and have no prior systemic therapy for advanced disease.

As described in Table 2 and Figure 1, the sponsor began enrollment (Part 1) in an all-comers population, and enrolled a total of 66 patients. Following this, the sponsor stopped enrollment in

Part 1 and began Part 2 of the study with the biomarker selected population (baseline  $CCND1 > 1.5$  or  $P16 < 0.8$ ). The sponsor enrolled a total of 99 patients in Part 2 of the study. Finally, after an interim analysis in Part 1 of the study, the sponsor combined Part 1 and Part 2 together to form the primary analysis population.

**Figure 1: PALOMA-1 Study Overview**



Source: CSR Section 7.1

In the P+L arm, the treatment schedule was Palbociclib administered once-daily for 3 weeks followed by 1 week off treatment. Letrozole was administered on a continuous dosing regimen once daily in both arms.

Throughout the study, the sponsor used two stratification factors for randomization: disease site (visceral disease vs. bone-only vs other) and time to recurrence (disease-free interval from adjuvant therapy to recurrence >12 months vs. ≤ 12 months). At the conclusion of the study, the sponsor found many instances where an incorrect stratification factor was used at the time of randomization. For time to recurrence, 22 individuals were misclassified at the time of randomization, and for disease-site 29 individuals were misclassified.

### 3.2.2 Endpoints and Analysis Populations

The final version of the SAP used investigator-assessed PFS in Part 1 and Part 2 combined. The SAP also detailed a hierarchical approach to testing PFS in Part 1 and Part 2 separately. The sponsor also outlined interim analyses for the primary analysis in the final SAP.

Documentation of a PFS event was made using the RECIST criteria. A screening scan was done within 4 weeks of starting treatment with follow-up scans every 8 weeks or if progression was suspected. Bone-lesion scans occurred every 12 weeks or when new metastases were suspected.

Additional secondary endpoints include BICR assessed PFS, OS, and ORR. The data cutoff for this study was in November 2013.

In the original protocol, the sponsor had an interim analysis for futility only. When the protocol and statistical analysis plan were modified in amendment #5, the sponsor removed the futility look and added two, and possibly three, interim efficacy looks. In amendment #7, the sponsor changed the timing of the final analysis. Due to the number of changes just mentioned, it is unclear how many times the sponsor looked at the study while it was ongoing.

Reviewer Comments:

- *It makes little sense to discuss the interim analyses for the primary analysis. The sponsor made several changes to the primary analysis population and changed the timing of the final analysis from 114 patients to 95 patients.*
- *Because this is an open-label small study with 165 total patients of investigator based PFS, FDA asked the sponsor to include a BICR analysis.*

### **3.2.3 Sample Size**

The sample size for this study was originally determined using a HR=0.67, with the P+L arm having a 13.5 month median PFS time and the L arm having a 9 month PFS time. A total of 114 events were needed to achieve 80% power to detect an HR=0.67 with a 1-sided alpha = 0.10. In protocol amendment 7, the sponsor changed the timing of the final analysis from 114 events to 95 events.

Reviewer Comments:

- *This study was not originally designed to be a registration study and thus the reason for setting alpha = 20% for a two-sided test.*
- *The sample size is too small for a confirmatory study.*

### **3.2.4 Statistical Methodologies**

The sponsor used a stratified log-rank test in the ITT population (Part 1 and Part 2 combined) with investigator-assessed PFS as the primary analysis. The sponsor used disease-site and time-to-recurrence as well as study part (Part 1 vs. Part 2) as stratification factors for the analysis.

When using the BICR-assessed PFS endpoint, the sponsor only used study part for a stratification factor.

For the secondary analyses, OS and TTP were analyzed using log-rank tests. ORR was assessed using a Cochran Mantel Hanzel test.

Reviewer Comments:

- *The sponsor pre-specified using only study part for the BICR analysis.*

### 3.3 Patient Disposition, Demographic and Baseline Characteristics

#### 3.3.1 Patient Disposition, Demographic and Baseline Characteristics

Patient demographics and baseline-disease characteristics can be found in Tables 3 and 4. For race and age, there appeared to be good balance between the two arms. There appeared to be good balance between the two arms for ECOG status, although ECOG status was not balanced for between Part 1 and Part 2. In addition, most of the patients from the United States enrolled in Part 1 (23/31). For baseline-disease characteristics, there was imbalance in histology status and progesterone receptor status. The median time since breast cancer diagnosis was higher in the letrozole only arm, but was not statistically significant, likely due to small sample size and wide range.

**Table 3: Demographics**

	Part 1		Part 2		Part 1 + Part2	
	P+L N=34	L N=32	P+L N=50	L N=49	P+L N=84	L N=81
Age (years)						
Median	65.5 (41 to 89)	64.0 (42 to 75)	62.0 (46 to 83)	63.0 (38 to 84)	62.5 (41 to 89)	64 (38 to 84)
≥ 65	17 (50%)	15 (46.9%)	20 (40%)	24 (49.0%)	37 (44.0%)	39 (48.1%)
Race						
White	31 (91.2%)	26 (81.3%)	45 (90.0%)	46 (93.9%)	76 (90.5%)	72 (88.9%)
Black	1 (2.9%)	1 (3.1%)	0	0	1 (1.2%)	1 (1.2%)
Asian	2 (5.9%)	1 (3.1%)	4 (8.0%)	3 (6.1%)	6 (7.1%)	4 (4.9%)
Other	0	4 (12.5%)	1 (2.0%)	0	1 (1.2%)	4 (4.9%)
ECOG						
0	23 (67.6%)	20 (62.5%)	23 (46.0%)	25 (51.0%)	46 (54.8%)	45 (55.6%)
1	11 (32.4%)	12 (37.5%)	27 (54.0%)	24 (49.0%)	38 (45.2%)	36 (44.4%)
Country						
USA	10 (29.4%)	13 (40.6%)	3 (6%)	5 (10.2%)	13 (15.5%)	18 (22.2%)

Source: CSR, Table 18, Table 19, Reviewer's Analysis

**Table 4: Baseline-Disease Characteristics**

	Part 1		Part 2		Part 1 + Part2	
	P+L N=34	L N=32	P+L N=50	L N=49	P+L N=84	L N=81
De Novo Disease (Yes)	19 (55.9%)	17 (53.1%)	25 (50%)	20 (40.8%)	44 (52.3%)	37 (45.7%)
Histology Grd						
1	6 (17.6%)	7 (21.9%)	2 (4.0%)	3 (6.1%)	8 (9.5%)	10 (12.3%)
2	9 (26.5%)	16 (50.0%)	22 (44.0%)	22 (44.9%)	31 (36.9%)	38 (46.9%)

3	14 (41.2%)	3 (9.4%)	17 (34.0%)	15 (30.6%)	31 (36.9%)	18 (22.2%)
Progest Recep						
Positive	23 (67.6%)	18 (56.3%)	42 (84.0%)	35 (71.4%)	65 (77.4%)	35 (65.4%)
Median Duration Since BC Diagnosis (Years)	0.9 (0 to 27)	3.4 (0 to 33.9)	1.5 (0 to 25)	2.1 (0 to 40)	1.3 (0 to 27)	2.4 (0 to 40)
Measurable Disease (Yes)	27 (79.4)	23 (71.9)	38 (76)	43 (87.8)	65 (77.4)	66 (81.5)

Source: CSR, Table 19, BC = Breast Cancer

### 3.3.2 Stratification Levels

Table 5 provides counts of the stratification factors used at the time of randomization and the levels based upon the case report forms. There were many misclassifications made for the stratification factors. For duration-since-diagnosis, there were a total of 22 misclassifications in Part 1 and 2 combined, equally balanced amongst the two arms. In 90% of these misclassifications, the randomization level used > 12 months, while the case report forms showed ≤ 12 months.

For disease-site, the P+L arm had 17 (20.2%) misclassifications, and the L arm had 12 misclassifications (14.8%). There did not appear to be any trends for misclassification in disease site.

#### Reviewer Comment

- *The number of misclassifications in the stratification factors is high. However, as shown later in this review, the misclassifications do not appear to affect the results.*

**Table 5: Stratification Factor Levels**

	Part 1		Part 2		Part 1 + Part2	
	P+L N=34	L N=32	P+L N=50	L N=49	P+L N=84	L N=81
Disease Site Based Upon Randomization						
Bone Only	8 (23.5%)	7 (21.9%)	9 (18.0%)	7 (14.3%)	17 (20.2%)	14 (17.3%)
Visceral	12 (35.2%)	11 (34.4%)	27 (28.0%)	29 (59.2%)	39 (46.4%)	40 (49.4%)
Other	14 (41.2%)	14 (43.8%)	14 (54.0%)	13 (26.5%)	28 (33.3%)	27 (33.3%)
Disease Site Based Upon Case Rpt Forms						

Bone Only	7 (20.6%)	6 (18.8%)	10 (20.0%)	6 (12.2%)	17 (20.2%)	12 (14.8%)
Visceral	10 (29.4%)	11 (34.4%)	27 (54.0%)	32 (65.3%)	37 (44.7%)	43 (43.1%)
Other	17 (50.0%)	15 (46.9%)	13 (26.0%)	11 (22.5%)	30 (35.7%)	26 (32.1%)
Disease Free Interval Based Upon Randomization						
≤ 12 mths	20 (58.8%)	19 (59.4%)	27 (54.0%)	26 (53.1%)	47 (56.0%)	45 (55.6%)
> 12 mths	14 (41.2%)	13 (40.6%)	23 (46.0%)	23 (46.9%)	37 (44.0%)	36 (44.4%)
Disease Free Interval Based Upon Case Report Forms						
≤ 12 mths	24 (70.6%)	22 (68.8%)	35 (70.0%)	29 (59.2%)	59 (70.2%)	51 (63.0%)
> 12 mths	10 (29.4%)	10 (31.2%)	15 (30.0%)	20 (40.8%)	25 (29.8%)	30 (37.0%)

Source: Reviewer's Analysis

### 3.4 Primary Endpoint: Progression Free Survival

The primary endpoint for this study was investigator-assessed progression-free survival. Table 6 and Figures 2 and 3 present the efficacy results based upon the sponsor's primary analysis plan. Due to the data-driven changes in SAP, p-values are only reported as less than 0.01 and are classified as nominal. Sensitivity analyses for the primary endpoint are detailed in the next several sections.

#### Reviewer Comment

- *The results of the primary analysis suggest longer a PFS with Palbociclib + letrozole compared to letrozole treatment. Because the goal of this trial was to explore, a number of data-driven changes were made to the protocol. Therefore, from a statistical point of view, inference cannot be drawn.*
- *Due to the number of data-driven changes to the SAP, p-values are only reported as approximate and nominal.*
- *The results clearly suggest an impact of SAP amendments on the outcome. Therefore, pooling Part 1 and Part 2 is problematic.*

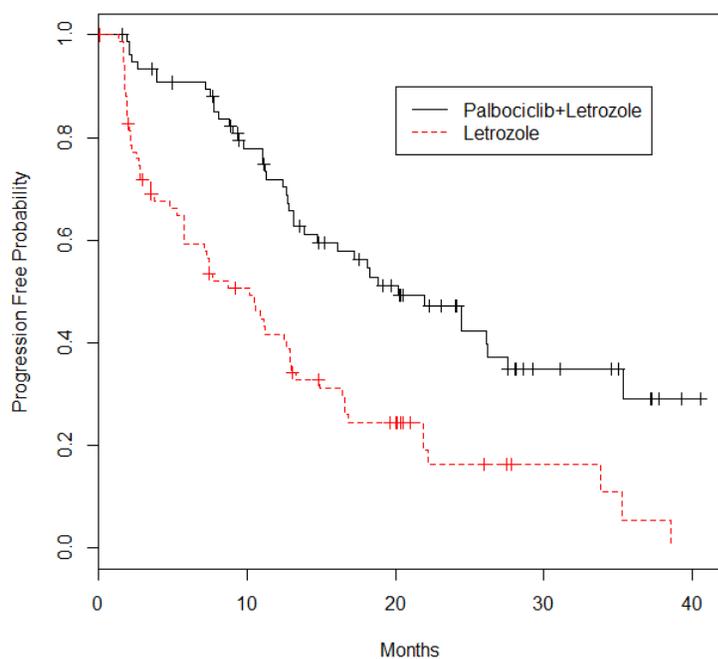
**Table 6: Investigator Assessed PFS**

	Part 1		Part 2		Part 1 + Part2	
	P+L N=34	L N=32	P+L N=50	L N=49	P+L N=84	L N=81
Number of events	15(44.1%)	25 (78.1%)	26 (52.0%)	34 (69.4%)	41 (48.8%)	59 (72.8%)
Censored	19 (55.9%)	7 (21.9%)	24 (48.0%)	15 (30.6%)	43 (51.2%)	22 (27.2%)
Median PFS	26.1	5.7	18.1	11.1	20.2	10.2

(months)				
95% CI	(11.2, NR)	(2.6, 10.5)	(13.1, 27.5)	(7.1, 16.4)
Hazard Ratio	0.299		0.508	0.488
95% CI	(0.156 – 0.572)		(0.303 – 0.853)	(0.319 – 0.748)
Nominal p-value	<0.01		<0.01	<0.01

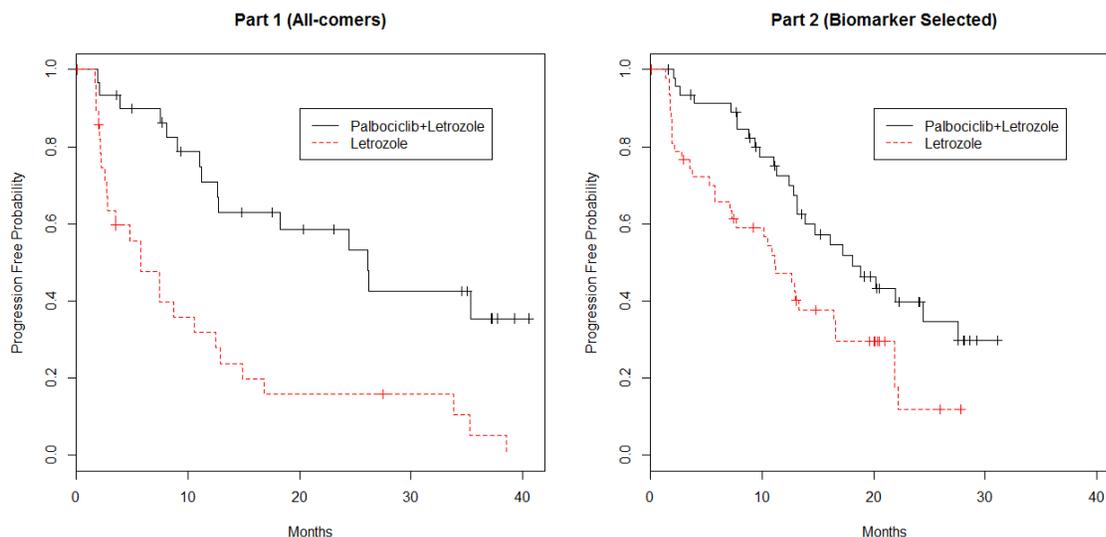
Source: Reviewer’s analysis

**Figure 2: Investigator Assessed PFS**



Source: Reviewer’s analysis

**Figure 3: Investigator Assessed PFS by Study Part**



Source: Reviewer's analysis

### 3.4.1 Issues with Investigator-assessed PFS

The primary analysis used investigator-assessed PFS in an open-label Phase 2 study. Due to the possibility of bias in an open-label study, and at the FDA's request, the sponsor conducted a 100% BICR review of all patients entered into the study. The BICR results (described in the next section) show that Palbociclib has a lesser estimated effect than the investigator assessed PFS analysis. In addition, the BICR analysis demonstrates some investigator bias towards the P+L arm over the L arm.

To help analyze differences between the investigator-assessment of PFS and the BICR-assessment of PFS, the FDA clinical review team reviewed the case report forms of 96 individuals where the investigator and BICR did not agree on the timing or censoring of the PFS events. As the results will show, the FDA analysis fell closer to the investigator-assessment of PFS. The FDA clinical review team did not review all patients in the study.

In addition to issues with PFS mentioned above, the study randomized many patients using incorrect stratification levels. Sensitivity analyses to follow assess the effect of these misclassifications on the primary analysis.

### 3.4.2 Blinded Independent Review Committee

Results of the BICR assessment of PFS can be found in Table 7 and Figure 4. As shown in Table 7, the p-value for Part 1 and Part 2 combined did not fall below 0.01, nor did either Part 1 or Part 2. In the sponsor's CSR, the sponsor used a 1-sided log-rank test, but to be fair, the p-values should be based upon the customary stratified 2-sided log-rank test. The difference in median PFS was 10 months, similar to the investigator-assessment of PFS.

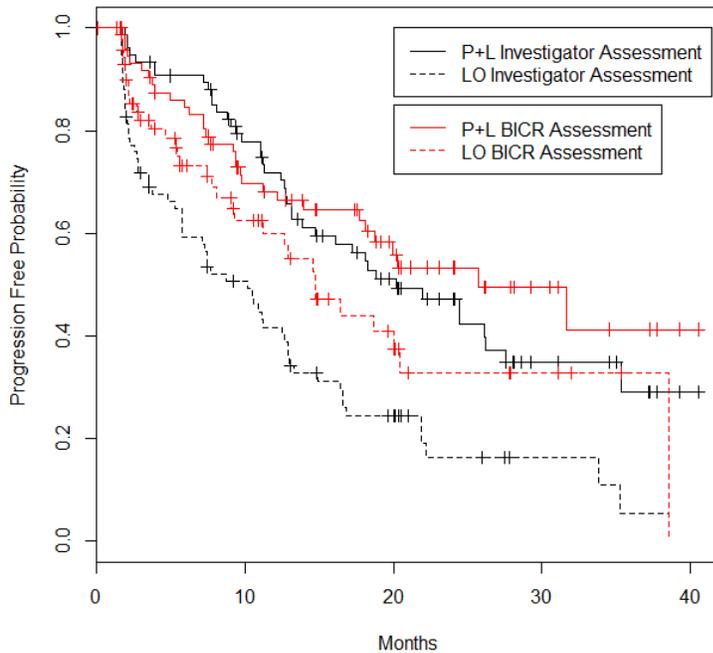
A Kaplan-Meier plot of the BICR-assessment of PFS is presented in Figure 4. Similar to the investigator-assessment of PFS, the BICR-assessment in Figure 4 illustrates longer PFS with P+L compared to L.

**Table 7: BICR Assessed PFS**

	Part 1		Part 2		Part 1 + Part2	
	P+L N=34	L N=32	P+L N=50	L N=49	P+L N=84	L N=81
Number of events	11 (32.4%)	9 (28.1%)	20 (40%)	24 (49%)	31 (36.9%)	33 (40.7%)
Censored	23 (67.6%)	23 (71.9%)	30 (60%)	25 (51%)	53 (63.1%)	48 (59.3%)
Median PFS (months)	31.6	38.6	20.3	14.6	25.7	14.8
95% CI	(11.2, NR)	(7.5, 38.6)	(12.2, NR)	(8.1, 20.0)	(17.7, NR)	(9.3, 20.4)
Hazard Ratio	0.731		0.576		0.621	
95% CI	(0.300, 1.779)		(0.316, 1.050)		0.378	
Nominal p-value	0.4902		0.0717		0.0595	

Source: Reviewer's analysis

**Figure 4: BICR Assessed PFS**



Source: Reviewer's analysis

#### Reviewer Comment

- *The BICR assessment shows a 10 month difference in PFS between the P+L and L arms.*

### **3.4.3 Differences between Investigator and BICR Assessment**

The BICR assessment and investigator assessment of PFS both showed a longer estimated PFS with P+L compared to L, but the results do not match up completely. One possible reason for this is due to the censoring rate, especially in the L arm. In the P+L arm, the investigator PFS censoring rate was 51.2%, while the BICR censoring rate was 63.1%, for a difference of 11.9%. But in the L arm, the investigator PFS censoring rate was 27.2%, while the BICR censoring rate was 59.3%, a 32.1% difference.

One reason for the difference between the two arms, per the clinical review team, may be due to progression determination in bone only disease. Under the RECIST criteria, progressive disease classification is difficult, and the BICR may have had a more difficult time classifying events as PD in the bone only disease setting.

Looking at the study results, in the P+L arm, there were a total of 6 bone-only disease participants with investigator determined progressive disease (PD) and BICR determined stable disease (SD), i.e. censoring. In the L arm, however, there were a total of 14 bone-only disease patients having investigator determined PD and BICR determined SD.

If we classify these 20 bone-only disease patients using the progressive events, the BICR censoring rate in the P+L arm equals 56.0%, a 4.8% difference now from the investigator censoring rate, and in the L arm 42.0%, a 14.8% difference. Moreover, a stratified log-rank sensitivity analysis using the reclassified events shows no difference in the overall conclusions.

#### Reviewer Comment

- *The bone-only disease patients partially help to explain censoring differences between the investigator-assessed PFS and BICR-assessed PFS.*
- *The analyses however do not fully explain differences.*

### **3.4.4 Early and Late Discrepancy Rate**

The difficulties surrounding bone only disease helps to partly explain the differences between the investigator assessment of PFS and the BICR assessment of PFS. But there still remains a discordance that cannot be explained.

In particular, there exists discordance between investigators claiming SD and the BICR determining PD. In the P+L arm, there were 9 events where the investigator censored an individual and the BICR determined progression. But in the L arm, there was only 1 event where investigator censored an individual and the BICR determined progression.

To help quantify the level of discordance between the investigator-assessed PFS and BICR-assessed PFS, we can use the difference in late discordance rate (LDR) and early discordance rate (EDR) (Amit O. et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis. *European Journal of Cancer*. 47, 1772-8.) The late discordance rate is the rate that the investigator determines progression later than the BICR as a proportion of the number of discordances. The EDR measures the rate of investigator determined progression earlier than the BICR, as a proportion of the number of investigator determined progression events. Importantly, if the LDR in the P+L arm is higher than in the L arm, or if the EDR in the P+L arm is lower than in the L arm, it signifies a level of investigator bias towards the treatment arm.

Table 8 displays the LDR and EDR rates in each arm for the full Phase 2 study and the Phase 2 study minus the bone only progressions, discussed in section 3.4.3. As shown, the EDR rates were nearly equal between the two arms. Moreover, a randomization test (using 10,000 randomizations of the assigned treatments), demonstrates that the EDR difference between the two groups falls near the center of the distribution.

For the LDR, however, there does appear to be investigator bias towards the treatment group. The difference in the LDR between the two arms was 22.5% and in a randomization test, this difference falls in the 98% percentile of the distribution.

Reviewer Comment

- *The issues with bone only disease helps partly explain the differences between the investigator assessment of PFS and BICR assessment of PFS.*
- *There still appears to be an unquantifiable investigator bias towards the P+L arm over the L arm.*

**Table 8: Early and Late Discordance Rates**

	Part 1 + Part 2		Part 1 + Part 2 – Bone Only	
	P+L N=84	L N=81	P+L N=78	L N=67
EDR	46.3%	50.8%	37.1%	35.6%
Difference		-4.5%		1.6%
Rand Test Quantile		32%		55%
LDR	55.8%	33.3%	64.9%	48.4%
Difference		22.5%		16.5%
Rand Test Quantile		98%		91%

Source: Reviewers Analysis. EDR = Early discrepancy rate, LDR = Late discrepancy rate. The rand test quantile represents the quantile of the observed EDR or LDR rate in 10,000 simulated samples of the randomized treatment assignments.

**3.4.5 FDA Sensitivity Analysis**

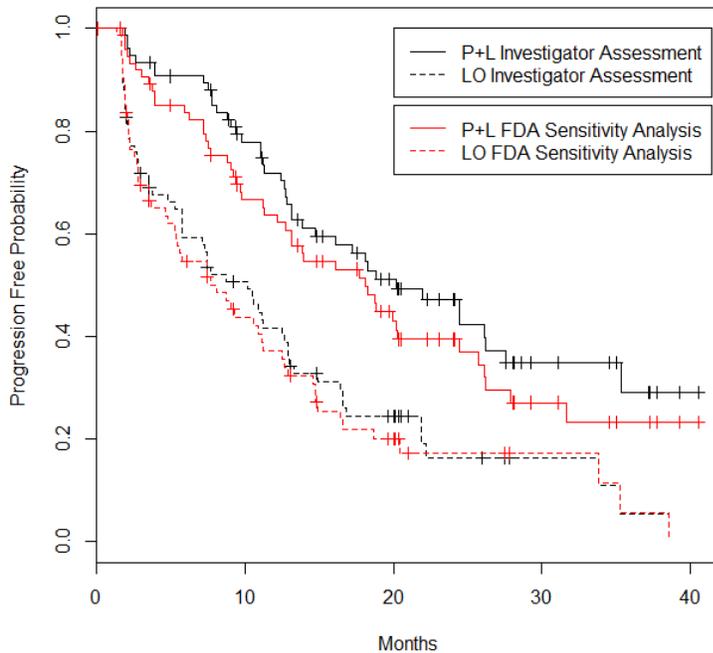
Since there appears to exist some level unexplained of investigator bias towards the investigational arm, the FDA clinical review team went back and reviewed the case report forms

and patient narratives for the 96 subjects where the BICR and investigator did not agree on the PFS time or event status. The results of their analysis can be found in Table 9 and Figure 5. The FDA sensitivity analysis (events as confirmed by the clinical review team) was consistent the investigator assessment of PFS.

**Table 9: FDA Sensitivity Analysis for PFS**

Part 1 + Part 2		
	P+L N=84	L N=81
Number of events	47 (55.9%)	57 (70.4%)
Censored	37 (44.1%)	24 (39.6%)
Median PFS (months)	18.1	8.1
95% CI	(12.1 – 24.4)	(4.8 – 11.2)
Hazard Ratio	0.574	
95% CI	(0.378-0.870)	
p-value	<0.01	

**Figure 5: FDA Sensitivity Analysis for PFS**



Source: Reviewer's analysis

#### Reviewer Comment

- *The FDA sensitivity analysis is an exploratory analysis. The Kaplan-Meier curves for the sensitivity analysis follow a similar trend as the investigator assessment of PFS.*
- *No statistical inference can be drawn based on these exploratory analyses.*

#### **3.4.6 Additional Sensitivity Analyses**

To further assess the primary endpoint, three additional sensitivity analyses were conducted. The first started with the investigator assessment of PFS. Then for all patients in the treatment arm, the sensitivity analysis classified patients as having progressive disease, unless they were censored at the end of study. No changes were made to patients in the L arm. In this sensitivity analysis, the median PFS time in the P+L arm was 13.1 months (11.0, 17.5), and the median PFS time in the L arm was = 10.2 months (5.7 12.6). The hazard ratio = 0.79.

#### Reviewer Comment

- *In this scenario, the P+L arm had a longer PFS time than the L arm.*

The final two sensitivity analyses focused on the wrong stratification factors being used at the time of randomization. The first analysis used an unstratified log-rank test using the investigator assessment of PFS. The hazard ratio for this analysis was HR=0.412. The second analysis used the stratification factors based upon the case report forms (HR = 0.459 (0.30, 0.70)).

#### Reviewer Comment

- *These two analyses demonstrate that the use of the wrong stratification factors did not appear to affect the PFS results seen elsewhere.*

#### **3.4.7 Overall Assessment of PFS**

Although no statistical inference can be made, Palbociclib + Letrozole treatment appears to have a longer PFS time than Letrozole treatment. The true increase in PFS time, however, remains uncertain due to the many issues mentioned above. There does appear to be some investigator bias towards the P+L arm.

### **3.5 Secondary Endpoints**

The sponsor used several additional secondary endpoints. This section discusses the overall survival endpoint, and objective response rate.

#### **3.5.1 Overall Survival**

For overall survival (OS) at the time of data cutoff, there were 30 death events in the P+L arm and 31 death events in the L arm. Table 8 displays results of log-rank test for the overall survival endpoint. As shown, there is a longer survival time (by 4 months) in the P+L arm. These results, however, are not fully mature. Figure 6 provides a Kaplan-Meier plot of the overall survival endpoint. No alpha was spent on the overall survival analysis, so no inference can be made.

*Reviewer Comment*

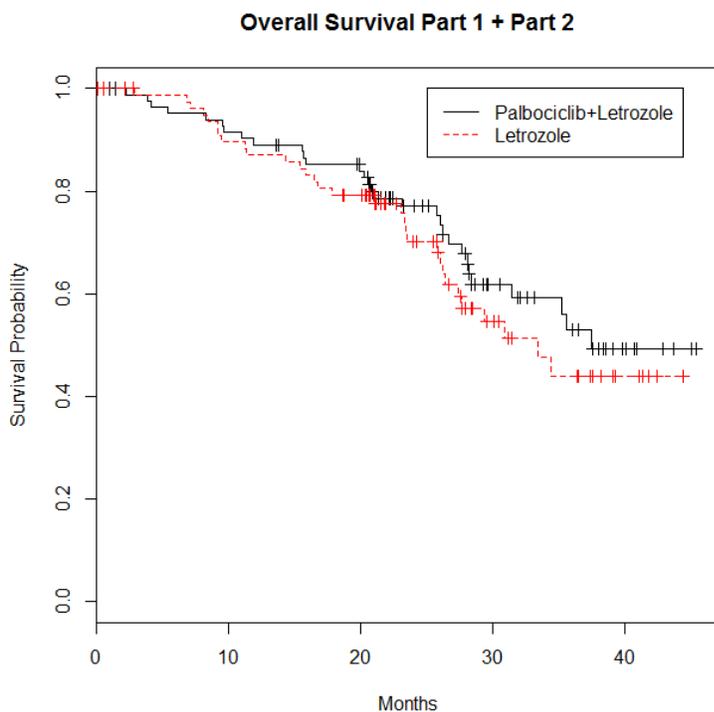
- *The overall survival data is not completely mature.*

**Table 10: Overall Survival**

	Part 1		Part 2		Part 1 + Part2	
	P+L N=34	L N=32	P+L N=50	L N=49	P+L N=84	L N=81
Number of event	16 (47.1%)	15 (46.9%)	14 (28.0%)	16 (32.7%)	30 (35.7)	31 (38.3)
Censored	18 (52.9%)	17 (53.1%)	36 (72.0%)	33 (67.3%)	54 (64.3%)	50 (61.7%)
Median PFS (months)	37.5	33.3	NR	NR	37.5	33.3
95% CI	(27.6, NR)	(26.0, NR)	(26, NR)	(23.4, NR)	(28.4, NR)	(26.4, NR)
Hazard Ratio	0.844		0.783		0.813	
Nominal p-value	0.32		0.25		0.21	

Source: Reviewer's analysis

**Figure 6: Overall Survival**



Source: Reviewer's analysis

### 3.5.2 Objective Response Rate

The objective response rate (complete response + partial response), as measured by the investigator, was 42.9% in the P+L arm [95% CI = (23.2, 44.7)] and 33.3% in the L arm [95% CI = (9.9%, 65.1%)]. The ORR in Part 1 was 44.1% in the P+L arm and 25.0% in the L arm. The ORR in Part 2 was 42.0% in the P+L arm and 38.0% in the L arm.

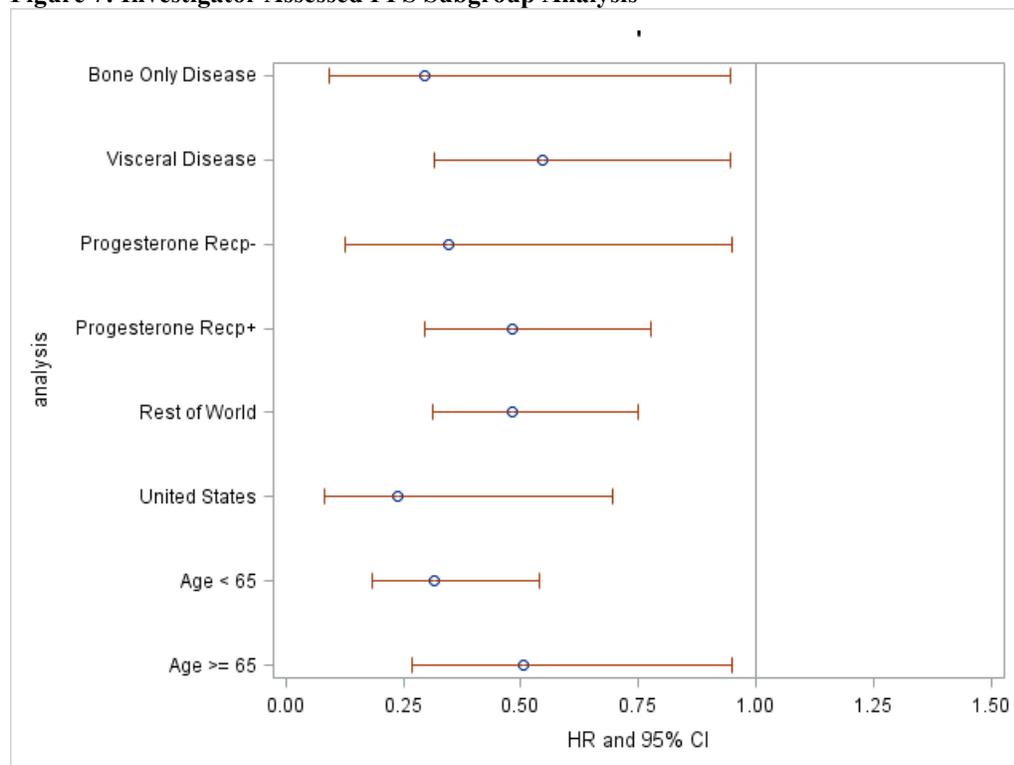
## 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

This section focuses on the efficacy results for investigator-assessed PFS by age, region, disease site (based upon the CRF), progesterone receptor status, and disease free interval (based upon the CRF). This section also focuses on sites where financial conflict of interests occurred, and on 1 site where OSI found major protocol violations. Finally, this section assesses the PFS results by CCDN1 and p16 biomarker status.

### 4.1 Demographic and Baseline Disease Characteristics

Results for investigator-assessed PFS for specific subgroups can be found in Figure 7. The results appear consistent across all subgroups analyzed. Results are not presented by gender since the study enrolled only females. Results are also not presented for race since 90% of participants were white.

**Figure 7: Investigator Assessed PFS Subgroup Analysis**



Source: Reviewer's analysis

## 4.2 Protocol Violations and Financial Conflicts of Interest

In the PALOMA-1 study, over 90% of study participants had at least one protocol violation. In addition, FDA's Office of Scientific Institutions (OSI) reported major study violations at one study site (Site 1001). To help assess whether the results at this site affected the analysis results, the primary analysis of investigator-assessed PFS was run again excluding this site (HR=0.463 95% CI = (0.302, 0.709) The updated analysis demonstrates that there does not appear to be a major bias by including this site.

In addition to OSI's findings, there were six sites with financial conflicts of interest. When removing these sites from the primary analysis, there was once again little effect on the overall conclusions (HR = 0.498 95% CI = (0.306, 0.812)).

### Reviewer Comment

- *The above sensitivity analysis suggests little bias occurs by leaving all study sites in the primary analysis.*

## 4.3 Biomarker Selection

As discussed in the study design section, PALOMA-1 was divided into two parts. Part 1 was an all-comers population that was stopped due to pre-clinical data. Part 2 was a biomarker selected population (CCND1 > 1.5 or P16 < 0.8) that the sponsor believed would have added benefit. Only later was Part 1 and Part 2 combined to form the full primary analysis population.

Using all participants in Part 2 and the biomarker positive participants in Part 1, an analysis of investigator assessed PFS finds a hazard ratio equal to 0.496, 95% CI = (0.299,0.824). Using only the biomarker negative participants in Part 1 (P+L arm had 22 patients with 10 PFS events, L arm 23 patients with 17 PFS events), an analysis of investigator assessed PFS finds a hazard ratio equal to 0.385, 95% CI = (0.153, 0.968). These results help to suggest little difference between the biomarker selected and unselected populations.

Delving into the biomarkers themselves, there appeared to be little difference between persons with CCND1 levels above 1.5 and below 1.5. For those with CCND1 levels above 1.5 (n=105), the HR = 0.408 95% CI = (0.242, 0.686) with a 11.5 month increase in median PFS. But for those with p16 levels below 0.8 (n=54), the HR equaled 0.326 95% CI = (0.150, 0.706) with a 11.4 month increase in PFS.

But for the p16 biomarker, there appeared to be a strong difference between those above 0.8 and those below 0.8. For those with p16 levels above 0.8 (n=112), the HR = 0.315 95% CI = (0.189, 0.523) with a 17.4 month increase in median PFS. But for those with p16 levels below 0.8 (n=33), the HR equaled 0.731 95% CI = (0.320, 1.667) with only a 1.5 month increase in PFS.

### Reviewer Comment

- *All the results for P16 are only exploratory results. But due to the large differences in magnitude between the two groups, FDA has requested a PMC to assess P16 and its relationship to Palbociclib in the confirmatory study.*

## **5. SUMMARY AND CONCLUSIONS**

### **5.1 Statistical Issues and Collective Evidence**

At this time, it appears that Palbociclib + letrozole treatment has a longer PFS time than letrozole treatment. Based upon the primary analysis of investigator-assessed PFS, the Palbociclib + letrozole has an estimated median PFS time of 20.2 months and the letrozole arm has a median PFS time of 10.2 months. Using the BICR assessment of PFS, the Palbociclib + letrozole has an estimated median PFS time of 25.7 months and the letrozole arm has a median PFS time of 14.8 months.

Nevertheless, due to poor study conduct, numerous protocol violations, data driven changes to the protocol, possible investigator bias towards the treatment arm, and a biomarker selected population in Part 2 of the study, the magnitude the difference in median PFS time remains uncertain at this time.

### **5.2 Conclusions and Recommendations**

The final decision on the benefit-risk evaluation of palbociclib plus letrozole treatment for the patient population studied is deferred to the clinical review team. The p-values have little meaning due to changes in the protocol, (b) (4)

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

ERIK W BLOOMQUIST  
01/15/2015

SHENGHUI TANG  
01/15/2015

RAJESHWARI SRIDHARA  
01/15/2015

## Statistical Review and Evaluation

**NDA #:** 207103                      **Sn #:** 19                      **Date Received:** 9/23/2014  
**Drug Name:** Palbociclib (Ibrance)                      **Indication:** mBC  
**Statistical Reviewer:** Erik Bloomquist  
**Medical Reviewer:** Beaver and Amiri  
**Sponsor:** Pfizer

**Special Protocol Assessment:** No

---

### Trial Specification:

**Trial Phase:** III    **Multicenter:** Yes  
**Blinding:** Double    **Control:** Placebo  
**Randomized:** Yes

**Treatment Arms:** 2  
**Treatment Schedule:** Palbociclib + Letrozole vs. Placebo + Letrozole

**Type of Hypothesis to be tested:** Superiority

**Primary Endpoint:** PFS

**Statistic:** Log-rank ; **Difference to be detected:** (b) (4) months

**Sample Size, N =** (b) (4) ;  $\alpha =$  ; One-sided / Two-sided;  $1 - \beta =$

**Interim Analyses:** Yes; **If yes, # of interim analyses =** 1 ; **Boundary Shape:** (b) (4)

**Primary Statistical Analyses:** Log-rank

### Summary:

After discussion with FDA on September 5, 2014, Pfizer submitted this written update to their statistical analysis plan for Study 1008. Study 1008 is the confirmatory study for palbociclib in post-menopausal, ER+, Her2-, metastatic breast cancer patients in the first-line setting. The study compares the add-on effect of palbociclib + letrozole versus letrozole + placebo.

Under the original study design, (b) (4) patients were to be enrolled using a 2:1 randomization schema. A total of (b) (4) PFS events would be required in the final analysis to have (b) (4)% power to detect a hazard ratio of (b) (4), or a (b) (4)% improvement in median PFS from (b) (4) months for the letrozole arm to (b) (4) months for the palbociclib plus letrozole arm (i.e., a (b) (4) month difference in median PFS). An interim analysis was to be performed after approximately (b) (4) patients had documented progressive disease or had died (approximately (b) (4)% of the total events expected).

After the study started, Pfizer changed the dosing information to include food and exclude proton pump inhibitors. After discussions with FDA, Pfizer agreed to (b) (4) the total sample size

from (b) (4) patients to (b) (4) so that a total of (b) (4) would be able to be enrolled under the new dosing strategy. The primary analysis would (b) (4) patients ((b) (4) PFS events), with a pre-planned secondary analysis in those not fasting or taking proton pump inhibitors.

Under this study with (b) (4) sample size, Pfizer planned an interim efficacy analysis when (b) (4) PFS ((b) (4)%) events had occurred. The sponsor proposed to use an (b) (4) when a HR of (b) (4) occurred at the interim analysis.

Based upon feedback from FDA, Pfizer has proposed to make this interim (b) (4) (b) (4) (b) (4) These calculations appear correct. Note that if Pfizer wished to achieve a HR equal to (b) (4) the pre-planned interim analysis would need the p-value to be less than (b) (4).

At the conclusion of their written document, Pfizer proposed that if the HR at the interim analysis exceeded (b) (4) months, this would be evidence of statistical and clinical significance.

#### **Comment to Sponsor**

The Agency agrees with the proposed amendment to the SAP and would like to meet with Pfizer when the results of the interim PFS analysis are available prior to stopping the study. Please let us know what your plans are for demonstrating efficacy in the proposed prescribed population (fed condition with no PPI) at the time of the interim analysis.

Erik Bloomquist, PhD  
Mathematical Statistician  
Date: 11/3/2014

Concur: Dr. Shenghui Tang

Cc:  
HFD-710/ Dr. Sridhara  
HFD-700/ Ms. Patrician

This review consists of 2 pages of text

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

ERIK W BLOOMQUIST  
11/05/2014

SHENGHUI TANG  
11/05/2014

NDA/BLA Number:  
NDA 207103

Applicant: Pfizer, Inc

Stamp Date: 6/30/2014

Drug Name: Pablociclib      NDA/BLA Type: Priority

On initial overview of the NDA/BLA application for RTF:

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	√			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	√			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	√			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	√			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? \_\_\_√\_\_\_

If the NDA/BLA is not fileable from the statistical perspective, state the reasons and provide comments to be sent to the Applicant.

NA

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	√			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	√			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			√	No interim analysis.
Appropriate references for novel statistical methodology (if present) are included.			√	The analysis methods used in the analysis are not novel.
Safety data organized to permit analyses across clinical trials in the NDA/BLA.	√			
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	√			

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

ERIK W BLOOMQUIST  
09/15/2014

SHENGHUI TANG  
09/15/2014