GEICAM

spanish breast cancer group

Statistical Analysis Plan (SAP)

GEICAM/2013-02

Phase III study of Palbociclib (PD-0332991) in combination with Endocrine therapy (exemestane or fulvestrant) versus chemotherapy (capecitabine) in Hormonal Receptor (HR) positive/HER2 negative Metastatic Breast Cancer (MBC) patients with Resistance to non-steroidal Aromatase inhibitors

"The PEARL study"

Sponsor: GEICAM (Spanish Breast Cancer Research Group Foundation)

Coordinating investigator: Dr. Miguel Martín. Hospital General Universitario Gregorio Marañón, Madrid, Spain

Author: Maribel Casas

Version: 2.0

Date: 31 March 2017

Protocol No.: GEICAM/2013-02

Protocol title: Phase III study of Palbociclib (PD-0332991) in combination with Endocrine therapy (exemestane or fulvestrant) versus chemotherapy (capecitabine) in Hormonal Receptor (HR) positive/HER2 negative Metastatic Breast Cancer (MBC) patients with Resistance to non-steroidal Aromatase inhibitors.

Protocol Version: Version 5.0, 09 March 2016

Sponsor: GEICAM

This Statistical Analysis Plan was created according the ICH Good Clinical Practice (1) (2), GEICAM policies and Standard Operating Procedures (SOP); and is consistent with the study protocol.

Version Date: 31-mar-2017

Version No.: 2.0

Previous Version: 1.0

Prepared by Statistician:

Name	Signature	Date
Maribel Casas	Maribel Casas	18-APR-2017

Reviewed by:

Statistician

Name	Signature	Date
JUAN JOSE MIRALLES		18-APR-2017

Project Manager

Name	Signature	Date
GRACIELA RODRIGUEZ	Graciela Rodriguez	18/APR/2017

Approved by:

Scientific Director

Name	Signature	Date
EVA CARRASCO		19-4-2017

Medical Coordinator

Name	Signature	Date
M MARTIN		18/4/17

TABLE OF CONTENTS

ABBREVIATIONS AND DEFINITIONS

AE	Adverse Event
AI	Aromatase Inhibitor
ALT/ALAT (SGPT)	Alanine Aminotransferase
ANC	Absolute Neutrophil Count
ASCO	American Society of Clinical Oncology
AST/ASAT (SGOT)	Aspartate Aminotransferase
CB	Clinical Benefit
CBR	Clinical Benefit Response
CDK	Cyclin-Dependent Kinase
cDNA	Circulating free DNA
CI	Confidence Interval
CISH	Chromogenic In Situ Hybridization
CMH	Cochran-Mantel-Haenszel test
CNS	Central Nervous System
CR	Complete Response
CT Scan	Computed Tomography Scan
CTCAE	Common Terminology Criteria for Adverse Events
DNA	Deoxyribonucleic Acid
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
ER	Estrogen Receptor
ESR1	Estrogen Receptor 1
FISH	Fluorescence In Situ Hybridization
FSH	Follicle-Stimulating Hormone
GCP	Good Clinical Practice
GEICAM	Spanish Breast Cancer Research Group

HER2	Human Epidermal Growth Factor Receptor 2
HR	Hormonal Receptor / Hazard Ratio
IDMC	Independent Data Monitoring Committee
IHC	Immunohistochemistry
IRB	Independent Review Board
ISH	In Situ Hybridization
ITT	Intent To Treat
MBC	Metastatic Breast Cancer
MID	Minimally important difference
NCI	National Cancer Institute
NSAI	Non-Steroidal Aromatase Inhibitor
OR	Objective Response
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease or Pharmacodynamic depending on the context
PFS	Progression-Free Survival
PgR	Progesterone Receptor
PK	Pharmacokinetic
PP	Per Protocol
PR	Partial Response
QOL	Quality of life
RD	Response Duration
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic Acid
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SC	Steering Committee
SD	Stable Disease

SISH	Silver In Situ Hybridization
SOP	Standard Operating Procedure
TdP	Torsade de Pointes
ULN	Upper Limit of Normal

1. INTRODUCTION

The aim of this Statistical Analysis Plan is to provide a detailed description of the planned statistical analyses to be performed for the GEICAM/2013-02 study (PEARL) (3).

The initially proposed Phase III study was designed to provide the opportunity to confirm the clinical benefit of palbociclib in combination with exemestane. The study was designed to demonstrate that this combination provides superior clinical benefit compared to capecitabine in postmenopausal women with HR+/HER2- metastatic breast cancer who are resistant to non-steroidal aromatase inhibitors.

Emerging data suggest that the choice of endocrine partner with palbociclib is particularly important to create optimal synergy in endocrine resistance setting. Studies of resistance to hormonal therapies and ER biology have highlighted the important role of ER receptor degraders for sequential endocrine treatment after patients progressed on aromatase inhibitors (AIs). Recent studies revealed that the acquisition of ESR1 mutations is a major mechanism of resistance to AIs. ESR1 mutations were found to be associated with exposure to AIs during the adjuvant and metastatic settings. Patients with ESR1 mutations had a substantially shorter PFS on subsequent AI-based therapy. In addition, the high incidence rate of ESR1 mutation has been reported among patients who have been treated with AIs in MBC patients (BOLERO-2, PALOMA-3). More recently, data from PALOMA-3 and other studies suggest that fulvestrant may be active among patients whose tumors had ESR1 mutation, while exemestane may be inactive. Thus among patients who have progressed on AIs, fulvestrant may be a better endocrine partner for palbociclib.

The present design provides the opportunity to confirm the clinical benefit of palbociclib in combination with endocrine therapy in relation to ESR1 mutational status. The primary study objectives are to demonstrate that: 1) fulvestrant plus palbociclib provides superior clinical benefit compared to capecitabine in postmenopausal women with HR+/HER2- metastatic breast cancer who are resistant to non-steroidal aromatase inhibitors and that 2) endocrine therapy (fulvestrant or exemestane) in combination with palbociclib provides superior clinical benefit compared to capecitabine in postmenopausal women with HR+/HER2- metastatic breast cancer and whose tumor had estrogen receptor (ESR1) mutational status as wild type at study entry.

2. STUDY OBJECTIVES

2.1 Primary Objectives

- To demonstrate that palbociclib in combination with fulvestrant is superior to capecitabine in prolonging Progression-Free Survival (PFS) in postmenopausal women with ER positive/HER2 negative MBC whose disease was resistant to non-steroidal aromatase inhibitors, regardless of the *ESR1* mutational status.
- To demonstrate that palbociclib in combination with endocrine therapy (exemestane or fulvestrant) is superior to capecitabine in prolonging PFS in postmenopausal women with ER positive/HER2 negative MBC whose disease was resistant to non-steroidal aromatase inhibitors and whose tumor had estrogen receptor *(ESR1)* mutational status as wild type at study entry.

  ESR1 mutational status will be determined by circulating free DNA (cDNA) and will be prospectively determined before the interims or final analyses. Patient tumor *ESR1* mutational status will be blinded to patients, investigators and study team.

2.2 Secondary Objectives

- To demonstrate that palbociclib in combination with endocrine therapy (exemestane or fulvestrant) is superior to capecitabine in prolonging PFS in postmenopausal women with ER positive/HER2 negative MBC whose disease was resistant to non-steroidal aromatase inhibitors regardless of *ESR1* mutational status at study entry.
- To demonstrate that palbociclib in combination with fulvestrant is superior to capecitabine in prolonging Overall Survival (OS) in postmenopausal women with ER positive/HER2 negative MBC whose disease was resistant to non-steroidal aromatase inhibitors.
- To demonstrate that palbociclib in combination with endocrine therapy (exemestane or fulvestrant) is superior to capecitabine in prolonging OS in postmenopausal women with ER positive/HER2 negative MBC whose disease was resistant to non-steroidal aromatase inhibitors and whose tumor had estrogen receptor *(ESR1)* mutational status as wild type at study entry.

- To demonstrate that palbociclib in combination with endocrine therapy (exemestane or fulvestrant) is superior to capecitabine in prolonging OS in postmenopausal women with ER positive/HER2 negative MBC whose disease was resistant to non-steroidal aromatase inhibitors regardless of *ESR1* mutational status at study entry.
- To compare other efficacy measures between the treatment arms: Objective Response Rate (ORR), Clinical Benefit Rate (CBR) and Response Duration (RD).
- To compare safety and tolerability between the treatment arms.
- To compare health-related quality of life between the treatment arms.
- To evaluate the Pharmacokinetics (PK) of the combination of exemestane with palbociclib (in selected sites and only in patients accepting to participate).

2.3 Exploratory Objectives

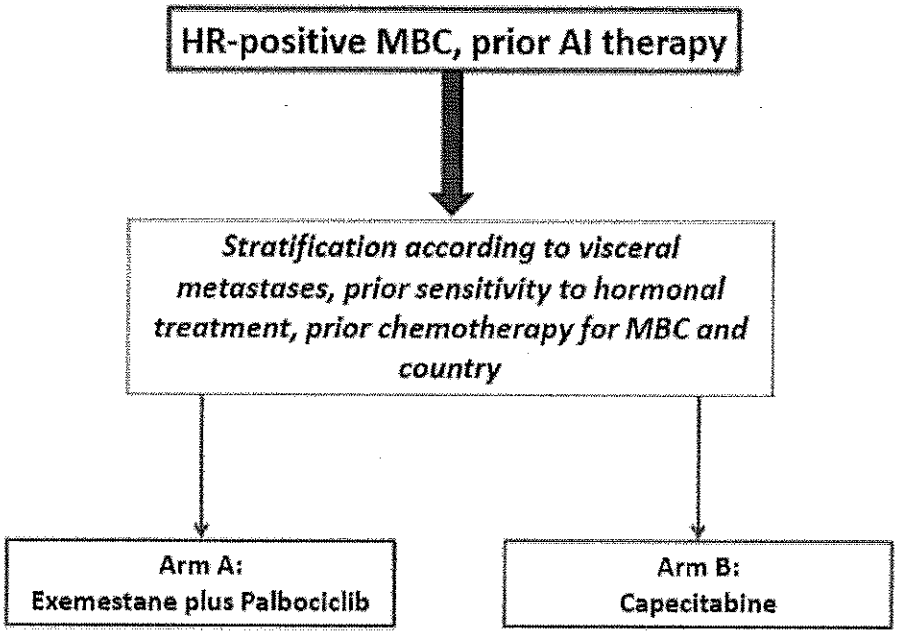
- To prospectively evaluate whether the magnitude of PFS and OS prolongation of palbociclib in combination with fulvestrant is the same in patients with *ESR1* mutational status as wild type and patients with *ESR1* mutational status positive at study entry.
- To prospectively evaluate whether PFS and OS of palbociclib in combination exemestane is better among those patients with *ESR1* mutational status as wild type than those patients with *ESR1* mutational status positive at study entry.
- To prospectively evaluate PFS and OS of capecitabine by *ESR1* mutational status.
- 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), tumor sensitivity and/or resistance (e.g. Ki67, pRb, PIK3CA mutation, CCNE1 expression) or breast cancer (e.g. PTEN, *ERBB2, BRCA 1 and BRCA2).*

3. STUDY DESIGN

This is an international, multicenter, open label, controlled, randomized phase III study comparing the efficacy and safety of palbociclib in combination with endocrine therapy (exemestane or fulvestrant) versus capecitabine in postmenopausal women with HR positive/HER2 negative MBC whose disease was resistant to previous non-steroidal aromatase inhibitors (letrozole or

anastrozole), defined as recurrence while on or within 12 months after the end of adjuvant treatment or progression while on or within 1 month after the end of treatment for advanced disease. It is not mandatory to have letrozole or anastrozole as the most recent treatment before randomization but recurrence or progression while receiving (or immediately after the end of) the most recent systemic therapy has to be documented before randomization. Patients must have measurable disease or at least one bone lesion, lytic or mixed (lytic + blastic), which has not been previously irradiated and is assessable by CT/MRI in the absence of measurable disease.

Figure 1. Study Design Cohort 1



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

- ✓ Palbociclib, 125 mg, orally once daily on Day 1 to Day 21 followed by 7 days off treatment on every 28 days cycles;

in combination with

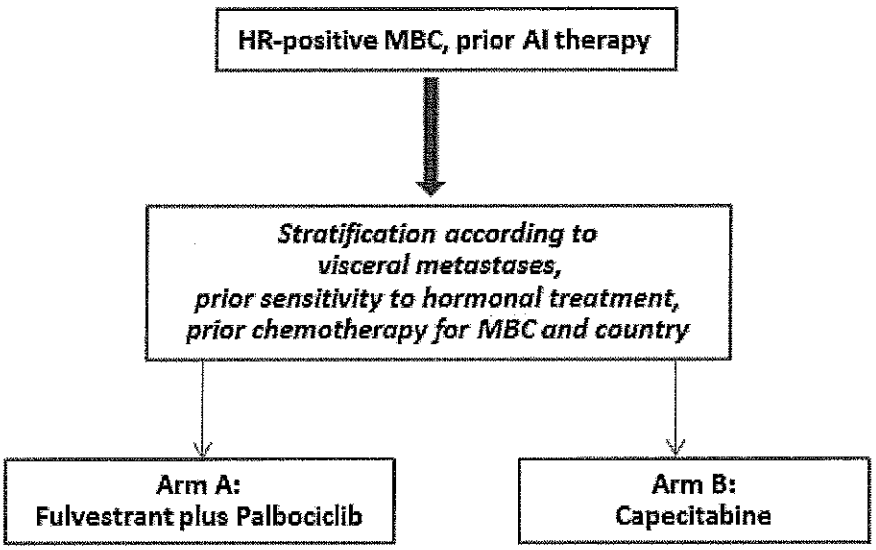
- ✓ Exemestane, 25 mg, orally once daily (continuously).

NOTE: In patients within the PK sub-study, exemestane will be administered daily during a 7-day lead-in period (day minus-7 through day minus-1) immediately preceding cycle 1.

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

- ✓ Capecitabine, 1,250mg/m^2 twice daily for 2 weeks followed by a 1 week rest period, given as 3 weeks cycles. Capecitabine must be administered at a dose of 1,000mg/m^2 twice daily for 2 weeks followed by a 1 week of rest period, given as 3 weeks cycles, in patients over 70 years of age.

Figure 2. Study Design Cohort 2



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

- ✓ Palbociclib, 125 mg, orally once daily on Day 1 to Day 21 followed by 7 days off treatment given as every 28 days cycles;

in combination with

- ✓ Fulvestrant 500 mg, two 5ml intramuscular injections (one in each buttock) on Days 1 and 15 (±3 days) of Cycle 1, and then on Day 1 of each subsequent 28 days Cycle (± 3 days).

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

- ✓ Capecitabine, 1,250mg/m^2 twice daily for 2 weeks followed by a 1 week rest period, given as 3 weeks cycles. Capecitabine must be administered at a dose of 1,000mg/m^2 twice daily for 2 weeks followed by a 1 week of rest period, given as 3 weeks cycles, in patients over 70 years of age.

In both Cohorts patients will continue to receive their assigned treatment until objective disease progression, symptomatic deterioration, unacceptable toxicity, death or withdrawal of consent, whichever occurs first. Patients with permanent discontinuation of palbociclib for a reason different than progression (i.e., toxicity due to palbociclib) will continue with exemestane/fulvestrant (in the active treatment phase) until objective disease progression, clinical progression (under investigator criteria), unacceptable toxicity, death or withdrawal of consent, whichever occurs first. Patients with permanent discontinuation of exemestane/fulvestrant will be discontinued from the active treatment phase and entered into the follow-up phase.

3.1 Sample Size

The primary objectives of this study are to demonstrate that the combination of palbociclib and fulvestrant is superior to capecitabine in prolonging PFS in postmenopausal women with HR+/HER2- metastatic breast cancer, whose tumors are resistant to prior non-steroidal aromatase inhibitors (Cohort 2) and to demonstrate that palbociclib in combination with endocrine therapy (exemestane or fulvestrant) is superior to capecitabine in prolonging PFS in postmenopausal women with ER positive/HER2 negative MBC whose disease was resistant to non-steroidal aromatase inhibitors and whose tumor had estrogen receptor (*ESR1*) mutational status as wild type at study entry (Cohort 1 + Cohort 2). All patients from Cohort 2 and patients with *ESR1* mutational status as wild type from Cohort 1 + Cohort 2 are eligible for the two primary inferential assessments of efficacy.

A modification of Hochberg's method (6) will be used for two primary treatment comparisons to provide control of experiment-wise Type 1 error rate at a 5% significance level. With this closed testing method, statistical significance applies to both p_1 for Cohort 2 and p_2 for Cohort 1+ Cohort 2 (ESR1 wild type) if $p_1<0.05$ and $p_2<0.05$; or it applies only to Cohort 2 if $p_1\leq0.025$ when Cohort 1+ Cohort 2 (ESR1 wild type) has $p_2>0.05$; or it applies to only Cohort 1+ Cohort 2 (ESR1 wild type) if $p_2\leq0.025$ when Cohort 2 has $p_1>0.05$ (assuming the analyses results represented by both p_1 for

Cohort 2 and p_2 for Cohort 1+ Cohort 2 (ESR1 wild type) demonstrating the combination of palbociclib and fulvestrant is superior to capecitabine in prolonging PFS).

Cohort 2 will have an 80% power to detect a difference between the control arm with a median PFS of 6 months and the experimental arm (palbociclib plus fulvestrant) with a median PFS of 9 months, for a hazard ratio of 0.667, with a 5% significance level. Assuming a non-uniform accrual accomplished over a period of about 21 months, and a follow-up period for the final PFS analysis of about 28 months from the start of study randomization of Cohort 2, a total sample size of approximately 300 patients (150 in each treatment arm) will be required and the necessary number of events for the final PFS analysis is determined to be 193.

Cohort 1 + Cohort 2 (*ESR1* wild type) should have an approximately 80% power to detect a difference between the control arm with a median PFS of 6 months and the experimental arm (palbociclib plus fulvestrant or exemestane) with a median PFS of 9 months, for a hazard ratio of 0.667, with a 5% significance level in *ESR1* mutation status as wild type patients. The sample size will be similar to Cohort 2. Approximately 308 patients and 193 PFS events will be accumulated if we assume 80% cDNA collection/detect rate and 70% of patients will have tumors with *ESR1* mutational status as wild type at study entry (approximately 140 mutational status as wild type from 250 patients from Cohort 1 and 168 mutational status as wild type from 300 patients from Cohort 2).

The sample size described above will also allow the assessment of differences in the secondary endpoint of overall survival (OS). The median OS for women with advanced or metastatic breast cancer treated with capecitabine is assumed to be 22 months. With an overall significance level of 10% and one interim analysis of OS (at the time of PFS final analysis), Cohort 2 will have approximately 80% power to detect a HR of 0.667 (representing a 50% increase in median OS from 22 months to 33 months) when approximately 152 deaths have occurred after an approximate follow-up of 50 months from the start of study randomization. Similar sample size determination also applies to Cohort 1 + Cohort 2 (*ESR1* wild type).

3.2 Randomization

- Approximately 250 patients will be randomized 1:1 between the experimental arm (Arm A: approximately 125 patients treated with palbociclib plus exemestane) and the control arm

(Arm B: approximately 125 patients treated with capecitabine) before the approval of the protocol version 5.0, 09 March 2016 (Cohort 1).

- Approximately 300 patients will be randomized 1:1 between the experimental arm (Arm A: approximately 150 patients treated with palbociclib plus fulvestrant) and the control arm (Arm B: approximately 150 patients treated with capecitabine) from the approval of the protocol version 5.0, 09 March 2016 (Cohort 2).

Patients will be stratified by:

- Site of disease: visceral vs non-visceral (visceral are all lesions not included in the following list: breast, skin, subcutaneous tissue, lymph node or bone).
- Prior sensitivity to hormonal treatment:
    - Yes defined as:
        - At least 24 months of endocrine therapy before recurrence in the adjuvant setting in patients who have not received previous endocrine therapy in the metastatic setting.
        - Response or stabilization for at least 24 weeks of the most recent line of endocrine therapy in patients who have received previous endocrine therapy in the metastatic setting.
    - No: all other options
- Prior chemotherapy for MBC: yes (chemotherapy administered as "second adjuvant therapy" for locoregional recurrence should be considered as prior chemotherapy for MBC) vs no.
- Country.

The randomization will be performed using stratified blocks of treatment. The size of each block of treatment will be 4.

4. STUDY POPULATION

4.1 Inclusion Criteria

Patients are eligible to be included in the study only if they **meet all** of the following criteria:

1. The patient has signed the informed consent document.

2. Females with histologically confirmed MBC whose disease was resistant to previous non-steroidal aromatase inhibitors (NSAI) (letrozole or anastrozole), defined as:
    - ✓ Recurrence while on or within 12 months after the end of adjuvant treatment with a NSAI or
    - ✓ Progression while on or within 1 month after the end of treatment with NSAI for advanced disease.

Note: Administrative letter n.2 dated 11Oct2016 confirmed a change in this inclusion criteria, in the countries where this document has been accepted by IRB and MoH, the inclusion criteria#2 is: ...

Females with histologically confirmed MBC whose disease was resistant to previous aromatase inhibitors (AI) (letrozole, anastrozole or exemestane), defined as:

- ✓ *Recurrence while on or within 12 months after the end of adjuvant treatment with a AI or*
- ✓ *Progression while on or within 1 month after the end of treatment with AI for advanced disease.*

3. Previous chemotherapy is permitted either in the (neo) adjuvant setting and/or first line therapy for MBC (chemotherapy administered as "second adjuvant therapy" for locoregional recurrence should be considered as first line chemotherapy for MBC).
4. It is not mandatory to have letrozole or anastrozole as the most recent treatment before randomization but recurrence or progression of breast cancer while receiving (or immediately after the end of) the most recent systemic therapy has to be documented before randomization.
5. Hormonal receptor positive (HR+) breast cancer based on local laboratory determination. HR+ defined as ≥ 1% positive cells by IHC for ER and/or PgR.
6. Documented HER2 negative breast cancer based on local laboratory determination on most recent tumor biopsy. HER2 negative tumor is determined according to recommendations of ASCO/CAP 2013 guidelines, as IHC score 0 or 1+ or negative by ISH (FISH/CISH/SISH) defined as a HER2/CEP17 ratio < 2 with an average HER2 copy number <4.0 or for single probe assessment a HER2 copy number < 4.
7. Measurable disease or at least one bone lesion, lytic or mixed (lytic+blastic), which has not been previously irradiated and is assessable by CT/MRI in the absence of measurable disease according to RECIST 1.1 criteria (4).

8. Patient is at least 18 years of age.
9. Eastern Cooperative Oncology Group (ECOG) Performance Status ≤ 1.
10. Life expectancy ≥ 12 weeks.
11. Adequate organ and marrow function defined as follows:
    - ✓ ANC ≥ 1,500/mm3 ($1.5x10^9$/L)
    - ✓ Platelets ≥ 100,000/mm3 ($100x10^9$/L)
    - ✓ Hemoglobin ≥ 9g/dL (90g/L)
    - ✓ Serum creatinine ≤ 1,5xULN and estimated creatinine clearance ≥ 60 mL/min as calculated using the standard method for the institution
    - ✓ Total serum bilirubin ≤ 1,5xULN (≤ 3.0xULN if Gilbert´s disease)
    - ✓ AST and/or ALT ≤ 3.0xULN (≤5.0xULN if liver metastases present)
    - ✓ Alkaline phosphatase ≤ 2.5xULN (≤5.0xULN if bone or liver metastases present)
12. Postmenopausal women defined as women with:
    - ✓ Prior bilateral surgical oophorectomy, or
    - ✓ Age > 60 years, or
    - ✓ Age < 60 years and medically confirmed post-menopausal status defined as spontaneous cessation of regular menses for at least 12 consecutive months with no alternative pathological or physiological cause or follicle-stimulating hormone (FSH) and estradiol blood levels in their respective postmenopausal ranges.
13. Resolution of all acute toxic effects of prior anti-cancer therapy or surgical procedures to NCI CTCAE version 4.0 Grade ≤ 1 (except alopecia or other toxicities not considered a safety risk for the patient at investigator´s discretion).
14. Willingness and ability to comply with scheduled visits, treatment plan, laboratory tests and other study procedures.

4.2 Exclusion Criteria

Patients will be excluded from the study if they **meet any** of the following criteria:

1. Have received more than 1 prior chemotherapy regimen for MBC. (NOTE: Chemotherapy administered as "second adjuvant therapy" for locoregional recurrence should be considered one prior chemotherapy for MBC). Other previous anticancer endocrine treatments for advanced disease are allowed.
2. Patients with advanced, symptomatic, visceral spread that are at risk of life-threatening complications in the short term (including patients with massive uncontrolled effusions [pleural, pericardial, peritoneal], pulmonary lymphangitis and over 50% liver involvement).
3. Known active uncontrolled or symptomatic CNS metastases, carcinomatous meningitis or leptomeningeal disease as indicated by clinical symptoms, cerebral edema, and/or progressive growth. Patients with a history of CNS metastases or cord compression are eligible if they have been definitively treated with local therapy (eg, radiotherapy, stereotactic surgery) and are clinically stable off anticonvulsants and steroids for at least 4 weeks before randomization.
4. Prior treatment with any CDK4/6, mTOR or PI3K inhibitor [any agent whose mechanism of action is to inhibit the PI3 kinase-mTOR pathway] or capecitabine.
5. a) Patients included in Cohort 1: Prior treatment with exemestane in the metastatic setting. If the patient has received exemestane in the adjuvant setting and developed MBC, she will be eligible for the study provided:
   - ✓ She has received letrozole/anastrozole as first-line MBC and progressed.
   - ✓ At least 1 year has elapsed since the end of adjuvant exemestane treatment.

   b) Patients included in Cohort 2: Prior treatment with fulvestrant in the metastatic setting. If the patient has received fulvestrant in the adjuvant setting and developed MBC, she will be eligible for the study provided:
   - ✓ She has received letrozole/anastrozole as first-line MBC and progressed.
   - ✓ At least 1 year has elapsed since the end of adjuvant fulvestrant treatment.
6. Patients treated within the last 7 days prior to randomization with:

- ✓ Food or drugs that are known to be CYP3A4 inhibitors
- ✓ Drugs that are known to be CYP3A4 inducers
- ✓ Drugs that are known to prolong the QT interval

7. Patients who received before randomization:
   - ✓ Any investigational agent within 4 weeks.
   - ✓ Chemotherapy within a period of time that is < the cycle length used for that treatment (e.g. < 3 weeks for fluorouracil, doxorubicine, epirubicine or < 1 week for weekly chemotherapy).
   - ✓ Previous endocrine therapy is permitted without any window.
   - ✓ Radiotherapy within 2 weeks (all acute toxic effects must be resolved to NCI CTCAE version 4.0 grade <1, except toxicities not considered a safety risk for the patient at investigator´s discretion) but patients who received prior radiotherapy to >25% of bone marrow are not eligible independent of when it was received.
   - ✓ Major surgery or other anti-cancer therapy not previously specified within 4 weeks, (all acute toxic effects must be resolved to NCI CTCAE version 4.0 grade < 1, except toxicities not considered a safety risk for the patient at investigator´s discretion).
8. Diagnosis of any other malignancy within 3 years prior to randomization, except for adequately treated basal cell or squamous cell skin cancer, or carcinoma in situ of the cervix.
9. QTc > 480msec, family or personal history of long or short QT syndrome, Brugada syndrome or known history of QTc prolongation, or Torsade de Pointes (TdP).
10. Uncontrolled electrolyte disorders that can compound the effects of a QTc-prolonging drug (eg, hypocalcemia, hypokalemia, hypomagnesemia).
11. Any of the following within 6 months of randomization: myocardial infarction, severe/unstable angina, ongoing cardiac dysrhythmias of NCI CTCAE version 4.0 Grade ≥ 2, atrial fibrillation of any grade, coronary/peripheral artery bypass graft, symptomatic congestive heart failure, cerebrovascular accident including transient ischemic attack, or symptomatic pulmonary embolism.

12. Difficulties to swallow tablets, malabsorption syndrome disease significantly affecting gastrointestinal function, resection of the stomach or small bowel, or active inflammatory bowel disease or chronic diarrhea.
13. Known hypersensitivity to exemestane, palbociclib, capecitabine, fulvestrant or any of their excipients.
14. Any of the following contraindications for chemotherapy with capecitabine:
    - ✓ Known deficiency or family history of deficiency of dihydropyrimidine dehydrogenase.
    - ✓ Requirement for concurrent use of the antiviral agent sorivudine (antiviral) or chemically related analogues, such as brivudine.
15. Only for patients in Cohort 2 any of the following contraindications for treatment with fulvestrant:
    - ✓ Bleeding diathesis (i.e., disseminated intravascular coagulation [DIC], clotting factor deficiency) or long-term anticoagulant therapy (other than antiplatelet therapy and low dose warfarin).
16. Known human immunodeficiency virus infection.
17. Other severe acute or chronic medical or psychiatric condition or laboratory abnormality that may increase the risk associated with study participation or investigational product administration or may interfere with the interpretation of study results and, in the judgment of the investigator, would make the patient inappropriate for entry into this study.
18. Recent or active suicidal ideation or behavior.

4.3 Discontinuations

4.3.1 Discontinuation of Study Drugs/Medications

The criteria for enrollment must be followed explicitly. If a patient who does not meet enrollment criteria is inadvertently enrolled, that patient should be discontinued from the study drugs/medications, but can be allowed to continue in the study in order to provide the follow up data needed for the analysis of the entire population. An exception may be granted if the patient, in the opinion of the investigator, is having benefit from the study drugs/medications. In these rare

cases, the investigator must obtain documented approval from GEICAM to allow the patient to continue to receive the study drugs/medications.

Patients can be discontinued from the study therapy in the following circumstances:

- Patient's own request.
- Unacceptable toxicity as defined in the protocol.
- Tumor progression as defined in the protocol.
- Any clinical adverse event (AE), laboratory abnormality or intercurrent illness which, in the opinion of the investigator, indicates that continued participation in the study is not in the best interest of the patient.
- Termination of the study by GEICAM.
- Physician's decision, including need of other anti-cancer therapy, not specified in the protocol.
- If the patient is non-compliant with study procedures.
- Loss of ability to freely provide consent through imprisonment or involuntarily incarceration for treatment of either a psychiatric or physical (e.g. infectious disease) illness.
- All patients will be discontinued from the active treatment phase and entered into the follow-up phase in case of a delay of more than 4 weeks or permanent discontinuation of exemestane/fulvestrant unless there is an obvious clinical benefit per the investigator's medical judgment and after discussion with GEICAM. Patients with a delay of more than 4 weeks or permanent discontinuation of palbociclib for a reason different than progression (i.e., toxicity due to palbociclib) will continue with exemestane/fulvestrant (in the active treatment phase) until objective disease progression, clinical progression (under investigator criteria), unacceptable toxicity, death or withdrawal of consent, whichever occurs first.

If possible, and after the permanent discontinuation of treatment, the patients will be assessed using the procedure normally planned for the last dosing day with the study treatment.

All permanent treatment discontinuation should be recorded by the Investigator in the eCRF when considered as confirmed.

4.3.2 Discontinuation of Study Sites

Study Site participation may be discontinued if GEICAM, the investigator or the IRB of the study site judges it necessary for any reason.

4.3.3 Discontinuation of Study

The study may be discontinued by GEICAM if this is medically reasonable and consistent with applicable regulations of Good Clinical Practice (GCP). Stopping the study for medical reasons may be required if patients experienced adverse reactions under the treatment with the study drug/medication or if new information about the safety or effectiveness of the study drug/medication justifies it.

4.4 ITT Population

The Intent to treat population (ITT) will include all patients who are randomized, with study drug/medication assignment designated according to initial randomization.
The ITT population will be the primary population for evaluating patient characteristics and efficacy.

4.5 ESR1 wild type Population

The *ESR1* wild type population will include all patients who are randomized, with study drug/medication assignment designated according to initial randomization and whose tumor had estrogen receptor (*ESR1*) mutational status as wild type at study entry.
The *ESR1* wild type population will be the primary population for evaluating patient characteristics and efficacy.

4.6 Per-protocol population

Per-protocol population is a subset of the ITT population that received at least one dose of study medication and completed the study without any major protocol violations.

4.7 Safety population

Safety population will include all patients randomized in the study who received at least one dose of treatment, according to the actual treatment received. This population is for the safety analysis.

4.8 PK population

PK population is a subset of included patients in the experimental arm with available PK samples.

4.9 PD population

PD population is a subset of included patients with available PD samples.

5. ENDPOINTS AND STUDY VARIABLES

5.1 Efficacy Endpoints

5.1.1 Primary Endpoint

Progression Free Survival (PFS)

The primary endpoint is PFS which is defined as the time from randomization to the first documented progressive disease based on the investigator's assessment, using RECIST version 1.1 (4), or death from any cause, whichever occurs first. PFS data will be censored on the date of the last tumor assessment on study for patients who do not have objective tumor progression and who do not die while on study. Patients lacking an evaluation of tumor response after randomization will have their PFS time censored on the date of randomization with 1 day duration. Additionally, patients who start a new anti-cancer therapy prior to documented PD will be censored at the date of the last tumor assessment prior to the start of the new therapy.

5.1.2 Secondary Endpoints

Overall Survival (OS)

The overall survival is defined as the time from the date of randomization to the date of death from any cause. In the absence of confirmation of death, survival time will be censored to last date the patient is known to be alive.

Objective Response (OR)

The objective response is defined as a Complete Response (CR) plus Partial Response (PR) based on the investigator's assessment according to the RECIST version 1.1 (4) in patients randomized with measurable disease. Tumor assessment will be performed at baseline, the same

method of measurement used at baseline will be used for further evaluations, that will be conducted every 8 weeks ±7days. The best response across treatment will be recorded.

Clinical Benefit (CB)

Clinical benefit is defined as CR plus PR plus stable disease based on the investigator's assessment lasting more than 24 weeks according to the RECIST version 1.1 (4) in all randomized patients (ITT population).

Response Duration (RD)

Response duration is defined as the time from the first documentation of objective tumor response (CR or PR) to the first documented progressive disease using RECIST version 1.1 (4) and based on the investigator's assessment, or to death due to any cause, whichever occurs first. RD data will be censored on the date of the last tumor assessment on study for patients who do not have objective tumor progression and who do not die due to any cause while on study. Additionally, patients who start a new anti-cancer therapy prior to documented PD will be censored at the date of the last tumor assessment prior to the start of the new therapy. RD will only be calculated for the subgroup of patients with an objective response.

5.2 Safety Endpoints

Safety will be assessed by standard clinical and laboratory tests (hematology, serum chemistry). Adverse events grade will be defined by the NCI CTCAE v4.0 (5)

5.3 Other Variables

5.3.1 Pharmacokinetic Variables

The PK will determine whether palbociclib influences the pharmacology of exemestane. Blood samples will be taken at the time defined in the protocol in patients included in the experimental arm of Cohort 1 (palbociclib plus exemestane) in selected sites. Enrollment of these patients will continue until approximately 20 PK-evaluable patients have completed the planned PK collections in the experimental arm of Cohort 1.

5.3.2 Biomarker, Pharmacogenomic and Pharmacodynamic Variables

Baseline biomarker values from most recently obtained tumor tissue (deeply recommended from metastatic tumor) will be used for central assessment of biomarkers related to breast tumor sensitivity and/or resistance to palbociclib (e.g., Ki67, p16/CDKN2A, pRb, CyclinD and others) or breast cancer (e.g. PTEN, *ERBB2, BRCA 1 and BRCA2)*.. A whole blood sample will be collected for potential pharmacogenomic analyses related to drug response or adverse drug reactions; For example, putative safety biomarkers, drug metabolizing enzyme genes, drug transport protein genes, or genes thought to be related to the mechanism of drug action may be examined. Correlative plasma samples will be collected for exploratory analysis to analyze the pharmacodynamic (PD) treatment effects on circulating free DNA or RNA explore specific breast cancer and efficacy predictive biomarkers. Samples will be collected from all patients, unless prohibited by local regulations.

5.3.3 Quality of Life Variables

Patient reported outcomes of health-related quality of life will be assessed using the EORTC QLQ-C30 and breast modules (QLQ-BR23) instruments and the EuroQol Health Utilities Index EQ-5D-3L instrument.

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

The EORTC-QLQ-BR23 is a 23-item breast cancer-specific companion module to the EORTC-QLQ-C30 and consists of two functional scales (body image and sexuality); three symptom subscales (arm/hand, breast, and systemic side effects) and single items covering sexual enjoyment, distress at hair loss, and future perspective.

The EuroQol-5D (EQ-5D) (version 3L) is a 6 item instrument designed to assess health status in terms of a single index value or utility score. It consists of 5 descriptors of current health state (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression); a patient is asked to rate each state on a three level scale (1=no problem, 2=some problem, and 3=extreme problem) with higher levels indicating greater severity/impairment. It also includes a visual analogue scale: the EQ VAS. The EQ VAS records the patient's self-rated health on a scale from 0 (worst imaginable health state) to 100 (best imaginable health state). Published weights are available that allow for the creation of a single summary score. Overall scores range from 0 to 1, with low scores representing a higher level of dysfunction and 1 as perfect health.

6. DATA SCREENING AND ACCEPTANCE.

6.1 Missing data

The frequency of missing data will be examined and reported for each variable in the analysis. We will not perform data imputation for missing data.

6.1.1 Missing date

If the day of the month is missing for any date used in a calculation, the 15th 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 are missing for any date used in a calculation, the date will be considered missing.

6.1.2 Missing Tumor Assessments

For the evaluation of response, the RECIST 1.1 (4) criteria will be used.

If baseline tumor assessment is inadequate according to RECIST 1.1 criteria (4), the patient may not be assessed for certain endpoints like progression free survival, objective response, response duration or clinical benefit response.

Inadequate baseline assessment may include:

- If patient has measurable disease: measurements not provided for one or more target lesions according to RECIST 1.1 (4) criteria.
- If patient has only unmeasurable disease:

- Patients with lytic or mixed (lytic+ blastic) bone lesions can be included in the study and will not be assessed for objective response but for progression free survival and CBR. Patients without baseline measurements of bone disease cannot be assessed for any of these objectives.
- Patients without lytic or mixed (lytic+ blastic) bone lesions cannot be included in the study and will not be assessed for any of the aforementioned endpoints (only for survival).

After baseline, 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 CBR and OR, patients who do not have on study radiographic (or photographical where applicable) tumor reevaluations will be counted as non-responders (and reported as not evaluable).

6.1.3 Missing Data in PFS Derivation

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

If a substantial number of patients have questionable failure or censorship dates for the 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.

6.2 Statistical software

The statistical analysis will be developed in SAS Enterprise Guide v5.1.

6.3 Database lock

1) The data base lock for the first interim analysis is planned after 150 events (in Cohort 1) have occurred.
2) The data base lock for the second interim analysis is planned after 116 events (in Cohort 2) have occurred.
3) The database lock for final PFS analysis is planned after 193 events have occurred in Cohort 2 (at this time, it will be expected to achieve a similar number of events for Cohort 1 + Cohort 2 (ESR1 wild type)).

7. INTERIM ANALYSIS

The study is designed to have two interim analyses and the final analysis.

The **first interim analysis** will be performed after approximately 150 patients have documented progressive disease or death in Cohort 1, only patients randomized to Cohort 1 will be included. The enrollment of Cohort 2 will not be affected by the outcome of the first interim analysis.

The **second interim analysis** will be performed on Cohort 2 and Cohort 1 + Cohort 2 (*ESR1* wild type) after approximately 116 patients have documented progressive disease or death in Cohort 2 (approximately 60% of the total events expected for Cohort 2). The number of PFS events for Cohort 1 + Cohort 2 (*ESR1* wild type) cannot be pre-determined.

7.1 Purpose of interim analyses

The purposes for the **first interim analyses** are to assess the safety of the patients and to potentially re-evaluate the assumption of the proportion of patients with ESR1 wild type randomized to the study which may affect the sample size determination for Cohort 1 + Cohort 2 (*ESR1* wild type) analyses.
The purposes for the **second interim analyses** are to allow for early stopping of the study for efficacy, futility, to assess the safety of the combination regimens and to potentially re-estimate the sample size of the trial.

7.2 Interim analyses procedure

The significance level will be allocated to the second interim and final analyses with a Bonferroni method, such as alpha=0.002 for the second interim analysis, and alpha=0.048 for the final analysis.

The same Hochberg method (6) described for the final analysis will be used for the second interim analysis. The rho spending function family (7) will be used for the futility boundary (rho=4, non-binding) for PFS at the second interim analysis. Non-binding for futility implies that the futility boundary will be constructed in such a way that it can be overruled if desired by the IDMC without inflating the Type I error rate. The total Type I error rate will be well preserved. The sample size of the study may be adjusted at the second interim analysis using the method outlined by Cui et al. (8) and applied to the time-to-event endpoint. Using the Cui method guarantees that the overall type I

error will still be preserved after a sample size increase. To protect the study integrity and for confidential reasons, the details of sample size re-estimation plan and procedure will be written in a separate document. If the results of the interim analyses indicate serious safety concerns, the IDMC will alert the SC with the recommendation to stop the clinical trial for safety reasons.

Efficacy and Futility Boundaries for the Second Interim Analysis Expressed as Z Scales and p-values

	Number of Event (%)	**Z Scale**	**p-value**
Efficacy	116 (60%)	3.044	0.002
Futility (non-binding)	116 (60%)	0.328	0.743

The efficacy and futility boundaries provided in the above table are defined for Cohort 2 if the second interim analysis is performed with exactly 116 events. If the interim analysis is performed with not exactly 116 events, the efficacy boundary expressed as the p-value stays the same (p=0.002), but the Z scale can be slightly different depending on the number of the events at the analysis. The same principle is also applied to Cohort 1 + Cohort 2 (*ESR1* wild type) interim PFS analysis where p=0.002 is the efficacy boundary if the number of the events is more or less 116 at the analysis.

One interim analysis of OS will be performed at the time of PFS final analysis. Given the number of events for the OS interim analysis is depended on the time of the final PFS analysis and cannot be pre-determined, the Haybittle Peto efficacy boundary of p=0.002 is selected. The Z scale will be calculated based the result and the number of the events at the analysis.

7.3 IDMC (Interim Data Monitoring Committee)

The study will use an IDMC. The IDMC membership and governance is outlined in a separate charter.

The IDMC will be responsible for ongoing monitoring of the efficacy and safety data from patients in the study according to the Charter. The IDMC will make recommendation as to whether or not the trial should continue based on ongoing reviews of safety data. In addition, the IDMC will also evaluate the interim efficacy data and make a recommendation regarding study continuation based on observed results of the study.

In addition to the use of the IDMC, the safety of the trial will also be monitored by the SC. The recommendations made by the IDMC to alter the conduct of the study will be forwarded to the SC for final decision. GEICAM will designate a biostatistician not affiliated with the project to prepare data for IDMC review. Clinical sites will be restricted from access to study results until the conclusion of the study.

7.3.1 Early Safety Review

The safety of the combination of palbociclib plus exemestane at the recommended doses will be assessed with a limited number of patients exposed to it.

These data will be evaluated by the Independent Data Monitoring Committee (IDMC). The IDMC will make a recommendation whether to interrupt/amend the protocol mainly based on the proportion of grade 4 neutropenia, grade 4 thrombocytopenia or grade 3 QTc prolongations. The Steering Committee (SC) will take the final decision.

PK analysis of the palbociclib and exemestane combination will be performed once the collection of samples from patients included in the PK cohort is completed (approximately 20 evaluable patients accepting to participate and enrolled in the experimental arm of Cohort 1, palbociclib plus exemestane, in selected sites).

7.4 Criteria for End of Study

This study will be considered complete following the data cut-off date and datalock for the final analysis of OS. The data cut-off date for the final analysis will occur when approximately 50% of enrolled patients in Cohort 2 (approximately 152 patients) have died.
If further data are collected that are not included as part of the final locked database, the postlock data will eventually be combined with the locked database and stored in a data library separate from the locked database.

8. STATISTICAL METHODS AND ANALYSES

8.1 Statistical Methods

Descriptive Analysis

Frequencies, percentages and ninety-five percent confidence intervals of interest will be calculated for categorical variables wherever possible (9). For continuous variables, standard descriptive statistics, such as total number of observations, number of available data, mean, standard deviation, minimum, percentil 25, median, percentil 75 and maximum will be calculated.

Chi-square χ^2 Test

The first type of chi-square test is *the goodness of fit test*. This is a test which makes a statement or claim concerning the nature of the distribution for the whole population. The data in the sample is examined in order to see whether this distribution is consistent with the hypothesized distribution of the population or not. One way in which the chi-square goodness of fit test can be used is to examine how closely a sample matches a population. The chi-square goodness of fit test can be used to provide a test for the representativeness of a sample.

Suppose that a variable has a frequency distribution with *k* categories into which the data has been grouped. The frequencies of occurrence of the variable, for each category, are called the observed values. The manner in which the chi-square goodness of fit test works is to determine how many cases there would be in each category if the sample data were distributed exactly according to the claim. These are termed the expected number of cases for each category. The total of the expected number of cases is always made equal to the total of the observed number of cases. The null hypothesis is that the observed number of cases in each category is exactly equal to the expected number of cases in each category. The alternative hypothesis is that the observed and expected number of cases differs sufficiently to reject the null hypothesis.

Let O_i is the observed number of cases in category *i* and E_i is the expected number of cases in each category, for each of the *k* categories *i = 1, 2, 3…., k*, into which the data has been grouped. The hypotheses are

$$H_0 : O_i = E_i \qquad \forall\, i$$
$$H_1 : O_i \neq E_i \quad for\ some\ i$$

and the test statistic is

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where the summation proceeds across all *k* categories, and it has *k-1* degrees of freedom.

The second type of chi-square test is the chi-square *test for independence of two variables*. The chi-square test of independence allows researchers to determine whether variables are independent of each other or whether there is a pattern of dependence between them. If there is dependence, the researcher can claim that the two variables have a statistical relationship with each other.

The only limitation on the use of this test is that the sample sizes must be sufficiently large to ensure that the expected number of cases in each category is five or more, if not, the Fisher exact test must be used. This rule can be modified somewhat, but as with all approximations, larger sample sizes are preferable to smaller sample sizes. There are no other limitations on the use of the test, and the chi-square statistic can be used to test any contingency or cross classification table for independence of the two variables.

The chi-square test for independence is conducted by assuming that there is no relationship between the two variables being examined. The alternative hypothesis is that there is some relationship between the variables.

$$\mathrm{H}_0 : \mathrm{X\ and\ Y\ are\ independent}$$
$$\mathrm{H}_1 : \mathrm{X\ and\ Y\ \ are\ dependent}$$

Using that we can estimate the expected frequencies in each cell as $E_{ij} = np_ip_j$ where n is the total count, p_i is the proportion in row i and p_j is the proportion in row j. Then the chi-quared statistic is, for an $r \times c$ table:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

The observed numbers of cases,O_{ij}, are the numbers of cases in each cell of the cross classification table, representing the numbers of respondents for each combination of the variables. The expected numbers of cases E_i for each of the cell can be obtained from the multiplication rule of probability for independent events.

Fisher Exact Test

When the assumptions for the Chi-squared test are not met, i.e. one of the expected values in a 2x2 table is less than 5, and especially when it is less than 1, then Fisher's Exact test must be

applied. The null hypothesis for the test is that there is no association between the rows and columns of the 2x2 table, such that the probability of a subject being in a particular row is not influenced by being in a particular column. If the columns represent the study group and the rows the outcome, then the null hypothesis could be interpreted as the probability of having a particular outcome not being influenced by the study group, and the test evaluate whether the two study groups differ in the proportions with each outcome. An important assumption for Fisher's Exact test, is that the binary data are independent. If the proportions are correlated then more advanced techniques should be applied.

The test is based upon calculating directly the probability of obtaining the results that we have shown (or results more extreme) if the null hypothesis is actually true, using all possible 2×2 tables that could have been observed, for the same row and column totals as the observed data. These row and column totals are also known as marginal totals. What we are trying to establish is how extreme our particular table (combination of cell frequencies) is in relation to all the possible ones that could have occurred given the marginal totals.

For tables *rxk*, with *r* or/and *k* major than 2, it could be used the extension of Freeman-Halton to the Fisher Exact Test. Other alternative is to obtain the p-value using Monte Carlo simulation (9).

Mann-Whitney-Wilcoxon Test

Mann Whitney U o Wilcoxon rank-sum test

The Mann-Whitney U test is a nonparametric method for comparing two independent samples. This test is also called the Wilcoxon rank-sum test. The Mann-Whitney U test assumes only that we have independent random samples from the two groups. It does not assume anything about a normal distribution.

The null hypothesis for the Mann-Whitney U test is that the population distribution of the response variable—whatever that distribution might be— is the same for both groups. The alternative hypothesis is that the response variable tends to be larger for one group than for the other group.

To calculate the test statistic for the Mann-Whitney U test, we start by summing the ranks in each group. Define

R_1 = sum of ranks for Group 1; R_2 = sum of ranks for Group 2:

Then just pick one group or the other—it really doesn't matter which. Let's say we pick Group 1. Then the Mann-Whitney U test statistic[1] is

$$U = R_1 - \frac{n_1(n_1 + 1)}{2}$$

where n_1 is the number of subjects in Group 1.

If the null hypothesis is true, then every possible arrangement of the ranks among the two groups is equally likely. We can use this fact to calculate the probability of getting various values of the test statistic U under the null hypothesis, which in turn lets us calculate the p-value of the test.

If the p-value is less than or equal to α, then we reject the null hypothesis and conclude that one group tends to have larger response variable values than the other (one-sided alternative) or merely that there is a difference one way or the other (two-sided alternative).

If the p-value is greater than α, then we fail to reject the null hypothesis, which means we think the null hypothesis is reasonable, which means we think it's reasonable that there's no difference between the response variable values for the two groups.

Kaplan-Meier Method

The Kaplan-Meier estimator is a non-parametric statistic used to estimate the survival function from lifetime data. Kaplan-Meier estimate is one of the best options to be used to measure the fraction of subjects living for a certain amount of time after treatment.
An important advantage of the Kaplan–Meier curve is that the method can take into account some types of censored data, particularly right-censoring, which occurs if a patient withdraws from a study, is lost to follow-up, or is alive without event occurrence at last follow-up.
The Kaplan-Meier survival curve is defined as the probability of surviving in a given length of time while considering time in many small intervals. There are three assumptions used in this analysis. *Firstly*, we assume that at any time patients who are censored have the same survival prospects as those who continue to be followed. *Secondly*, we assume that the survival probabilities are the same for subjects recruited early and late in the study. *Thirdly*, we assume that the event happens at the time specified. The Kaplan-Meier estimate is also called as "product limit estimate". It involves computing of probabilities of occurrence of event at a certain point of time. We multiply

[1] There are a lot of other ways to calculate U, all of which give the same answer.

these successive probabilities by any earlier computed probabilities to get the final estimate. Total probability of survival till that time interval is calculated by multiplying all the probabilities of survival at all-time intervals preceding that time (by applying law of multiplication of probability to calculate cumulative probability) (10).

$$Prob\ (T > t) = S\ (t_i) = S\ (t_{i-1}) x\ S\ (\ ^{t_i}/_{t_{i-1}})$$

Two or more survival curves can be compared statistically by testing the null hypothesis i.e. there is no difference regarding survival among two or more interventions.

H_0: Survival functions are equals

H_a: Survival functions are different

This null hypothesis is statistically tested by log-rank test. The test statistic is

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where

k=number of groups.

O_i= total of observed events in the group i.

E_i= total number of expected events in the group i.

The test statistic and the significance can be drawn by comparing the calculated value with the critical value (using chi-square table) for degree of freedom equal to k-1. If the test statistic value is major than the critical value (using chi-square table) for degree of freedom equal to k-1 then we can say that there is significant differences between the groups regarding the survival.

Cox Proportional Hazards Regression Model

Cox proportional hazards regression model is used to assess the association between several risk factors, considered simultaneously, and survival time. The Cox proportional hazards model is called a semi-parametric model, because there is several important assumptions for appropriate use of the model that including: independence of survival times between distinct individuals in the sample, a multiplicative relationship between the predictors and the hazard, and a constant hazard ratio over time.

The Cox proportional hazards regression model can be written as follows:

$$h_i(t) = \lambda_0(t)e^{\beta_1 x_{i1}+\cdots+\beta_k x_{ik}}$$

Where

$h_i(t)$=the expected hazard at time t

$\lambda_0(t)$=the baseline hazard (represents the hazard when all of the predictors are equal to zero)

β_k=estimated coefficients of the model

x_{ik}=predictors of the model

In a Cox proportional hazards regression model, the measure of effect is the hazard rate, which is the risk of failure (i.e., the risk or probability of suffering the event of interest), given that the participant has survived up to a specific time. The exp (β_k) produces a hazard ratio (HR). The influence of each predictor into survival time must be tested through the hypothesis:

$$H_0: \beta_k = 0 \quad \text{or} \quad H_0: HR_k = 1$$

$$H_1: \beta_k \neq 0 \quad \text{or} \quad H_1: HR_k \neq 1$$

If the predictor is associated with survival (i.e. the null hypothesis above is rejected or confidence interval (HR) does not include the 1) then hazard ratio must be interpreted as follow: If the hazard ratio is less than 1, then the predictor is protective (i.e., associated with improved survival) and if the hazard ratio is greater than 1, then the predictor is associated with increased risk (or decreased survival).

The best model can be obtained using methods of variables selection (stepwise, backward and forward). The good fit of the final model and residuals must be evaluated (11).

8.2 Statistical Analyses

An exploratory and descriptive analysis will be developed for each variable in the study. All continuous variables will be summarized using the following descriptive statistics: n, mean, median, standard deviation, maximum and minimum and confidence interval (CI) will be performed when this information is considered relevant to describe a unique variable. The frequency and percentages of observed levels will be reported for all categorical measures. Ninety-five percent confidence intervals will be provided for estimates of interest wherever

possible. All summary tables will be structured with a column for each treatment and will be annotated with the total population size relevant to that table/treatment, including any missing observations.

The assumptions of normality and homoscedasticity of the variables for the use of parametric tests will be studied. All statistical tests will be performed with a significance level of 5%, unless otherwise specified. For qualitative comparison of independent samples the Chi-squared test (or Fisher's exact test for 2x2 tables) will be used, while for the quantitative samples Mann-Whitney-U will be used. The Kaplan-Meier limit-product method will be used to estimate PFS, OS and RD. The comparison of those parameters between the two treatment groups will be performed using the Log-Rank test. The Kaplan-Meier survival curve will be presented graphically. Median PFS, OS and RD with the 95% confidence interval will be reported. Cox regression models will be used to estimate unadjusted and adjusted hazard ratio and its 95% confidence interval. The Wald test will be used to establish the prognostic importance of each covariable. Variables that appear unbalanced in the baseline values will possibly be added to the Cox model.

8.2.1 Patient Disposition

A detailed description of patient disposition will be provided. It will include:

- summary of patients entered and by site
- total number of patients entered
- total number of patients enrolled
- summary of reasons for patients entered, but not enrolled
- total number of patients treated
- summary of reasons for patients enrolled, but not treated.

A detailed summary of reasons for patient discontinuation from study treatment will be provided.

A summary of all identified important protocol violations will be provided.

8.2.2 Patient Characteristics

Patient characteristics will include a summary of the following:

- patient demographics

- baseline disease characteristics
- preexisting conditions/secondary conditions
- prior therapy

Other patient characteristics will be summarized as deemed appropriate.

Standard descriptive statistics, such as the mean, median, range and proportion, will be used to summarize the patient sample and to estimate parameters of interest. Ninety-five percent confidence intervals will be provided for estimates of interest where possible.

8.2.3 Concomitant Therapy

A summary of concomitant therapies will be generated in the safety population.

8.2.4 Treatment Compliance

Treatment information will be collected at each dose administration. The estimate of percent compliance will be given by:

$$\text{Percent Compliance} = \frac{\text{Actual dose administered per week}}{\text{Dose expected to be administered per week}} \times 100$$

No minimal level of compliance will be defined for patient inclusion in efficacy analyses. To be considered compliant patients should have received at least 80% of the planned number of doses. Exploratory analysis of the impact of compliance on selected efficacy endpoints may be performed if deemed necessary.
The number and frequency of cycles received in each group of treatment will be reported.
The Relative Dose Intensity (RDI) will be calculated for all drugs and median and range will be mainly described.
All cases of study treatment dose modification will be listed, grouped by reasons. The number and frequency of delays, reductions and omissions will also be reported.

8.2.5 Efficacy Analyses

All efficacy analysis will be based on the ITT and *ESR1* wild type populations. Additional PFS and OS analyses will be performed on the per-protocol population. All analyses will be conducted at a 5% level of significance. No adjustments are planned for multiple testing/comparisons in those secondary and supportive hypothesis tests other than the primary analyses of PFS and secondary analyses of OS.

8.2.5.1 Analyses of Primary Endpoint

Progression-Free Survival (PFS)

A modification of Hochberg's method (6) will be used for two primary treatment comparisons to provide control of experiment-wise Type 1 error rate at a 5% significance level. With this closed testing method, statistical significance applies to both p_1 for Cohort 2 and p_2 for Cohort 1+ Cohort 2 (ESR1 wild type) if $p_1<0.05$ and $p_2<0.05$; or it applies only to Cohort 2 if $p_1\leq0.025$ when Cohort 1+ Cohort 2 (ESR1 wild type) has $p_2>0.05$; or it applies to only Cohort 1+ Cohort 2 (ESR1 wild type) if $p_2\leq0.025$ when Cohort 2 has $p_1>0.05$ (assuming the analyses results represented by both p_1 for Cohort 2 and p_2 for Cohort 1+ Cohort 2 (ESR1 wild type) demonstrating the combination of palbociclib and fulvestrant is superior to capecitabine in prolonging PFS).

The primary analyses of PFS will be performed in the ITT population for Cohort 2 and *ESR1* wild type population for Cohort 1 + Cohort 2. A stratified log-rank test will be used to compare PFS time between the 2 treatment arms at the interim and/or final analyses with the overall significance level preserved at 5%. PFS time associated with each treatment arm will be summarized for the ITT/*ESR1* wild type populations using the Kaplan-Meier method and displayed graphically where appropriate. Confidence intervals (CIs) for the 25th, 50th and 75th percentiles of the event-free time will be reported. The Cox Proportional hazards model will be fitted to compute the treatment hazard ratio and the corresponding 95% CI. Additionally a similar analysis will be also performed in the per-protocol population.

The secondary analysis of PFS will be performed in the ITT population for Cohort 1 + Cohort 2. Statistical methods for this analysis will be the same that outlined above without adjusting for the multiplicity.

8.2.5.2 Analyses of Secondary Endpoints

Overall Survival (OS)

OS will be analyzed in the ITT population for Cohort 2 and *ESR1* wild type population for Cohort 1 + Cohort 2. A stratified log-rank test will be used to compare OS time between the 2 treatment arms at the interim and/or final analyses with the overall significance level preserved at 0.10. The stratification factor(s) will be the same as for the PFS analysis. OS for the two arms will be assessed using Kaplan-Meier methods and displayed graphically where appropriate. The median event times and 95% CIs will be estimated. Cox regression models will be used to estimate the treatment hazard ratio and its 95% confidence interval. Additionally a similar analysis will be also performed in the per-protocol population.

Objective Response Rate (ORR)

A patient will be considered to have achieved an OR if the patient has a sustained complete response (CR) or partial response (PR) according to RECIST v.1.1 definitions (4). Otherwise, the patient will be considered as non-responders in the OR rate analysis. Additionally, patients with inadequate data for tumor assessment (eg, no baseline assessment or no follow-up assessments) will be considered as non-responders in the OR rate analysis.

The OR rate (ORR) on each randomized treatment arm will be estimated by dividing the number of patients with objective response (CR or PR) by the ITT/*ESR1* wild type patients with measurable disease by treatment arm ("response rate").

$$\text{Objective Response Rate} = \frac{\text{Number of CRs + PRs}}{\text{ITT or ESR1 wild type population with measurable disease}}$$

The ORR will be reported, including a 95% confidence interval. ORR comparison between the two treatment arms will be assessed using Cochran-Mantel-Haenszel (CMH) test with the same stratification factors as for the PFS analysis.

In addition, the best objective response for each patient will be summarized by treatment arm.

Response Duration (RD)

RD will only be calculated for the subgroup of patients with an objective response. RD for the two treatment arms will be summarized using Kaplan-Meier methods and displayed graphically, where appropriate. The median event time and 95% CI for the median will be provided for each endpoint. RD comparison between the two treatment arms will be assessed using Log-rank test.

Clinical Benefit Rate (CBR)

Clinical benefit (CB) rate (CBR) on each randomized treatment arm will be estimated by dividing the number of patients with CR, PR, or SD ≥24 weeks by the ITT/*ESR1* wild type populations by treatment arm. A 95% CI for the CBR will be provided. CBR comparison between the two treatment arms will be assessed using CMH test with the same stratification factors as for the PFS analysis.

$$\text{Clinical Benefit Response Rate} = \frac{\text{Number of CRs} + \text{PRs} + \text{SD} \geq 24 \text{ weeks}}{\text{ITT or ESR1 wild population}}$$

All of the above secondary analyses will be conducted at a two-sided 0.05 level of significance.

8.2.6 Safety Analyses

The toxicity and tolerability of study drugs/medications will be evaluated in the safety population. Safety analyses will include summaries of the incidence of adverse events by maximum CTCAE grade (v4.0; NCI 2010) (5) that occur during the study treatment period or within 30 days of the last dose of study treatment, regardless of causality and according to the relationship to study drug/medication as assessed by the investigator. Additionally, the following safety-related outcomes will be summarized:

- study treatment discontinuations due to adverse events.
- deaths
- SAEs
- hospitalizations and transfusions
- use of key concomitant medications or growth factors.

Analyses for data with discrete dates, for example, deaths, transfusions, and concomitant medications, will be performed through 30 days after each patient's last dose of study treatment. Adverse events will also be analyzed in this timeframe; that is, if an event starts within 30 days of discontinuation from study treatment, but after 30 days after the last dose of study treatment, it will not be included.

Adverse events data and serious adverse events will be presented in frequency tables by grade. Hematological and clinical biochemistry toxicities will be assessed from laboratory test parameters. The safety analysis will be performed in the safety population.

An early safety review will be performed by the IDMC. In addition, the interim analyses for PFS will provide the potential to identify early any unexpected safety issues associated with the palbociclib combinations.

8.2.7 Other Analyses

8.2.7.1 Pharmacokinetic Analysis

Within-patient average trough concentrations will be listed by patient by analyte. Summary statistics will be provided for palbociclib trough concentrations by study cycle and for within-patient average trough concentrations. Summary statistics will be provided for exemestane trough concentrations by nominal collection time, for within-patient average trough concentrations during the exemestane lead-in period, and for the within-patient average trough concentrations from Day 14 of Cycles 1 and 2.

Statistical Analysis Plan for PK analysis will be presented in an independent document.

8.2.7.2 Biomarker, Pharmacogenomic and Pharmacodynamic Analysis

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 will be used to investigate any possible relationship of biomarker levels with palbociclib plus endocrine treatment anti-tumor efficacy. Once they are decided it will be developed an appendix to this SAP explaining the methods before the analysis takes place.

8.2.7.3 Patient Reported Outcomes

Breast cancer-specific quality of life scores and change from baseline scores will be compared between the treatment arms using a mixed model repeated measures approach adjusting for specified covariates. In addition, analyses will be performed to determine if the change from

baseline scores achieve the appropriate minimally important difference (MID) cut-off for the scale being examined.

In addition to the above analyses, an examination of the time to deterioration composite endpoint will be carried out using survival analysis methods. A composite definition for deterioration based on death, tumor progression, and/or breast cancer-specific quality of life subscale MIDs may be used.

8.2.7.4 Subgroup Analyses

Exploratory Subgroup analysis for the primary endpoint will be performed for the following groups using Cox-proportional hazard model.

- Visceral vs. non-visceral
- Prior sensitivity to hormonal treatment (yes vs. no)
- Prior chemotherapy for MBC (yes vs. no)
- Number of involved sites (one vs. multiple)
- Lesion measurable vs. Lesion Not Measurable
- Age Group 65 vs. ≥65
- Treament line (1^{st} vs 2^{nd} vs 3^{rd})

Other exploratory subgroups may be used if deemed appropriate.

Forest-plot will present graphically the subgroup analysis for PFS.

8.2.7.5 Multivariate Analyses

A multiple Cox regression analysis will be carried out for PFS and OS, in order to adjust the comparison of the treatment for the principal prognostic factors. These factors include age, visceral disease, measurable disease, number of involved sites, prior sensitivity to hormonal treatment, prior chemotherapy for MBC and treatment line. The Wald test will be used to establish the prognostic importance of each covariate. Covariates that appear unbalanced in the baseline values will possibly be added to the Cox model.

9. TABLES Y FIGURES

Table/Figure No.		Table/Figure title
1.		Recruitment per site
2.		Consort Diagram
	2.1	Deviations from protocol
	2.2	Study population
3		Demographic characteristics
4		Disease characteristics
5		Treatment
	5.1	Cycles administered
	5.2	Received Dose Intensity
	5.3	Dose Modifications: Delays, reductions, omissions
	5.4	Reasons of treatment discontinuation
6		Safety analysis
	6.1	Adverse events per Patient
	6.2	Serious Adverse Events (SAEs)
7		Efficacy analysis
	7.1	Progression Free Survival
	7.2	Overall survival
	7.3	Causes of death
	7.4	Objective Response
	7.5	Response Duration
	7.6	Clinical Benefit Response
8		Subgroup Analysis
	8.1	Forest Plot
9		Multivariate Analysis
	9.1	Univariate analysis
	9.2	Multivariate Analysis

The numbers of tables and figures do not have to match exactly with those of the statistical report.

10. APPENDIX

10.1 Mock Tables

See attachment "Mock Tables.pdf" *(Note: This is a reference document, the tables do not have to match exactly with those of the statistical report).*

11. BIBLIOGRAFHY /REFERENCES

1. *ICH E9, Statistical Principles for Clinical Trials.*
2. *ICH E3, Structure and Content of Clinical Study Reports.*
3. *GEICAM. Protocol GEICAM/2013-02 Version 5.0.*
4. *Eisenhauer et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1).* s.l. : European Journal of Cancer 45 228-247, 2009.
5. *Services, U.S. Department of Health and Human. Common Terminology Criteria for Adverse Events (CTCAE) version 4.0.* May 28, 2009.
6. *Gary G. Koch and Todd A. Schwartz. An Overview of Statistical Planning to Address Subgroups in Confirmatory Clinical Trials.* s.l. : Journal of Biopharmaceutical Statistics, 2014, Vols. 24: 72–93.
7. *Kim K, DeMets, DL. Confidence intervals following group sequential tests in clinical trials.* s.l. : Biometrics, 1987, Vols. 43 (4): 857-64.
8. *Cui L, Hung HMJ, Wang S, et al. Modification of sample size in group-sequential clinical trials.* s.l. : Biometrics, 1999.
9. *P08 - 2008 Confidence Interval Calculation for Binomial Proportions: Keith Dunnigan Statking Consulting, Inc.*
10. *Hope, Adery C.A. A Simplified Monte Carlo Significance Test Procedure.* 3, 1968, Journal of the Royal Statistical Society, Vol. 30, págs. 582-598.
11. *Manish Kumar Goel, Pardeep Khanna and Jugal Kishore. Understanding survival analysis: Kaplan-Meier estimate.* s.l. : Int J Ayurveda Res., 2010 Oct-Dec, Vols. 1 (4): 274-278.
12. *David W. Hosmer and Stanley Lemeshow. Applied Survival Analysis: Regression Modeling of Time to Event Data.*