[redacted]

Statistical Analysis Plan (SAP)

Final Version 2.0
04th March 2020

EudraCT Number	2014-004698-17
Protocol Number	MedOPP067 (PARSIFAL)
Protocol Version Date	20th December 2016
Title	A randomized, multicenter, open-label, phase II trial to evaluate the efficacy and safety of palbociclib in combination with fulvestrant or letrozole in patients with HER2 negative, ER+ metastatic breast cancer
Sponsor	Medica Scientia Innovation Research MedSIR ARO Rambla de Catalunya 2, 2D 08007 Barcelona (Spain)
Principal investigator	Antonio Llombart MD, PhD Medical Oncologist Chief of Oncology Department Arnau de Vilanova Hospital, Valencia (Spain)

STATISTICAL ANALYSIS PLAN (SAP)

Performed by:

SAIL Servicio de Asesoría a la Investigación y Logística

Avenida Meridiana 350, 9 D
08027 Barcelona (Spain)
www.sail-biometria.com

Signature Page

EudraCT Number: **2014-004698-17**
Protocol Number: **MedOPP067**

Written by:

Signature: 

Date: 06/03/2020

Xavier Masramon
Biostatistician
SAIL
Barcelona, Spain

Sponsor Signature:

The undersigned hereby declare that they have examined the Statistical Analysis Plan document and agree to its form and content.

Represented by:

Signature:		Signature:	
Date:	06/03/2020	Date:	06/03/2020
Name:	Lourdes Gómez	Name:	Miguel Sampayo
Title:	International Trial Lead, MEDSIR	Title:	Statistician

SAP Revision History:

Version Number	Date	Changes
1.0	15th February 2018	New
2.0	09th December 2019	The primary endpoint is PFS assessed by investigator criteria, instead of PFS per RECIST v.1.1 criteria.

LIST OF ABBREVIATIONS

Abbreviation	Definition
ABC	Advanced Breast Cancer
AE	Adverse event
ARO	Academic Research Organization
ATC	Anatomical Therapeutic Chemical
BC	Breast cancer
BPM	Beats per minute
CB	Clinical benefit
CBR	Clinical Benefit Rate
CI	Confidence interval
CPMP	Committee for proprietary medicinal products
CR	Complete response
CTCAE	Common Terminology Criteria for Adverse Events
DOR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
ECRF	electronic Case Report Form
EMEA	European Medicines Agency
EOS	End of Study
ER	Endocrine receptors
EWP	Efficacy Working Party
FDA	Food and Drug Administration
GCP	Good Clinical Practice
HER2	human epidermal growth factor receptor 2
HR	Hazard ratio
ICH	International Conference on Harmonization
ID	Identification
IQR	Interquartile Range
IMP	Investigational Medicine Products
ITT	Intention-to-Treat population
MBC	Metastatic Breast Cancer
MEDDRA	Medical Dictionary for Regulatory Activities
ORR	Overall Response Rate
OS	Overall Survival
PD	Progression Disease
PFS	Progression Free Survival
PP	Per Protocol Population
PR	Partial Response
PT	Preferred term
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic acid
RR	Relative Risk
RDI	Relative Dose Intensity
SAE	Serious Adverse Event
SAIL	SAIL S.L.L.
SAP	Statistical Analysis Plan

Abbreviation	Definition
SAS	Statistical Analysis Software
SC	Steering Committee
SD	Stable Disease
SI	International System of Units
SOC	System Organ Class
STD	Standard deviation
TEAE	Treatment emergent adverse event
TLF	Tables, listings and figures
TNM	Tumor Node Metastasis
TTP	Time to progression
TTR	Time to response
UAE	United Arab Emirates
WHO	World Health Organization

TABLE OF CONTENTS

1 INTRODUCTION

1.1 General

The present study (PARSIFAL) is an open-label, randomized, controlled multicenter phase II study with the aim of assessing the efficacy and safety of the combination palbociclib plus fulvestrant vs. palbociclib plus letrozole in terms of PFS in women with ER-positive advanced breast cancer (ABC).

The purpose of this statistical analysis plan (SAP) is to provide a protocol specific description of the statistical analysis that will be performed to produce an integrated clinical/statistical report.

This SAP is based upon the following study documents:

- Initial protocol version dated: 28th January 2015 (applicable globally)
- Protocol version dated: 30th July 2015 (applicable in France, Russia, Saudi Arabia & UAE)
- Protocol version dated: 7th October 2015 (applicable in Germany)
- Protocol version dated: 20th January 2016 (applicable globally)
- Protocol version dated: 30th June 2016 (applicable globally)
- electronic Case Report Form (eCRF), Version 3.0 (September 16, 2016)
- Protocol version dated: 20th December 2016 (applicable globally)

1.2 Type of Study

This is an international, randomized, open-label, controlled, multicenter phase II study of parallel groups.

1.3 Study Protocol Amendments

- Initial Version dated 28th January 2015 (applicable globally)
- Version dated 30th July 2015 (applicable in France, Russia, Saudi Arabia & UAE)
- Version dated 7th October 2015 (applicable in Germany)
- Version dated 20th January 2016 (applicable globally)
- Version dated 30th June 2016 (applicable globally)
- Version dated 20th December 2016 (applicable globally)

1.4 Study Population

Postmenopausal women and premenopausal women receiving LHRH analogues, aged ≥ 18 years with ER positive and HER2 negative locally advanced or metastatic breast cancer (MBC) that had not received any therapy for the metastatic disease. Subjects must have histologic confirmation of the estrogen and/or progesterone-positive and HER2 negative receptors breast cancer. Evidence of measurable or evaluable metastatic disease is required.

1.5 Study Design

At least 486 eligible patients will be randomized 1:1 to receive either palbociclib plus fulvestrant (interventional arm) or palbociclib plus letrozole (control arm).

The total of 486 randomized patients was reached on the 8th of January 2018.

The randomization code was generated by SAIL prior to the study using PROC PLAN available in SAS® version 9.4 statistical program. The seed number, for each list, was chosen randomly using RANUNI (function available in SAS), that it was chosen because random numbers are known to follow uniform distribution.

Forty codes were generated, with 20 blocks of length 2, for each stratum. Taking to account 2 sites of disease (visceral and non-visceral), 2 metastatic diseases (diagnosed de novo and non-diagnosed de novo) and 100 centers, the number of stratums were 400. Therefore, a total of 16,000 codes were generated.

The stratum "metastatic disease" for the first randomized patients (see the initial protocol) was the nature of prior (neo)adjuvant anticancer treatment received (prior hormonal therapy; no prior hormonal therapy) instead of metastatic diseases (diagnosed de novo and non-diagnosed de novo). The protocol amendment #1 describes the replacement of the stratification criterion "prior vs. non-prior hormonal therapy" by "de novo vs. non de novo metastatic disease". Therefore, after protocol amendment #1, the stratum "metastatic disease" was already the metastatic diseases (diagnosed de novo and non-diagnosed de novo).

In addition, it has been observed some patients that the stratums (sites of disease and/or metastatic diseases) were modified after the assigned treatment (randomization).

All efficacy endpoints are going to be analyzed adjusting by the site of disease (visceral and non-visceral) and by the metastatic disease (diagnosed de novo and non-diagnosed de novo).

Patients in the interventional arm (A) will receive fulvestrant 500 mg/5mL i.m. **injection** administered on Days 1 of a 28-day cycle (loading dose on cycle 1 requires administration also on Day 14). Palbociclib 125 mg capsules will be taken orally once daily beginning on Day 1 of fulvestrant and continuing through Day 21 of every 28-day cycle.

Patients in the control arm (B) will receive letrozole 2.5 mg tablet **orally** once daily beginning on Day 1 and continuing through Day 28 of a 28-day cycle. Palbociclib 125 mg capsules will be taken orally once daily beginning on Day 1 of letrozole and continuing through Day 21 of every 28-day cycle.

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 discontinuing the active treatment phase will enter a treatment follow-up period during which survival and new anti-cancer therapy information will be collected every 6 months from the last dose of investigational product. The treatment follow-up period will continue up to 12 months after last patient randomization.

1.6 Study Schedule

The schedule of visits for this study and procedures to be performed at each visit are shown in the following table (see appendix 1 of the protocol).

Study Period	**Screening**	**Treatment period**	**Treatment follow up period**	
Day	**-28 to -1**	**Each cycle (every 28 days)**	**28±7 last dose treatment**	**Every 6 months**
Informed Consent	X			
ER and HER2 status	X			
Baseline sings/symptoms	X			
Check of inclusion/exclusion criteria	X			
Post-menopausal status confirmation	X			
Medical History	X			
Physical Examination and ECOG	X	X	X	
Weight and Vital signs	X	X	X	
Concomitant Medication Reporting	X	X	X	
Review Patient Diary		X		

Study Period		**Screening**	**Treatment period**	**Treatment follow up period**	
Day		**-28 to -1**	**Each cycle (every 28 days)**	**28±7 last dose treatment**	**Every 6 months**
AE reporting		X	X	X	
12-lead ECG		X	X	X	
Tumor Assessments		X	X	X	X
Samples for translational sub-study:					
Primary tumor biopsy		X			
Blood samples for translational sub-study		X	X	X	
Biopsies from metastatic lesions		X	X		
Standard Laboratory Procedures:					
Hematology		X	X	X	
Biochemistry		X	X	X	
Treatment Administration:					
Arm A	Fulvestrant 500 mg/5mL (i.m. injection)		D1 & D14 (cycle 1) D1 (other cycles)		
	Palbociclib 125 mg total dose (capsules)		D1 to D21		
Arm B	Letrozole 2.5 mg total dose (tablets)		D1 to D28 continuously		
	Palbociclib 125 mg total dose (capsules)		D1 to D21		

1.7 Sample Size

The sample size calculations were described in the protocol, section 8.1, using the following wording:

Superiority analysis:

Based on published efficacy data for palbociclib and fulvestrant in similar target population, the investigator hypothesis (H1) is that median PFS in the palbociclib plus fulvestrant arm (31.3 months) will be higher (Hazard Ratio = 0.7) than in the palbociclib plus letrozole group

(22 months). Therefore, we will test the null hypothesis (H0) that median PFS survival is equal in both groups.

The analysis will be performed with Log-Rank test. We assumed an exponential survival function. We estimate a 24 months (mo.) accrual period and a 12 mo. treatment period (maximum follow-up of 36 mo.). We planned a randomization (1:1). Regarding type I and type II errors, we assumed a power of 80% and a two-sided overall alpha error of 5%. An interim analysis will occur at 22 mo. with 89 events (35% of total events expected). The final analysis will be performed at approximately 36 mo. with 254 events and 486 patients included (52% PFS event rate assumed).

According to Lan-DeMets O'brien-Fleming approximation spending function, the twosided local type I error for testing the null-hypothesis within one interim and final analysis will be 0.001 and, 0.0498, respectively.

Therefore, we should include 243 patients in the control arm and 243 patients in the interventional arm. A total of 486 patients will be included in this study.

The randomization will be stratified. It can be expected that including factors of prognostic importance in the cox regression model as defined for the confirmatory analysis will increase the power as compared to the log-rank test. However, we preferred to take a conservative approach and we accept the sample size calculated without adjusting for prognostic factors.

Non-inferiority analysis:

As per EMEA guidelines, we will switch to non-inferiority analyses if the superiority criteria cannot be met (CPMP/EWP/482/99 EMEA guideline) (1)*. We will declare noninferiority if the upper bound 95% confidence interval (95%CI) of HR between median PFS in palbociclib plus letrozole and palbociclib plus fulvestrant arms, will fall within the non-inferiority margin of 1.21. Non-inferiority margin is justified according with the FDA guidance in non-inferiority studies (FDA GUIDANCE: Non-Inferiority Clinical Trials to Establish Effectiveness)* (2)*. They propose to estimate the average effect of the active control over placebo in historical studies and selecting the 95%CI lower bound. Finally, this value should be adjusted to retain at least 50% of the historical effect of active control versus placebo arms. Accordingly the combined effect of PALOMA-1 and 2 studies are (HR: 1.79. 95%CI: 1.47 to 2.18) and the non-inferiority margin is 1.21. With 254 PFS events, if the HR ≤ 0.94 or median PFS in palbociclib plus letrozole is 22 mo. vs. palbociclib plus fulvestrant 23.3 mo. or better, the upper bound 95% of CI will fall within the non-inferiority margin of 1.21, allowing for the determination that both combinations have a similar treatment effect.*

The sample size estimation was made using R (package gsDesign), according to the formulas published by Lachin JM and Foulkes MA (1986) (3)*.*

Meta-analysis methods to combine Paloma 1 and Paloma 2 trials:

The hazard ratios of Paloma 1 (and Trio-18) (*Finn RS et al. 2015)* (4).

And Paloma 2 (*Finn RS et al. 2016*) (5) trials between Palbociclib plus Letrozole against Letrozole arm were (HR = 0.488, 95%CI: 0.319 to 0.748) and (HR = 0.58, 95%CI: 0.46 to 0.72), respectively. In accordance with previous articles (*Tierney JF et al. 2007*) (6) we combine in meta-analysis the ln(HR) from both studies. The variance of ln(HR) was estimated as:

$$= \left[\frac{\ln(\text{upper CI}) - \ln(\text{lower CI})}{2 \times \text{z score for upper CI boundary}} \right]^2$$

The combined effect of ln(HR) were calculated in a random effects meta-analysis. The restricted maximum-likelihood random effect model was used to derive the overall estimates and the 95% confidence intervals (CIs). We assessed heterogeneity through the use of both the I-square test statistic and Chi-square test (I2 = 0%, p-value = 0.4819). There was not relevant (I2<25%), nor significant heterogeneity between studies (p>0.1).

2 STUDY OBJECTIVES

2.1 Primary Objective

To compare the efficacy of the combination of palbociclib plus fulvestrant versus palbociclib plus letrozole in terms of progression free survival (PFS) in patients with hormone-sensitive HER2-negative metastatic or locally advanced breast cancer.

2.2 Secondary Objectives

- To evaluate the safety and tolerability of the combination of palbociclib plus fulvestrant or letrozole.
- To correlate the safety profile of palbociclib combined with fulvestrant or letrozole with baseline patient characteristics.
- To compare the time to progression (TTP) of the combination of palbociclib plus fulvestrant with palbociclib plus letrozole.
- To compare the clinical response (in terms of clinical benefit and overall response) of the combination of palbociclib plus fulvestrant with palbociclib plus letrozole.
- To compare the duration of response (DoR) of the combination of palbociclib plus fulvestrant with palbociclib plus letrozole.
- To compare time to response (TTR) of the combination of palbociclib plus fulvestrant with palbociclib plus letrozole
- To compare the overall survival (OS) of the combination of palbociclib plus fulvestrant with palbociclib plus letrozole.

- To identify potential biomarkers to predict the benefit from Palbociclib combined with endocrine therapy
- To identify mechanisms of resistance to palbociclib combined with endocrine therapy

3 ANALYSIS POPULATIONS

Patients will be classified into one of the following populations: Screening Population, Safety Population, Intent-To-Treat (ITT) Population and Per Protocol (PP) and exploratory evaluable Population.

Exclusions from efficacy analyses, in accordance with definitions of the populations, will be reviewed by clinical and statistical teams prior to database release.

3.1 Screening Population

All patients who were present at the screening visit will be included in the screening population.

3.2 Safety Population

All randomized patients who will receive one drug exposure of study treatment will be included in the safety analysis set, except for patients who will receive one drug exposure but are immediately lost to follow-up.

All safety analyses will be based upon the Safety population.

3.3 Intent-To-Treat Population

A patient will be included in the Intent-to-Treat (ITT) population if the patient has been randomized. The patient will be analyzed in the arm to which they were randomly assigned, regardless of their adherence with the entry criteria, regardless of the treatment they received, and regardless of subsequent withdrawal from treatment or deviation from the protocol.

3.4 Per Protocol Population

A patient will be included in the Per Protocol (PP) population if the subject has been included in the ITT population, who fulfills all eligibility criteria, has started treatment on the assigned arm, and without major protocol deviations.

3.5 Exploratory evaluable population

Exploratory analyses will be performed on those patients in the ITT (all randomized) population who consented to participate in the exploratory research program and were evaluable for biomarker status.

3.6 Protocol Deviations

Protocol deviations will be recorded in the SAIL database. A by-patient listing of protocol deviations and a by-patient listing of inclusion and exclusion criteria not met will be provided.

Major protocol deviations are defined as those deviations from the protocol likely to have an impact on the perceived efficacy and/or safety of study treatments.

To determine the Per Protocol Population, all protocol deviations will be reviewed prior to database lock to determine which ones should be classified as major deviations. Patients with major protocol deviations will be excluded from the Per Protocol Population.

Major protocol deviations may be discussed in the data review meeting. Criteria for determining the "per protocol" group assignment would be established by the Steering Committee (SC) before the statistical analysis begins.

Major protocol deviations and any action to be taken regarding the exclusion of subjects or affected data from specific analyses are defined below.

A summary of the number and percentage of subjects with a major protocol deviation by type of deviation will be provided. Also, a by-subject listing of major and minor protocol deviations will be provided.

PD	Description	Category	Grade	Rationale	Comments
01	Postmenopausal women	Inclusion/Exclusion criteria	Major	Efficacy	Automatic check
02	ECOG	Inclusion/Exclusion criteria	Case by case	Safety	Automatic check
03	Tumor lesion must have ER+/Her2-	Inclusion/Exclusion criteria	Major	Efficacy	Automatic check
04	Patients should not be candidates for surgery treatment	Inclusion/Exclusion criteria	on case by case	Efficacy	Study treatment is not the therapeutic approach for these patients
05	Patients must have evaluable disease	Inclusion/Exclusion criteria	Major	Efficacy	Manual review
06	Life expectancy ≥ 12 months	Inclusion/Exclusion criteria	on case by case	Efficacy	Manual review
07	Adequate organ function	Inclusion/Exclusion criteria	on case by case	Efficacy Safety	If criteria are not met, patients would not be treated at full dose or high probability of early

PD	Description	Category	Grade	Rationale	Comments
					discontinuation
08	Adequate follow-up	Inclusion/Exclusion criteria	Major	Efficacy	If no post-treatment evaluation according to protocol
09	Informed consent form not signed	Inclusion/Exclusion criteria	Major	GCP	Patient data could not be used for efficacy assessment
10	No other malignancies within 5 last years	Inclusion/Exclusion criteria	on case by case	Efficacy	Manual review on medical history
11	No relevant toxicities at baseline	Inclusion/Exclusion criteria	on case by case	Efficacy	Manual review on medical history
12	No receptors histology	Inclusion/Exclusion criteria	Major	Efficacy	Automatic checks
13	DFI to previous endocrine therapy ≥12 months	Inclusion/Exclusion criteria	Major	Efficacy	Manual review of previous anticancer treatments
15	Major surgery or sequalae within 4 weeks prior to randomization	Inclusion/Exclusion criteria	on case by case	Safety	Same as above
16	Active bleeding disorders	Inclusion/Exclusion criteria	on case by case	Efficacy Safety	If clinically relevant**
17	Serious concomitant systemic disorder, including conditions that could interfere with absortion	Inclusion/Exclusion criteria	on case by case	Efficacy Safety	Same as above
18	Unable to swallow tablets	Inclusion/Exclusion criteria	on case by case	Efficacy	Major if compliance lower than 80%
19	Chronic treatment with corticosteroids	Inclusion/Exclusion criteria	on case by case	Efficacy Safety	To define threshold for major PD Manual review on ConMed
20	Prolonged QTc	Inclusion/Exclusion criteria	on case by case	Safety	If clinically relevant
21	Uncontrolled electrolyte disorders	Inclusion/Exclusion criteria	Major	Safety	If clinically relevant**
22	IMP administered does not correspond to randomization arm	Randomization	Major	Efficacy	
23	IMP no administered but randomization completed	Randomization	Major	Efficacy	Study discontinuation prior to treatment start
24	Hypersensitivity to IMP	Investigational Medicinal Product	on case by case	Efficacy Safety	major if affects compliance
25	IMP overdose	Investigational Medicinal Product	Major	Efficacy Safety	If compliance more than 20% vs. prescribed/expected dose
26	IMP underdose	Investigational Medicinal Product	Major	Efficacy Safety	If compliance less than 80% vs. prescribed/expected dose
27	IMP administration dosing/schedule	Investigational Medicinal Product	on case by case	Efficacy Safety	If administered dose is systematically deviating from expected dose
28	IMP toxicity	Investigational Medicinal Product	On a case by case	Safety	If toxicity has not managed appropriately Listing with AE's grade 3/4 with action = none before the EOT date

PD	Description	Category	Grade	Rationale	Comments
29	Tumor assessment not done	Study procedure	Major	Efficacy Safety	
30	Tumor assessment out of the window	Study procedure	Minor	Efficacy	If done more than every 12 weeks +7 days
31	No baseline tumor assessment	Study procedure	on case by case	Efficacy	If not baseline values any post-baseline data indicating presence of tumor lesions will be considered evidence of PD Flag if baseline data is previous to 28 days before day 1 cycle 1. For bone scan flag if more than 60 days
32	No post-baseline tumor assessment	Study procedure	Major	Efficacy	if no post baseline assessment or if only post-baseline assessment is not at least after 8 weeks of treatment start.
35	Prohibited medication was taken	Concomitant medication	On a case by case	Efficacy Safety	If CYP3A inducers/inhibitors
36	Prohibited medication was taken	Concomitant medication	Major	Efficacy	If any therapies intended for the treatment of cancer prior to EoT (e.g., antineoplastics)
37	Prohibited medication was taken	Concomitant medication	On a case by case	Safety	If it cause QT interval prolongation
38	Subject not withdrawn as per protocol	Administrative and Other	on case by case	GCP	If data is obtained after withdrawal

4 DEFINITION OF ENDPOINTS

The overall response at each radiological assessment, used in the definition of the following efficacy endpoints, will be according to RECIST criteria guidelines version 1.1.

In accordance with RECIST v1.1 guideline confirmation of response is only required in non-randomized trials where response is the primary endpoint. Our study does not meet these criteria, so confirmatory measurement of complete and partial response has not been requested to investigators.

4.1 Primary Efficacy Endpoints

The primary efficacy endpoint is progression free survival (PFS) analyzed in the ITT and PP population.

The PFS is defined as the time from randomization until death by any cause or objective tumor progression or clinical disease progression, as assessed by investigator criteria. Patients with no progression or death will be censored at the date of their last evaluable imaging previous 35 days after last dose administration date.

Censoring rules are specified below:

Situation	Date of progression or censoring	Outcome
Progression documented between scheduled visits	Earliest of: • Date of assessment by investigator (if progression is based on clinical criteria); or • Date of assessment showing new lesion (if progression is based on new lesion); or • Date of last radiological assessment of measured lesions (if progression is based on increase in sum of measured lesions).	Progressed
Death before first progression disease (PD) assessment	Date of death.	Progressed
Death between adequate assessment visits	Date of death.	Progressed
No progression	Date of last radiological assessment of measured lesions.	Censored
Treatment discontinuation for undocumented progression	Date of last radiological assessment of measured lesions.	Censored
Treatment discontinuation for toxicity or other reason	Date of last radiological assessment of measured lesions.	Censored
Death or progression after more than one missed visit	Date of last radiological assessment of measured lesions.	Censored

4.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints will be:

- The PFS as assessed per RECIST v.1.1 criteria.
- The unconfirmed Best Overall Response, defined as the best overall response recorded from the start of the study treatment until 35 days after last dose administration date and confirmation of response is not required.
    - When CR or PR is the best response across all time points, then best overall response will be CR or PR respectively.
    - When SD is the best response for ≥ 24 weeks the best overall response will be SD ≥ 24 w.
    - When SD is the best response for < 24 weeks the best overall response will be SD < 24 w.
    - When non-target disease only and Non-CR/Non-PD is the best response for ≥ 24 weeks the best overall response will be SD ≥ 24 w.

  - When non-target disease only and Non-CR/Non-PD is the best response for < 24 weeks the best overall response will be SD < 24 w.
  - When PD is the best response across all time points, best overall response will be PD.
  - When there is no evaluable tumor assessments best overall response will be NE.

- The confirmed Best Overall Response, defined as the best overall response recorded from the start of the study treatment until 35 days after last dose administration date and confirmation of response is required.
  - When CR or PR is the best response for ≥ 4 weeks then best overall response will be CR or PR respectively.
  - When SD is the best response for ≥ 24 weeks the best overall response will be SD ≥ 24 w.
  - When SD is the best response for < 24 weeks the best overall response will be SD < 24 w.
  - When non-target disease only and Non-CR/Non-PD is the best response for ≥ 24 weeks the best overall response will be SD ≥ 24 w.
  - When non-target disease only and Non-CR/Non-PD is the best response for < 24 weeks the best overall response will be SD < 24 w.
  - When PD is the best response across all time points, best overall response will be PD.
  - When there is no evaluable tumor assessments best overall response will be NE.
- Unconfirmed Objective Response Rate (ORR) is defined as the proportion of patients with best overall response of unconfirmed CR or unconfirmed PR.
- Unconfirmed Clinical Benefit Rate (CBR) is defined as the proportion of patients with best overall response of unconfirmed CR or unconfirmed PR or SD ≥ 24 w.
- Confirmed Objective Response Rate (ORR) is defined as the proportion of patients with best overall response of confirmed CR or confirmed PR.
- Confirmed Clinical Benefit Rate (CBR) is defined as the proportion of patients with best overall response of confirmed CR or confirmed PR or SD ≥ 24 w.
- The Overall Survival (OS) is defined as the time from randomization until death from any cause. Patients with no death will be censored on the last available follow-up date.
- The duration of response (DoR) is defined as the time from documentation of first tumor response (either CR or PR) to disease progression or death due to any cause. The DoR will be calculated for the participants with unconfirmed CR or PR.

- The duration of Clinical Benefit (DoCB) is defined as the time from documentation of first unconfirmed clinical benefit (either CR or PR or SD ≥ 24 w) to disease progression or death due to any cause. The DoCB will be only calculated for the participants with unconfirmed clinical benefit.

- The Time to Progression (TTP) is defined as the time from randomization to objective tumor progression or clinical disease progression (TTP does not include deaths). Patients with no progression will be censored at the date of their last evaluable imaging previous 35 days after last dose administration date.

- The Time to Response (TTR) is defined as the time from randomization to unconfirmed ORR date. Patients without ORR will be censored at the date of their last evaluable imaging previous 35 days after last dose administration date.

4.3 Secondary Endpoints-Translational Sub-Studies

- Presence of different pattern of expression of ESR1 mutations and other CDK4/6 related biomarkers (i.e. Rb, Akt, PIK3, p53 CA cyclin D1, cyclin A2, E2F1...) in liquid biopsies and tissue samples.

- Proteomics analysis to evaluate differential pattern of protein expression from tissue samples.

- Exome and RNA sequencing will be performed in selected samples

The analyses of these endpoints will be exploratory. Therefore, no pre-specified analyses are detailed in the SAP.

4.4 Safety Endpoints

The safety endpoints will be:

- Baseline characteristics
- The extent of exposure
- Concomitant medications
- Adverse events (AEs)
- Physical examination
- Vital signs
- ECOG Performance Status
- 12-lead Electrocardiogram
- Hematology

- Biochemistry

5 STATISTICAL METHODS

5.1 Interim Analysis

One interim analysis will be performed after 35% of the total PFS events (89 events) have been observed.

Interim analysis will evaluate both safety and efficacy data (the primary end-point and all safety and efficacy secondary objectives). All them are pre-specified at section 5 statistical methods.

Efficacy and safety results of the interim analysis will be evaluated by the Steering Committee (SC) that will decide about the suitability of continuing with the study.

The trial may be stopped for superiority or non-inferiority at interim according with mentioned decision criteria (See section 1.7 Sample Size). However, the decision to stop the trial should be agreed by the SC after reviewing the interim safety and efficacy data. The trial may also stop for inferiority if palbociclib plus fulvestrant arm is significantly worse than palbociclib plus letrozole.

Interim analysis will include both safety and efficacy data to ensure that SC assessment could take into account risk vs. benefit data of the combinations used in the trial. In order to correctly balance the safety and efficacy information, it is needed to know if the combination has proven superiority prematurely, not only futility. Trial would not be suspended in circumstances only because the treatment efficacy has been early demonstrated. Efficacy data would be evaluated together with safety data by the SC.

In accordance with sample size assumptions with 89 PFS events at interim analysis the median PFS in palbociclib plus fulvestrant group should be ≥ 40 months (HR≤0.52) so that the non-inferiority hypothesis will be accepted. This is a median PFS much higher than the observed median PFS in the active treatment arms in previous studies (20 to 24 months). Thus, it is not likely that neither superiority nor non-inferiority boundaries will be crossed at interim analysis.

In addition, there are other secondary objectives in the study that would discourage early study termination. The information provided by translational sub-studies has a great interest at clinical and scientific levels. The early discontinuation of the trial would impact negatively in the sub-studies validity.

5.2 General Methodology

Definition of baseline: For each safety or efficacy parameter, the last valid assessment made before first study drug administration will be used as the baseline for all analyses of that safety or efficacy parameter unless otherwise specified.

Continuous data will be summarized in terms of the number of observations, mean, standard deviation (STD), median, minimum, maximum, and first and third quartiles, unless otherwise stated. Where data are collected over time, both the observed data and change from baseline will be summarized at each time point.

The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, and first and third quartiles will be reported to one more decimal place than the raw data recorded in the database. The STD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), frequency counts and percentages. Changes from baseline in categorical data will be summarized using shift tables where appropriate.

Percentages will be presented to one decimal place. A percentage of 100% will be reported as 100%. Percentages will not be presented for zero counts. Unless otherwise stated, percentages will be calculated using n as the denominator, for frequency tables not assessed by time point the population will be used as denominator. If sample sizes are small, the data displays will show the percentages, but any textual report (e.g. clinical study report) will describe frequencies only.

P-values greater than or equal to 0.001, in general, will be presented to three decimal places. However, if a p-value is only presented to four decimal places (by SAS) it will not be rounded again, but will be presented to four decimal places. P-values less than 0.0001 will be presented as "<0.0001".

Confidence intervals will be presented to one more decimal place than the raw data. A two-sided significance level of 5% will be used for confidence intervals.

All report outputs will be produced using SAS® version 9.4 version in a secure and validated environment. All report outputs will be provided to the Sponsor in a single Microsoft Word document.

5.3 Subject Disposition

Descriptive statistics will be provided for the following:

- Overall number of subjects in the screening population, number of screen failures, and the number of patients randomized.
- Number and percent of subjects in each of the analysis populations (Safety, ITT and PP) by treatment and center.
- Listing of subjects excluded from each of the analysis populations along with reason for exclusion.
- Listing of protocol deviations.
- Study termination:
    - Number and percent of subjects who completed the study.
    - Frequency of premature termination reasons.
    - Listing of all dropouts along with reason for termination, treatment group and time of termination.

No statistical tests are planned for these data.

5.4 Baseline Characteristics

Baseline characteristics will be provided by treatment group and overall for the ITT and PP population.

Descriptive statistics, including number of subjects in each treatment group, mean, standard deviation (STD), median and range for continuous variables and frequency and percent for categorical variables will be provided by treatment group:

- Demographic characteristics
- Medical history
- Prior concomitant medication
- Tumor history
- TNM Staging for Primary Tumor
- Primary Tumor Treatment
- Previous surgeries for breast cancer
- Metastatic Disease
- TNM Staging for Metastatic Disease
- Receptors for Metastatic Disease (optional for patients not diagnosed de novo)

- Target and Non-Target Lesions (tumor radiological assessment)
- Vital Signs
- ECOG Performance Status
- 12-lead Electrocardiogram
- Hematology
- Biochemistry

No statistical tests are planned for these data.

5.4.1 Summary by Treatment Group and Overall

A baseline global table will be generated (ITT population), showing the following demographic and baseline characteristics by treatment group and overall. This table will be used to present in a possible article. The presented data will be median (range) or n (%).

- Age (Median - Range)
- Race (Caucasian, Black, Asian, Other)
- Countries (Spain, France, UK, Italy, Russia, Germany, Czech Republic)
- ECOG (0 ,1, 2)
- Menopausal status (Pre or perimenopause, Postmenopause)
- Estrogen receptor (ER) status
- Progesterone receptor (PgR) status
- Human epidermal growth factor receptor 2 (HER2) status
- Disease stage at initial diagnostic (I, II, III, IV)
- Diagnosis of advance cancer (Local or regional, Distant)
- Disease free interval (median – range)
- Disease free interval (<12 months, 12 to 24 months, >24 months)
- De novo / non-de novo
- Measurable disease (No, Yes)
- Metastatic site (Bones, Lymph node, Lung, Liver, Brain, Pleura, Skin, Chest,…)
- Number of metastatic disease sites
- Prior chemotherapy (Adjuvant, Neoadjuvant)
- Prior endocrine treatment (yes / no)
- Neo(adjuvant) prior chemotherapy families (Nitrogen mustard analogues, Anthracyclines and related substances, Pyrimidine analogues, Taxanes, Other immunosuppressants)
- Neo(adjuvant) prior chemotherapy (Cyclophosphamide, Fluorouracil, Doxorubicin, Epirubicin, Docetaxel, Paclitaxel, Methotrexate, Capecitabine, Vinorelbine, …)

- Prior endocrine therapy (Adjuvant, Neoadjuvant)
- Neo(adjuvant) prior endocrine therapy families (Anti-estrogens, Aromatase inhibitors, Gonadotropin releasing hormone analogues, Selective estrogen receptor modulators, …)
- Neo(adjuvant) prior endocrine therapies (Tamoxifen, Letrozole, Anastrozole, Exemestane, Goserelin, Raloxifene, Toremifene, Triptorelin, …)
- Prior targeted therapy (Neoadjuvant, Adjuvant, …)
- Neo(adjuvant) prior targeted therapy
- Prior medication with analgesics or corticoids

A by-subject listing of all demographic and other baseline characteristics will be provided by treatment group.

5.5 Efficacy

The intervention arm (palbociclib in combination with fulvestrant) will be compared against the control arm (palbociclib in combination with letrozole) for all efficacy analysis.

The study is designed to have one interim analysis and the final analysis. The interim analysis and final analysis will be based on the primary endpoint. The interim analysis (See section 5.2.) will be performed after 89 PFS events have been observed.

5.5.1 Primary Efficacy Analysis

The progression-free survival will be compared between the two groups using the Kaplan-Meier method, log-rank test and multivariable Cox proportional hazards model, adjusting for site of disease (visceral vs. non-visceral), the onset of metastatic disease diagnoses (de novo metastatic vs. non-de novo patients).

The median PFS survival probability in each treatment group and the Hazard Ratio (HR), with its corresponding confidence intervals, will be calculated.

We will test the primary endpoint at a nominal level of 0.001 and 0.0498 at interim and final analysis, respectively.

In the interim analysis after 89 PFS events:

- We will declare that the PFS of palbociclib plus fulvestrant arm is superior from control group if the upper limit of 99,9% confidence interval for the hazard ratio is lower than 1.
- Otherwise, non-inferiority will be declared if the upper limit of 99,9% confidence interval for the hazard ratio is lower than non-inferiority margin of 1.21.

In the final analysis after 254 PFS events:

- We will declare that the PFS of palbociclib plus fulvestrant arm is superior from control group if the upper limit of 95% confidence interval for the hazard ratio is lower than 1.

- Otherwise, non-inferiority will be declared if the upper limit of 95% confidence interval for the hazard ratio is lower than non-inferiority margin of 1.21.

The primary efficacy analysis will be based on the ITT and PP population. The ITT population will be considered the primary population for superiority analysis, with appropriate support provided by the PP population. In non-inferiority analysis the ITT and PP population have equal importance and their use should lead to similar conclusions for a robust interpretation (1).

5.5.2 Secondary Efficacy Analysis

All secondary efficacy analysis will be based also on the ITT and PP population. The ITT population will be considered the primary population for superiority analysis.

The statistical comparison of PFS, OS, TTP, DoR and TTR between treatment groups, will be performed using the Kaplan-Meier method and Log-Rank test followed by multivariable Cox proportional hazards model for adjusting for stratified randomization variables.

The binary outcomes (ORR and CBR) will be evaluated using the Fisher's exact test followed by multivariate logistic regression for adjusting for stratified randomization variables.

The Odds Ratio (OR) and Hazard ratios (HR), with its corresponding 95% confidence intervals, will be calculated to compare dichotomous and time to event variables, respectively.

To assess the impact of potential clustering for patients cared by the same site, we will use mixed cox regression models.

No adjustment will be made for multiple testing.

5.5.3 Methods of analysis for each variable

Variable/Outcome	Hypothesis	Outcome measure	Method of analysis
1) Primary efficacy			
PFS (investigator criteria)	Intervention > control arm (ITT population)	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
	Intervention > control arm (PP population)	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
2) Secondary efficacy (ITT population)			
PFS (RECIST v1.1 criteria)	Intervention > control arm	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
ORR	Intervention > control arm	Binary	Fisher's exact test and Logistic model
CBR	Intervention > control arm	Binary	Fisher's exact test and Logistic model
OS	Intervention > control arm	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
DoR	Intervention > control arm	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
DoCB	Intervention > control arm	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
TTP	Intervention > control arm	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
TTR	Intervention > control arm	Time to event	Kaplan-Meier survival analysis (Log-Rank test and Cox model)
3) Safety (Safety Population)			
Adverse events	Intervention ≠ control arm	MedDRA (categorical)	Descriptive methods
Toxicities G3-G4	Differences according baseline characteristics	MedDRA (categorical)	Chi-squared / Mann Whitney test/ Logistic regression.
4) Sensitive analysis			
Per protocol analysis	Intervention > control arm (PP population)	All outcomes	Kaplan-Meier survival analysis (Log-Rank test)/ Chi-squared test
Adjusting for baseline and stratified randomization factors	Intervention > control arm (ITT population)	Time to event endpoints	Multivariable Cox regression models
Adjusting for baseline and stratified randomization factors	Intervention > control arm (ITT population)	Binary endpoints	Multivariable logistic regression models

5.5.4 Efficacy Endpoints by Baseline Characteristics

To assess whether the results are consistent across subgroups forest plots of HR or OR with 95% CIs will be provided. All efficacy endpoint (PFS, OS, ORR, CBR, DoR, TTP and TTR) will be compared between treatments arms (ITT population) using forest plots in each of the baseline characteristics pre-specified at section 5.5.1 and by stratified factors.

We will adjust study arms effect by stratification factors, baseline characteristic analyzed and an interaction term between study arms and baseline characteristic analyzed. The hazard ratio (time-to-event) or odds ratio (event) presented in the table should correspond with the exponential coefficient of study arm term in the model. There will be one regression model by baseline characteristic analyzed, however we will present one HR/OR for each variable category. We will obtain the HR/OR in each specific category of the same baseline characteristic codifying this category as reference (value 0) in the regression model. The p-value will correspond with the p-value of the interaction term. There will be only one p-value by baseline characteristic. Neo(adjuvant) prior chemotherapies families should be codified as four different baseline characteristics (yes, no). Neo(adjuvant) prior endocrine families should be codified as two different baseline characteristics (yes, no).

5.6 Translational Sub-Studies

These statistical analyses will be exploratory. Therefore, no pre-specified analyses are detailed in the SAP. All secondary analysis will be based also on the exploratory evaluable population.

The objective of the statistical analyses of biomarkers is the identification of those markers or combinations of markers which show best association with positive or negative clinical outcome of palbociclib plus fulvestrant or letrozole.

The biomarker analyses will be exploratory. They aim at exploring the potential to predict clinical benefit, by each marker separately and/or by suitable combinations.

Further data on markers will be analyzed, dependent on its availability. According to experience many biomarkers show a skewed statistical distribution across subjects and within subject. Appropriate transformations will be applied to transform these measurements into distributions with an approximate Gaussian shape. These transformations do not change the order of the values, such that non-parametric analyses based on ranks or cut-offs remain unchanged by the transformation. The basic statistics and interdependencies of the different markers will be descriptively investigated.

Markers will be evaluated on a univariate level regarding their potential for prediction (e.g. search or adaptation of cut-offs) of the clinical endpoints. Further multivariate techniques (e.g. Multiple Logistic Regression, Principal Component Analysis with Rotation, Cluster Analysis) will be employed in order to study combinations of markers. Biomarker and

Response correlations with clinical covariates will be investigated. It will be checked whether covariates can improve the prediction and whether there is an interaction with the biomarkers. Relevant covariates could become a part of the statistical prediction model. Candidate groupings derived from biomarkers will be checked with time to event variables (e.g. Kaplan-Meier curves, Cox proportional hazard model, log-rank test).

5.7 Safety

All safety tables will list or summarize subjects by treatment group on the entire Safety Population. These safety assessments will be subjected to clinical review and summarized by appropriate descriptive statistics.

5.7.1 Duration and Extent of Exposure

Extent of Exposure will be based on the ITT, PP and safety population

- b: "Actual Cycle Duration" is the treatment duration for a cycle per CRF. It is the length of time (days) between actual and next cycle start date dose. At the last cycle is the difference between start and stop date dose.
- c: "Actual Cycle Dose Days" is the number of days with dose administration in the cycle, considering the interruptions.
- d: "Actual Total Dose per Cycle" is the total dose a patient actually took in a cycle, considering interruptions and reductions.
- e: "Intended Daily Dose per Cycle" is equal to 125 mg for Palbociclib, 2.5 mg for Letrozole, and 1000 mg for Fulvestrant at cycle 1 and 500 mg at other cycles.
- f: "Intended Cycle Duration" is equal to 28 days, except for the last cycle that is the minimum of 28 and Actual Cycle Duration.
- g: "Intended Cycle Dose Days for Palbociclib" is equal to 21 days, except for the last cycle that is the minimum of 21 and Actual Cycle Duration. "Intended Cycle Dose Days for Letrozol" is equal to 28 days, except for the last cycle that is the minimum of 28 and Actual Cycle Duration. "Intended Cycle Dose Days for Fulvestrant" is equal to 2 days at cycle 1, and 1 day for all other cycles, except for the last cycle that is the minimum of 2 and Actual Cycle Duration at cycle 1, and the minimum of 1 and Actual Cycle Duration for all other cycles.
- A: "Total number of cycles".
- B: "Treatment Duration" = Sum over all cycles of (b).
- C: "Days on drug" = Sum over all cycles of (c).
- D: "Total Actual Total Dose" = Sum over all cycles of (d).
- E: "Mean Intended Daily Dose" = Mean over all cycles of (e).

- F: "Total Intended Duration" = Sum over all cycles of (f).
- G: "Total Intended Dose Days" = Sum over all cycles of (g).
- H: "Intended Total Dose" = G*E
- I: "Actual Average Daily Dose on Dose Days" = D/C
- J: "Ratio For Dose Interruption" = C/G
- K: "Ratio For Cycle Duration" = F/B
- L: "Actual Average Daily Dose Intensity"= I*J*K
- M: "Relative Dose Intensity (RDI)" = L/E*100

The treatment duration (days), days on Drug and Treatment compliance (%) will be summarized in terms of the number of observations, mean, standard deviation (STD), median, minimum and maximum, to each treatment and both arms.

Extent of exposure measured as RDI in different treatments and both arms will be described with median, interquartile range (IQR) and range. The RDI will be dichotomized in different cutoffs (≥50%, ≥70%, ≥80%, ≥90%, ≥100%) and described with frequencies and percentages. The treatments RDI will be correlated with best overall tumor assessment, PFS and OS based on ITT and PP populations. The correlation between RDI and best overall response will be described with 95%CI spearman correlation test and scatter plots. The correlation between RDI dichotomized and time-to-event outcomes (PFS, OS) will be analyzed with Kaplan-meyer curves and log-rank test.

5.7.2 Concomitant Medications

The number and percent of unique patients taking concomitant medications will be summarized by therapeutic classification, coded term and treatment group. Elective surgeries/procedures performed during the study will be presented in a listing.

The following are conventions that will be used to classify individual medications as prior and/or concomitant:

- Medications with stop dates prior to randomization will be considered prior.
- Medications with missing stop dates or stop dates the day of or after randomization will be considered concomitant, regardless of start date. Additionally, if the start date is prior to randomization or missing, the medication will also be considered prior.

Frequencies and by-subject listing of all prior and concomitant medications will be provided, containing variables listed on Prior/Concomitant Assessment eCRF, their corresponding categories (Prior or Concomitant), and WHO Anatomical Therapeutic Chemical (ATC) level 2 and preferred term if applicable.

5.7.3 Adverse Events

All AEs will be recorded on the eCRF "Adverse Events" page and will be coded using the current version of MedDRA® to give a system organ class (SOC) and preferred term (PT) for each event. All adverse event safety data will be updated to the version of MedDRA that is current at the time of the database lock and statistical analyses. Adverse events will be coded with grades defined according to CTCAE V4.0 criteria.

Treatment-emergent AEs (i.e. those events occur after the first study medication administration and were not present at baseline or worsened in severity following the start of treatment) will be tabulated. The TEAE will be tabulated according to intensity and causality. If intensity of an AE or causality of an AE to the study medication is missing, a worst-case scenario will prevail (severe in intensity or probably related will be assumed). In the summary tables the number of subjects with events and the number of events will be presented.

The onset date of an AE will be compared to the date of first dose of study drug to determine whether or not the AE is treatment-emergent. Adverse events with an onset date on or after the date of first dose of study drug will be classified as treatment-emergent.

All deaths and SAEs, regardless of cause, from treatment start until 28 days after final dose of treatment. Non-fatal AEs occurring after treatment start regardless of cause, up until 28 days after final dose of treatment or until start of new anti-cancer treatment, whichever is first. Disease progression is not considered a treatment emergent adverse event unless the patient dies of disease prior to 28 days after discontinuation of treatment. Events that are continuations of baseline abnormalities are considered treatment emergent adverse events only if there is an increase in grade over baseline, or if there is an increase following a decrease during the study.

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

The following summaries will be provided:

- An overview of adverse events by treatment group (number of subjects with at least one AEs, number of subjects with at least one TEAE, number of subjects with serious TEAE, number of subjects with non-serious TEAE, number of deaths, number of subjects with TEAE leading to discontinuation of study treatment, number of subjects dropped out due to AE).
- A summary of the number and percentage of subjects reporting a treatment-emergent adverse event by treatment group, SOC, and PT.

- A summary of the number and percentage of subjects reporting a treatment-emergent adverse event related to study drug by treatment group, SOC, and PT.
- A summary of the number and percentage of subjects reporting a treatment-emergent adverse event by treatment group, by maximum intensity, SOC and PT.
- A summary of the number and percentage of subjects reporting a serious treatment-emergent adverse event, by treatment group, SOC and PT.
- A summary of the number and percentage of subjects reporting a treatment-emergent adverse event resulting in death during the study, by treatment group, SOC and PT.
- A summary of the number and percentage of subjects with adverse events leading to discontinuation of study drug, by treatment group, SOC and PT.

For adverse events, we will report intensity, casualty, relation, body system, action taken, and outcome.

Serious adverse events, deaths and study discontinuations will be described and examined in each study group.

Analysis of safety-related data will be considered at four levels:

- First, the extent of exposure (dose, duration, number of patients) will be examined to determine the degree to which safety can be assessed from the study.
- Second, we will describe and compare clinically relevant test, concomitant medications and adverse events reported in every study group. For adverse events, we will report intensity, causality, body system, action taken, and outcome.
- Third, serious adverse events, deaths and study discontinuations will be described and examined in every study group.
- Finally, patient grade 3 and 4 toxicities in every study group will be classified by MedDRA system organ class and compared between patient baseline characteristics.

The relation between baseline characteristics and serious adverse events (classified in MedDRA SOCs) will be analyzed with chi-squared test followed by multivariate logistic regression with appropriate interaction terms (baseline characteristic × treatment group).

The occurrence and maximal grade of toxicity for the whole duration of treatment will be listed and tabulated by type and dose level. Adverse events reported as non-drug related by the responsible investigator will be reported as well.

5.7.4 Clinical Laboratory Parameters

All hematology and biochemistry parameters will be presented by descriptive statistics in a tabulated summary by time point of assessment per treatment group together with the respective changes from baseline. In addition, a frequency table for clinically significant values will be presented by time point of assessment.

A by-subject listing for hematology and clinical chemistry will be provided. These listings will be presented by treatment group and time point and will include: center, subject identifier, laboratory parameter, parameter values (in SI units), SI unit, normal range and a flag with respect to normal range (below, within and above normal range).

5.7.5 Vital Signs

Weight, systolic and diastolic blood pressure, heart rate and respiratory rate will be presented by descriptive statistics in a tabulated summary by time point of assessment per treatment group together with the respective changes from baseline. In addition, frequency tables for the number of patients with increases or decreases from baseline in systolic/diastolic blood pressure of >20 mmHg and pulse rate of >15 bpm will be provided by time point of assessment and overall. A by-subject listing for all vital signs per treatment group and time point will be provided.

5.7.6 Physical Examination

A frequency table per treatment group, time point and body system will be provided for assessment results of normal, abnormal and not done.

A by-subject listing for all body systems per treatment group and time point will be provided. Only subjects with at least one abnormal finding will be included in this listing.

5.8 Handling of Missing Data

The analysis of the efficacy endpoints (progression and death) will be based on a log-rank or Cox regression tests and, therefore, not affected by patient withdrawals (as they will be censored) provided that dropping out is unrelated to prognosis.

Patients with missing information in other outcomes, such as clinical benefit rate or overall response, will be considered as no responders. Furthermore, we will report reasons for withdrawal for each randomization group and compare the reasons qualitatively.

5.9 Deviations from SAP

Any deviations from the original statistical plan will be described and justified in the final clinical study report.

6 APPENDICES

6.1 Appendix 1 – Primary Efficacy Analysis SAS Codes

Kaplan-Meier method and log-rank test.

```
proc lifetest data=ITT;
     time PFS_TIME*PFS(0);
     strata RAND_TREATMENT;
run;
```

Multivariable Cox proportional hazards model, adjusting for site of disease and the onset of metastatic disease diagnoses.

```
proc phreg data=ITT;
     class RAND_TREATMENT VISCERAL DENOVO;
     model PFS_TIME*PFS(0)=RAND_TREATMENT VISCERAL DENOVO;
     hazardratio RAND_TREATMENT;
run;
```

6.2 List of Tables, Listings, Figures

A complete list of tables, listings and figures (TLFs) will be given in a separate document which can be updated without updating the SAP. The list will serve as a reference for both the Sponsor, the trial statistician and the statistical programmer and includes the totality of statistical output to be produced.

All output will be headed with an appropriate heading specifying the study ID and abbreviated study title.

All output will be dated and have page numbers in the form 'Page [x / y]' where x denotes the current page within an output and y the total number of pages of that output.

All statistical output will identify the underlying analysis populations and indicate the number of patients/events in this population (N) and the number of patient/events actual contributing to the output (n). All statistical output will be presented per treatment (if applicable).

All patient listings will contain additionally to the patient identification the analysis population and the treatment.

7 BIBLIOGRAPHY

1. EMEA. Points to consider on switching between superiority and non-inferiority [Internet]. 2000. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003658.pdf

2. FDA. FDA GUIDANCE: Non-Inferiority Clinical Trials to Establish Effectiveness [Internet]. 2016. Available from: https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf

3. Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. Biometrics. 1986 Sep;42(3):507–19.

4. Finn RS, Crown JP, Lang I, Boer K, Bondarenko IM, Kulyk SO, et al. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. The Lancet Oncology. 2015 Jan 1;16(1):25–35.

5. Finn RS, Martin M, Rugo HS, Jones S, Im S-A, Gelmon K, et al. Palbociclib and Letrozole in Advanced Breast Cancer. N Engl J Med. 2016 17;375(20):1925–36.

6. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007 Jun 7;8:16.