STATISTICAL ANALYSIS PLAN

Phase III study evaluating palbociclib (PD-0332991), a Cyclin-Dependent Kinase (CDK) 4/6 Inhibitor in patients with hormone-receptor-positive, HER2-normal primary breast cancer with high relapse risk after neoadjuvant chemotherapy

"PENELOPE^{B}"

GBG-78 - BIG 1-13 - NSABP-B-54-I

EudraCT No.: 2013-001040-62

Version:	**5.0 final**
Date:	**September, 21st 2020**
Biostatistician:	[redacted]

Protocol number: GBG-78 - BIG 1-13 - NSABP-B-54-I

Protocol title: Phase III study evaluating palbociclib (PD-0332991), a Cyclin-Dependent Kinase (CDK) 4/6 Inhibitor in patients with hormone-receptor-positive, HER2-normal primary breast cancer with high relapse risk after neoadjuvant chemotherapy

[redacted]

Version: **5.0 final**

Version Date: **September, 21st 2020**

[redacted]	Date
[redacted]	Date
[redacted]	Date

1. TABLE OF CONTENTS

2. ABBREVIATIONS

ATDI	Actual total dose intensity
AE	Adverse event
AI	aromatase inhibitor
AESI	Adverse Events of special interest
ALAT	Alanine aminotransferase
ASAT	Aspartate aminotransferase
CHW	Cui, Hung and Wang method
CI	Confidence interval
CP	Conditional power
CPS-EG	Clinical-pathological stage – estrogen/grade
CRF	Case report form
CTC	Common Terminology Criteria
(NCI-)CTCAE	(National Cancer Institute) Common Terminology Criteria for Adverse Events
DMP	Data management plan
EIA	Efficacy Interim Analysis
EP	Evaluable Patients
EOT	End of therapy
EORTC	European Organisation for Research and Treatment of Cancer
ER	Estrogen receptor status
ET	Endocrine treatment / therapy
FDA	Food and Drug Administration
GBG	German Breast Group
HR	Hazard ratio
iDFS	Invasive Disease Free Survival
ITT	Intent-to-treat
LHRH	Luteinizing Hormone Releasing Hormone
LRRFI	Loco-regional relapse-free interval
[redacted]	[redacted]
MedDRA	Medical Dictionary for Regulatory Activities
n.a.	not applicable
NACT	neo-adjuvant chemotherapy
NYHA	New-York Heart Association
QoL	Quality of Life
OS	Overall Survival
PP	Per protocol
RTD	relative total dose
RDTI	relative total dose intensity
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Statistical Analysis Systems
SOC	System Organ Class
SOP	Standard Operating Procedure

SSR	Sample Size Re-Estimation
t.b.d.	to be determined

Table of abbreviation will be updated upon finalization of statistical report.

3. INTRODUCTION

The PENELOPE^{B} study is a prospective, international, multicenter, randomized, double-blinded, placebo-controlled, parallel-group Phase III study comparing the efficacy and the safety of thirteen cycles adjuvant treatment with palbociclib versus placebo in high risk patients (CPS-EG score >= 3 or (CPS-EG score = 2 and ypN+)) without pathological complete response after neoadjuvant chemotherapy for hormone-receptor-positive / HER2-normal primary breast cancer. Patients will receive standard adjuvant endocrine treatment after completion of adequate local surgical and radio therapeutic treatment.

There will be two interim analyses to monitor futility and to test for overwhelming efficacy and the study is designed for sample size adaptation. The final analysis will be performed to test for differences between the treatment arms in the primary endpoint invasive disease-free survival (iDFS).

This SAP describes the statistical procedures required within the group-sequential, adaptive design for sample size adaptation. The planned safety and compliance analyses will be described. Templates for tables and figures will be given.

Details on the cost analyses and health economic evaluation, on the analysis of biomarkers, and on the analysis of pharmacokinetics will be described in separate analyses plans.

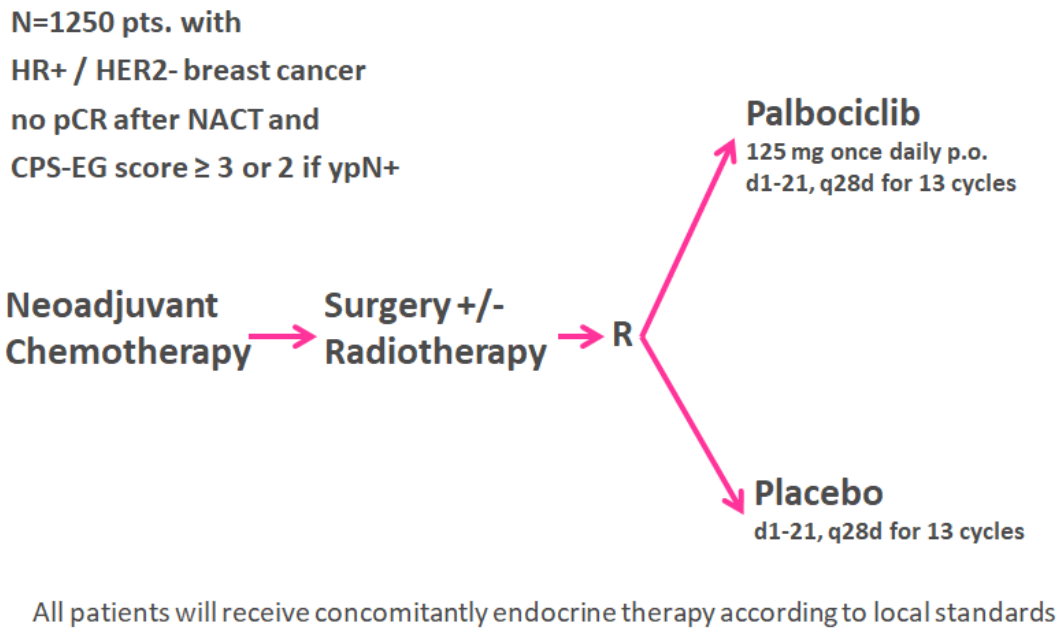
4. STUDY DESIGN

4.1 Overall design and treatment

The PENELOPE^{B} study is a prospective, international, multicenter, randomized, double-blinded, placebo-controlled, parallel-group Phase III study comparing the efficacy and the safety of thirteen cycles (1 year) adjuvant treatment with palbociclib versus placebo in high risk (either CPS-EG score >= 3 or CPS-EG score = 2 and ypN+) patients without pathological complete response after neoadjuvant chemotherapy for hormone-receptor-positive / HER2-normal primary breast cancer. Patients will receive standard adjuvant endocrine treatment after completion of adequate local surgical and radio therapeutic treatment.

The study has an adaptive design with two interim efficacy analyses to allow for sample size adaptation and to assess safety, including unexpected toxicity. The trial is designed to stop due to futility in each interim analysis and due to efficacy in the second interim analysis. The adaptive design procedure will be described in detail in section 7.

Figure 1: Penelope^{B} Study Design



4.2 Randomization/stratification

Patients will be randomized in a 1:1 ratio to:

- Arm A: Palbociclib at a dose of 125 mg once daily, Day 1 to Day 21 followed by 7 days off treatment in a 28-day cycle for thirteen cycles;
- Arm B: Placebo once daily Day 1 to Day 21 followed by 7 days off treatment in a 28-day cycle for thirteen cycles.

Randomization will be stratified using block randomization by:

- histological lymph node status at surgery (ypN 0-1 vs ypN2-3)
- age at first diagnosis (<=50 vs >50 yrs.)
- Ki-67 value (>15% vs <=15%)
- global region of participating site (Asian vs non Asian)
- risk status (CPS-EG Score >=3 vs CPS-EG Score=2 and ypN+)

Due to the prognostic impact of the randomization criterion risk status, randomization to the stratum CPS-EG score=2 and ypN+ will be limited to 50% of the patients.

4.3 Sample size calculation

Sample size was calculated based on the analysis of the primary endpoint, iDFS.

During the study the sample size estimation was amended twice.

Original protocol version until and including protocol version C8

For the placebo plus endocrine agent arm a 3-year iDFS rate of 0.72 was assumed based on an analysis of the GBG meta-database for patients with CPS-EG score ≥3. For the palbociclib plus endocrine agent arm a 3-year iDFS rate of 0.80 was assumed. This corresponds to a palbociclib/placebo HR=0.67. To detect a clinical relevant difference of 0.08 with a power of 85% a minimal number of 233 events for iDFS endpoint are required using a two-sided stratified log-rank test at an overall two-sided significance level of 0.05. Additional to the above assumptions, an exponential distribution with a common hazard rate (0.04383 subjects per year) for both arms was assumed for time to drop-out. Applying a 1:1 randomization and a non-uniform enrollment rate of 25 subjects per month at the peak (the peak enrollment rate is estimated to be reached 1 year after enrollment start), 800 patients will be enrolled within 36 months. The maximum study duration is estimated to be 5.9 years. A median follow up time at the final analysis on iDFS is estimated to be around 50 months.

Power calculation is also done for the endpoint of iDFS excluding second primary of non-breast cancer. Assuming that not more than 15% of iDFS event are second primary of

non-breast cancer, there is 79% power to detect statistical significance between the 2 arms with a 2-sided significance level of 0.05.

Protocol version D9 and E10

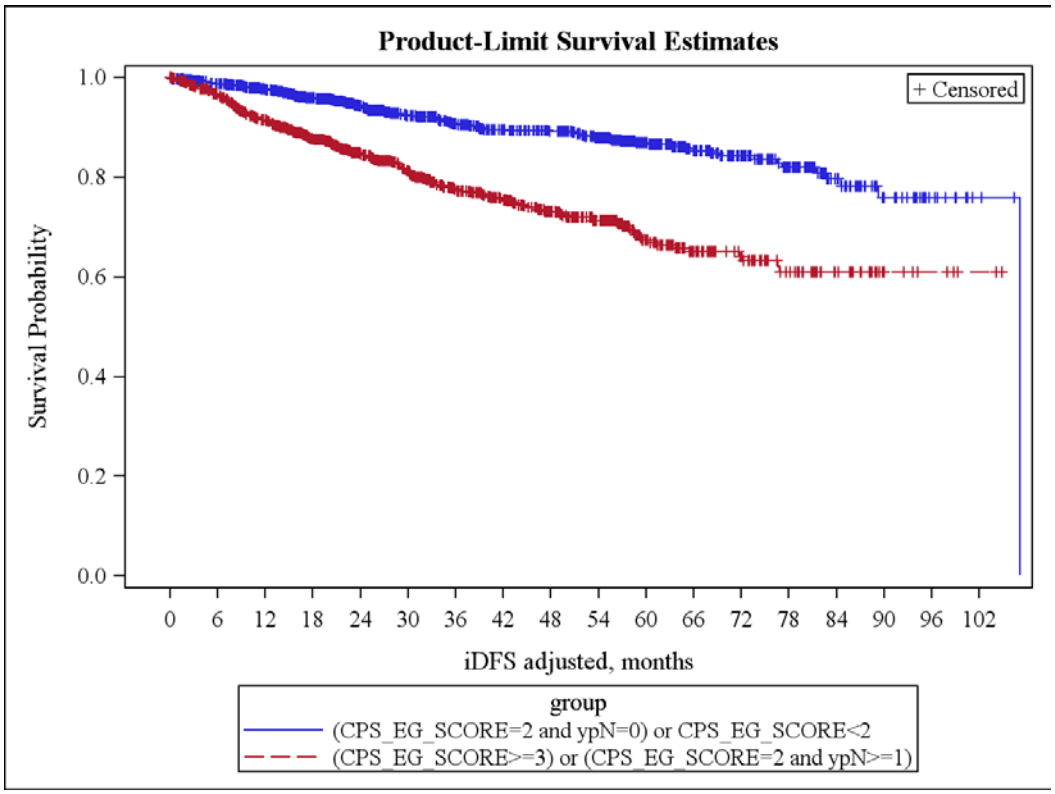
For the arm placebo plus endocrine agent a 3-year iDFS rate of 0.77 was assumed based on an analysis of the GBG meta-database for patients with CPS-EG score ≥3 or Score 2 and ypN+ disease. For the palbociclib plus endocrine agent arm a 3-year iDFS rate of 0.836 was assumed. This corresponds to a palbociclib/placebo HR=0.685. To detect this HR with a power of 85% a minimal number of 255 events for iDFS endpoint are required using a two-sided stratified log-rank test at an overall two-sided significance level of 0.05. Additional to the above assumptions, an exponential distribution with a common hazard rate (0.04383 subjects per year) for both arms was assumed for time to drop-out. Applying a 1:1 randomization and a non-uniform enrollment rate of 40 subjects per month at the peak, it was originally estimated that 1100 subjects will need to be enrolled.

Rationale of assumptions on sample size determination

The clinical-pathologic stage – estrogen/grade (CPS-EG) (1,2) combining clinical stage before neoadjuvant treatment, pathological stage after neoadjuvant treatment, grading and estrogen-receptor status can be used to identify high-risk patients. The score was further validated on 2454 patients with HR-positive/HER2-normal tumors from the German neoadjuvant studies' meta-database (3).

Estimation of 3-year iDFS rate was performed on the basis of a meta-database containing 2659 HR+/HER2- patients who received neoadjuvant therapy (about 18-24 weeks), surgery (average period of around 6 weeks after last cycle of chemotherapy) and radiation (about 6 weeks). In the PENELOPE^{B} study patients will be randomized after radiation or rehabilitation, if applicable, which is estimated to be approximately 9 months after start of neoadjuvant chemotherapy. Given these assumptions a 3-year iDFS rate of 0.77 was estimated for the control arm of patients with CPS-EG score >= 3 or CPS-EG score = 2 and ypN+. The assumption of exponential distribution on time-to-iDFS event for the control arm may affect the duration of study. This assumption cannot be assessed due to lack of patient-level data on surgery date.

Figure 2: Survival after completion of radiotherapy (iDFS adjusted) in 2659 patients with HR-positive/HER2-normal tumors from the German neoadjuvant studies' meta-database by CPS-EG score. The lower curve includes patients with 3 or more score points or score 2 and ypN+ status and the upper curve patients with less than 2 score points or score 2 and ypN.



Protocol version G11

The first efficacy interim analysis in April 2017 resulted in an increase of both sample size (n=1250) and number of events (290 events) based on calculated HR and a pre-defined threshold for conditional power. The sample size of 1250 patients is included in study protocol G11. The final analysis will take place when 290 iDFS events are observed which is estimated to occur about 6.5 years after first patient randomized (for details please refer to section 7.1).

Power calculation is also done for the secondary endpoint of iDFS excluding second primary of non-breast cancer. Assuming that not more than 15% of iDFS events are second primary of non-breast cancer, there is 79% power to detect statistical significance between the 2 arms with a 2-sided significance level of 0.05.

5. STUDY OBJECTIVES, ENDPOINTS AND COVARIATES

5.1 Objectives and endpoints

5.1.1 Primary efficacy

The primary objective of the study is to compare invasive disease-free survival (iDFS) for palbociclib vs. placebo in patients with residual invasive breast cancer and high CPS-EG score - after neoadjuvant chemotherapy receiving standard adjuvant endocrine therapy for HR-positive/HER2-normal primary breast cancer.

The primary endpoint is invasive disease-free survival (iDFS) and is defined according to Hudis (4) as the time in months between randomization and first event. The following events will be considered as first events: ipsi- or contralateral invasive in-breast or loco-regional recurrence, distant recurrence, death from breast cancer, death from non-breast cancer cause, death from unknown cause, invasive contralateral breast cancer, or second primary invasive cancer (non-breast).

Events are documented on the relapse form (RF) or on the death report form (DRF). The following dates of events are available on RF and DRF:

- date of loco-regional recurrence (for ipsi- or contralateral invasive in-breast or loco-regional recurrence)
- date of distant recurrence
- date of second primary invasive cancer (non-breast)
- date of death

The iDFS time of patients ***with iDFS event*** is calculated as the interval between date of randomization and the first date of event.

iDFS (in months) will be calculated as (first event date − randomization date +1)/30.4.

The cut-off date for the analysis of iDFS is event-driven. After the first efficacy interim analysis the number of events was increased and now the final analysis will be performed after 290 events have been collected.

The iDFS time of patients without iDFS event will be censored at the date of last disease assessment (where the patient was known to be alive and tumor-free/without new signs of disease). For patients who discontinued study treatment and have no follow-up visit, the maximum of the examination dates on [redacted] without significant findings and the examination dates on [redacted] will be used as date of last disease assessment. For patients with follow-up visit, the date of last disease assessment is documented on Follow up form FU01.

- If the patient status was “alive and tumor-free/without new signs of disease” in the FU assessment, the date of follow-up assessment will be used as censoring date.

- If the patient status was “lost to follow-up” in the FU assessment, the date of follow-up assessment of a previous FU assessment will be used as censoring date where the patient was known to be alive and tumor-free/without new signs of disease.
- If the patient status was “Withdrawal of informed consent” in the FU assessment, the date of follow-up assessment of a previous FU assessment will be used as censoring date where the patient was known to be alive and tumor-free/without new signs of disease.

Patients who were randomized but had disease relapse prior to randomization will have their iDFS times censored at study Day 1 (i.e., randomization). Patients alive who do not have post baseline contact will have their iDFS times censored at Day 1 (i.e., randomization). The main iDFS analysis will be performed regardless of the initiation of new anticancer therapies, i.e., patients receiving further anticancer therapy before disease relapse or occurrence of second primary invasive non-breast cancers or death will be assigned as an iDFS events at the first date of relapse or occurrence of a second primary cancer (non-breast) or death. Sensitivity analyses will be performed as described in section 10.7.5.

5.1.2 Key Secondary

The key secondary objective is to compare iDFS excluding second primary invasive non-breast cancers for palbociclib vs. placebo in patients with residual invasive breast cancer and high CPS-EG score - after neoadjuvant chemotherapy receiving standard adjuvant endocrine therapy for HR-positive/HER2-normal primary breast cancer.

The key secondary endpoint is iDFS excluding second primary invasive non-breast cancers and is defined as the time in months between randomization and first event. The following events will be considered as first events: ipsi- or contralateral invasive in-breast or loco-regional recurrence, distant recurrence, death from breast cancer, death from non-breast cancer cause, death from unknown cause, or invasive contralateral breast cancer.

Patients who were randomized but had disease relapse prior to randomization will have their iDFS times censored at study Day 1 (i.e., randomization). Patients alive who do not have post baseline contact will have their iDFS times censored at Day 1 (i.e., randomization). The main analysis of iDFS excluding second primary invasive non-breast cancers will be performed regardless of the initiation of new anticancer therapies, i.e., patients receiving further anticancer therapy before disease relapse or occurrence of a second primary invasive non-breast cancers or death will be assigned as an iDFS events at the first date of relapse or death. Two sensitivity analyses will be performed (D and E as described in section 10.7.5).

5.1.3 Secondary efficacy

5.1.3.1 Overall survival (OS)

Overall survival (OS) is defined as the time period in months between randomization and death due to any cause. OS (in months) is calculated as (date of death − randomization date +1)/30.4. Patients alive will be censored at the date of the last contact. Patients lacking survival data beyond randomization will have their OS times censored at study Day 1 (i.e., randomization).

An interim OS analysis will be conducted at the time of final iDFS analysis. Final survival will be analyzed 3 years after the final analysis of iDFS.

5.1.3.2 Locoregional recurrences-free interval

This endpoint is described in the study protocol as locoregional recurrences-free **survival** (LRRFS) including death as event. However, locoregional recurrences, death and distant recurrences are events which prevent each other from happening, i.e. are competing. This competing risk situation will be handled appropriately by reformulating the endpoint LRRFS into locoregional recurrences-free **interval** (LRRFI) which is defined as the time period in months between randomization and diagnosis of any stand-alone loco-regional (ipsilateral breast (invasive or DCIS), local/regional lymph nodes) recurrence of disease, or any stand-alone invasive contralateral breast cancer, whichever occurs first. Distant recurrence, second primary invasive non-breast cancers and death are considered competing events. In case loco-regional recurrence of disease or invasive contralateral breast cancer are documented on the same date as a competing event, it will be considered the respective competing event in the analysis. If more than one competing event is documented on the same date as an event of interest, the following hierarchy applies: 1) distant recurrence, 2) second primary invasive non-breast cancers and 3) death.

Locoregional events are documented on the relapse form (RF) with date of loco-regional recurrence (for ipsi- or contralateral invasive in-breast or loco-regional recurrence). LRRFI (in months) will be calculated as (first LRRFI event date − randomization date +1)/30.4.

LRRFI time of patients without LRRFI event will be censored at the date of last disease assessment. Patients alive who do not have post baseline disease assessment will have their LRRFI times censored at Day 1 (i.e., randomization). The LRRFI analysis will be performed regardless of the initiation of new anticancer therapies.

LRRFI will be analyzed at the time of final iDFS analysis.

5.1.3.3 Distant disease free survival

Distant disease free survival (DDFS) is defined as the time period between randomization and diagnosis of any distant recurrence of disease, any second primary invasive cancer

(non-breast) or death due to any cause, whichever occurs first. Distant disease events are documented on the relapse form (RF) with date of distant recurrence. DDFS (in months) will be calculated as (first DDFS event date − randomization date +1)/30.4. Distant disease events will be considered in the analysis regardless of loco-regional recurrence being documented on the same date.

The DDFS time of patients without DDFS event will be censored at the date of last disease assessment. Patients alive who do not have post baseline disease assessment will have their DDFS times censored at Day 1 (i.e., randomization). The DDFS analysis will be performed regardless of the initiation of new anticancer therapies.

DDFS will be analyzed at the time of final iDFS analysis.

5.1.3.4 iDFS in patients with luminal-B tumors

Analysis and censoring of iDFS in the subgroup of patients with luminal-B tumors will be performed similar to the analysis of the primary endpoint. However, no sensitivity analysis will be performed. This subgroup was initially planned to be determined by PAM50 (cut-point t.b.d.) or any other commercially available test (t.b.d.) before unblinding data. In the translational research committee (TRC) meeting during SABCS 2018 it was decided to use HTG-Seq due to a broader panel with 2000 genes and less consumption of tissue. The assessment was performed in a laboratory which had no access to the treatment allocation of the patients and was provided to GBG in June 2020. Count data from the HTG measurements of tissue from post-neoadjuvant residual invasive disease of the breast were used to calculate the AIMS (Absolute Assignment of Breast Cancer Intrinsic Molecular Subtype) subtype. The AIMS variable is categorical with five categories: “Basal”, “Her2”, “LumA”, “LumB”, and “Normal”. The subtype of highest posterior probability was assigned. Patients with the AIMS variable “LumB” should constitute the subgroup of interest. It turned out that tissue from post-neoadjuvant residual invasive disease of the breast of 906 patients was available and that 64 (7%) of these patients were “LumB”. The pre-planned secondary endpoint iDFS in patients with luminal-B tumors can not be analyzed due to this small subgroup. Further exploratory analysis will be performed which will be described in a separate biomarker SAP.

5.1.4 Safety

Safety will be assessed on the basis of any grade 1-4 adverse event (AE) including adverse events of special interest and serious adverse events (including AEs of grade 5). AEs are documented by type and severity as graded by National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) v.4.0; seriousness, and relationship to study drug (palbociclib/placebo).

An AE is defined as any untoward medical occurrence in a patient administered a medicinal product which does not necessarily have a causal relationship with this

treatment. It also includes any undesirable clinical or laboratory changes which do not commonly occur in the patient.

Unexpected serious adverse reactions (SUSAR) will be reported as unexpected toxicities.

Adverse Events of special interest (AESI) are defined as a) drug-induced liver injury, b) pregnancy, and c) overdose with IMP, defined as the intake of more than 23 capsules during a cycle (e.g. more than two capsules on more than two days of a cycle, more than two capsules during the week of break) and in addition more than 2 capsules on the same day. Non-relevant overdoses with IMP (defined as intake of 22 or 23 capsules in the same cycle) were considered as AESI until and including protocol version D9. Treatment modification due to an AESI will be documented (no action taken, dose delay, dose reduction, interruption, permanent discontinuation of treatment).

Documentation of AEs by the investigator are scheduled for Day 1 and 14 of cycles 1 and 2, on Day 1 of each of the following cycles, and at EOT. Moreover, the investigator will document an AE whenever he/she finds out about an occurrence.

Hematological parameters will be documented on Day 1 and 14 of cycles 1 and 2, on Day 1 of each of the following cycles, and at EOT, or if clinically indicated. Blood chemistry and liver function tests will be performed on Day 1 of cycles 1, 2, 3, 7, 11, and at EOT. Post-baseline blood chemistry values for ASAT, ALAT, bilirubin, alkaline phosphatases, serum creatinine, and serum albumin are not recorded on the CRF but corresponding CTCAE grades will be collected.

HbA1c clinical abnormality (yes/no) will be collected and documented as glucose intolerance on the [redacted] in case of an abnormal finding.

Predefined AEs will be documented on the following CRF [redacted]. [redacted] is not designed for documentation of multiple infections during one cycle. In such cases, the [redacted] will be used, on which in general AEs additional to the predefined AEs will be documented as free text AEs.

Death will be reported on [redacted] with cause of death.

5.1.5 Compliance

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

In this study, study drug refers to palbociclib or placebo and study treatment refers to one of the following:

- Arm A: Palbociclib at a dose of 125 mg once daily, Day 1 to Day 21 followed by 7 days off treatment in a 28-day cycle for thirteen cycles

- Arm B: Placebo once daily Day 1 to Day 21 followed by 7 days off treatment in a 28-day cycle for thirteen cycles

5.1.5.1 Extent of exposure

The **duration of** palbociclib/placebo **exposure** is defined as: last dose date – first dose date + 7 days regardless of unplanned intermittent discontinuations.

The **total number of cycles** includes all cycles including the last cycle in which doses were given even if the cycle was not completed.

The **cumulative dose** (mg) is defined as the sum of all doses from cycle 1 to and including the last cycle, where the last cycle is based on the investigator's report in the CRF. Overdoses are documented on the SAE form. The information concerning excessive doses will be included in the calculation of the cumulative dose.

RTDI is the total dose intensity within the entire treatment achieved by a patient relative to intended dose intensity based on the planned schedule of the treatment. RTDI will be calculated according to the following step-by-step algorithm:

Planned dose (PnD) is the amount of a drug planned to be given in a cycle. PnD is calculated separately for each component of a therapy.

Planned total dose (PTD) is the planned cumulative dose over the entire treatment duration; i.e. if a patient was planned to receive *n* cycles of drug, then sum the total amount of drug planned during those *n* cycles.

i.e. $$PTD(\text{mg}) = \sum_{i=1}^{n} PD_i$$

Planned total dose intensity (PTDI) is the planned average dose intensity over the entire treatment duration (described as planned dose intensity in the study protocol),

i.e. $$PTDI(\text{mg/week}) = \frac{PTD}{planned\ \ duration\ \ of\ \ therapy\ (weeks)}$$

Actual dose (AD) is the total amount of drug that the patient has received over one cycle.

Actual total dose (ATD) is the cumulative dose of the drug that has been given over the treatment duration of *n* cycles,

i.e. $$ATD\ (\text{mg}) = \sum_{i=1}^{n} AD_i$$

ATD is similar to the **cumulative dose.**

Actual total dose intensity (ATDI) is defined as the dose intensity of what has actually been administered over the *n* cycles (described as ADI in the study protocol).

i.e. $$ATDI(\text{mg/week}) = \frac{ATD}{duration\ of\ therapy\ (weeks)}$$

If a patient has not received some drug at all, ATDI for this drug is considered to be 0.

Relative Total Dose (RTD) is the ratio of ATD and PTD, expressed as a percentage.

i.e. $$RTD(\%) = \frac{ATD}{PTD} \times 100$$

Relative Total Dose Intensity (RTDI) is the ratio of ATDI and PTDI, expressed as a percentage (described as RDI in the study protocol).

i.e. $$RTDI(\%) = \frac{ATDI}{PTDI} \times 100$$

Note that RTDI expresses the effect of **reductions, interruptions and delays** as well as premature discontinuation of a treatment.

The algorithm will be modified for patients who die during study treatment or who discontinue study treatment due to disease relapse/second primary invasive non-breast cancers. In the calculation of the planned total dose and the planned total dose intensity for these patients, the planned number of weeks of the last cycle will be equal to the number of weeks actually completed.

For all other patients who discontinued treatment prematurely, the actual dose of the remaining weeks will be considered zero doses without delay and the last planned dose will be used as planned dose in the remaining weeks.

5.1.5.2 Modification of treatment schedule

According to the protocol dose modifications may occur as dosing interruption, dose reduction and dose delay.

A ***dosing interruption*** is defined as the omission of capsules within a cycle.

The recommended starting dose of palbociclib/placebo and the reduced dose levels are given in table 5-1.

Table 5-1: Available Dose Levels

Dose Level	Palbociclib/placebo for 3 out of 4 weeks (3/1 schedule)
Starting Dose	125 mg/d
-1	100 mg/d
-2	75 mg/d
Discontinue Study Treatment or consider 75mg/d (2/2 schedule)	

* Palbociclib/placebo dose reduction below 75 mg/d is not allowed, the schedule may move to 75mg/day two weeks on followed by two weeks off.

A special case of interruptions are **omitted cycles** where the interruption is extended to a complete cycle, but this omitted cycle is followed by at least one cycle, even if this next cycle was discontinued prematurely. Omitted cycles are not the last cycle of study treatment. Those are considered premature treatment discontinuations.

A ***dose reduction*** is defined as a decrease of the actual dose level by at least one dose level. Once a dose has been reduced for a given patient, all subsequent doses should be administered at that dose level, unless further dose reduction is required. The reduced doses in those subsequent cycles are not defined as dose reductions. Dose re-escalation is not allowed.

Relevant overdoses of study medication are defined as the intake of more than 23 capsules instead of 21 capsules during a cycle, e.g.

- More than two capsules on more than two days of a cycle
- More than two capsules during the week of break
- More than two capsules on the same day

Dose delays: A cycle was deemed to have been delayed if (start date of the current cycle – start date of previous cycle – 28) > 3. Delays of ≤ 3 days should have been documented in the category "organizational reasons". In case that those categorizations could not be followed, the actual dose delays will be displayed. i.e. all dose delays will be displayed within the group "Patients with any delay".

5.1.5.3 Treatment discontinuation

Permanent discontinuation of treatment will be documented with reasons.

5.1.5.4 Changes in endocrine treatment

Changes in the endocrine treatment and reasons for each treatment arm and overall will be summarized.

5.1.5.5 Other

Impact of drug exposure based on the PK analysis (described in a separate analysis plan) on efficacy and/or safety findings will be investigated.

5.1.6 Quality of life

Quality of life (QoL) will be evaluated using a general QoL questionnaire (EORTC QLQ-C30), a breast cancer questionnaire (QLQ-BR23) module, a fatigue questionnaire (EORTC QLQ FA13 Fatigue), a mood questionnaire (GAD7) and the EQ-5D instrument. The QoL questionnaires will be completed in cycle 1, 3, 5, 7, 9, 11, and End of Treatment and every six months until final primary endpoint analysis.

The EORTC QLQ-C30 incorporates five functional scales (physical, role, cognitive, emotional, and social), three symptom scales (fatigue; nausea, vomiting and pain) a global health status / QoL scale, and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). Scale scores can be obtained for the multi-item scales.

The fatigue module is measuring cancer related fatigue based on a multidimensional approach including physical, emotional and cognitive aspects of fatigue. One global item is assessing the interference of fatigue with activities of daily living. The reduced fatigue module with 13 items (FA13) will be used in combination with the EORTC QLQ-C30 (5).

The QLQ-BR23 breast cancer module comprises 23 questions assessing disease symptoms, side effects of treatment (surgery, chemotherapy, radiotherapy, and hormonal treatment), body image, sexual functioning, and future perspective. The breast cancer module incorporates five multiple-item scales to assess systemic therapy side effects, arm symptoms, breast symptoms, body image, and sexual functioning. In addition, single items assess sexual enjoyment, hair loss, and future perspective.

The EQ-5D is a widely used, brief self-administered health status instrument consisting of two parts. In the first part patients are asked to describe their health state on five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The score on the five dimensions are summarized to create a single summary score. As the questions may be answered differently in different countries / regions due to different local customs and social perspectives, published weights from the Euroquol group are used in determining the country appropriate summary scores (EQ-5D User's guide). The summary score is called the summary index or the health utility value. In the PENELOPE-B study the UK summary score which ranges from -0.594 to 1 with lower scores corresponding to higher levels of dysfunction will be used. The second part is a visual analogue scale which allows individuals to directly rate their self-perceived overall health state using values from 0 (worst imaginable) to 100 (best imaginable).

The GAD-7 is a brief self-report scale to identify general anxiety disorders (GAD) consisting of 7 items. Response options are “not at all”, “several days”, “more than half the days” and “nearly every day” scored as 0, 1, 2, and 3, respectively. GAD-7 Anxiety Severity will be calculated as the total score of the 7 items.

5.2 Baseline variables

Randomization is stratified by the following factors, as recorded in [redacted] where randomization was performed:

- histological lymph node status at surgery (ypN 0-1 vs ypN2-3)
- age at first diagnosis (<=50 vs >50 yrs.)
- centrally measured Ki-67 value (>15% vs <=15%)
- global region of participating site (Asian vs non Asian)
- risk status (CPS-EG Score >=3 vs CPS-EG Score=2 and ypN+)

Other baseline characteristics such as patient age at first diagnosis, race/ethnicity, height, weight, BMI, ECOG performance status, menopausal status, ER and HER2 status, Ki-67, primary diagnosis (histological confirmation of breast cancer in the core biopsy before start of neoadjuvant chemotherapy), cancer pretreatment, prior medication, medical history are collected. These baseline characteristics are not planned to be included as stratification variables or covariates in statistical models unless otherwise specified in subsections of 5.2.

5.2.1 BMI

BMI will be calculated using the following formula: weight (kg)/(height (cm)*height (cm)), rounded to one decimal place.

BMI classes will be calculated according to the WHO recommendation:

BMI<18.5	Underweight
18.5<=BMI<25	normal range
25<=BMI<30	preobese
30<=BMI<35	obese class 1
35<=BMI<40	obese class 2
BMI>=40	obese class 3

5.2.2 Ki-67

Ki-67 in Korea is given as a range of 10 points, e.g. 20-30. The mid-range is used to achieve comparable data to those from the other countries, in this example 25.

5.2.3 HER2 status

5.2.3.1 Calculation prior to study protocol version C8

HER2 status will be determined using central classification on tissue from post-neoadjuvant residual invasive disease of the breast or if not possible from residual nodal invasion as documented on [redacted]

5.2.3.2 Calculation with study protocol version C8 and later versions

HER2 status will be determined using central classification

- **preferably on tissue from post-neoadjuvant residual invasive disease of the breast or if not possible from residual nodal invasion as documented on** [redacted]
- **on tissue from core biopsy of the breast as documented on** [redacted] **or**

If no surgical specimen and no residual nodal invasion is available, tissue from core biopsy of the breast will be used.

5.2.4 Hormone-receptor status

5.2.4.1 Calculation prior to study protocol version [redacted]

Hormone receptor status will be determined using central classification of ER and PgR (based on percentage of positive stained cells) on tissue from post-neoadjuvant residual invasive disease of the breast or if not possible from residual nodal invasion as documented on [redacted] . Hormone receptor status is considered as negative if both ER and PgR were negative and as positive if at least one was positive.

5.2.4.2 Calculation with study protocol version [redacted] *and later versions*

Hormone receptor status will be determined **using central classification of ER and PgR**

- **preferably on tissue from post-neoadjuvant residual invasive disease of the breast or if not possible from residual nodal invasion as documented on [redacted] or**
- **on tissue from core biopsy of the breast as documented on [redacted] or**

If no surgical specimen and no residual nodal invasion is available, tissue from core biopsy of the breast will be used. Hormone receptor status is considered as negative if both ER and PgR were negative and as positive if at least one was positive.

Patients are eligible for the study with a positive HR status according to one of the assessments.

5.2.5 CPS-EG score

CPS-EG score will be calculated according to the point assignments given in the study protocol as follows:

Table 1. Point Assignments for the CPS + EG Staging System

Stage	Points
Clinical stage	
I	0
IIA	0
IIB	1
IIIA	1
IIIB	2
IIIC	2
Pathologic stage	
0	0
I	0
IIA	1
IIB	1
IIIA	1
IIIB	1
IIIC	2
Tumor marker	
ER negative	1
Nuclear grade 3	1

Abbreviations: CPS + EG, clinical-pathologic staging system incorporating ER-negative disease and nuclear grade 3 tumor pathology; ER, estrogen receptor.

ANATOMIC STAGE/PROGNOSTIC GROUPS			
Stage 0	Tis	N0	M0
Stage IA	T1*	N0	M0
Stage IB	T0	N1mi	M0
	T1*	N1mi	M0
Stage IIA	T0	N1**	M0
	T1*	N1**	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1*	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
Stage IIIC	Any T	N3	M0

5.2.5.1 Calculation prior to study protocol version C8

Clinical stage

Clinical tumor stage will be determined using the maximum diameters from palpation [redacted], breast ultrasound [redacted], mammography [redacted], and breast MRI [redacted] on baseline imaging [redacted] in hierarchical order: preferably MRI will be used followed by mammography, US, and palpation. Clinical tumor stage will be classified according to the TNM staging system. Clinical signs that result in tumor stage classification T4 are:

- extension to chest wall [redacted]
- skin involvement [redacted]
- inflammation [redacted]

Clinical lymph node status will be determined using the nodal status of palpation [redacted] and breast ultrasound [redacted] on baseline imaging [redacted] in hierarchical order: preferably nodal status of US will be used followed by palpation.

Clinical stage will be classified according to the right part of table 1 “ANATOMIC STAGE/PROGNOSTIC GROUPS”.

Pathologic stage

Pathologic tumor stage will be determined using the maximum diameter of the tumor [redacted] as documented on surgery [redacted]. If no tumor size is available the pathologic tumor stage will be determined on the basis of the TNM classification documented on the [redacted].

Pathologic lymph node status will be determined using the total number of positive lymph nodes [redacted] and classified according to TNM staging system. If the total number of positive lymph nodes is not available the pathologic lymph node status will be determined on the basis of the TNM classification documented on the [redacted]. If the lymph node status at surgery cannot be determined, it will be considered zero for the calculation of the CPS-EG score.

Pathologic stage will be classified according to the right part of table 1 “ANATOMIC STAGE/PROGNOSTIC GROUPS”.

ER status

ER status will be determined using the percentage of positive stained cells as locally assessed on core biopsies taken before start of neoadjuvant treatment and documented on the [redacted]. If the percentage of positive stained cells is not available ER status will be determined using the estrogen receptor classification [redacted].

Tumor grading

Tumor grading will be determined using grading [redacted] as locally assessed and documented on the [redacted].

Points for the calculation of the CPS-EG score will be assigned according to the left part of table 1. Tumor grading will be used for calculation of CPS-EG score instead of nuclear grade.

5.2.5.2 Calculation with study protocol version C8

Changes in the calculation of the CPS-EG score with study protocol version [redacted] will be highlighted in bold print.

Clinical stage

Clinical tumor stage will be determined using **the maximum of all diameters** from palpation ([redacted]), breast ultrasound ([redacted]), mammography ([redacted]), and breast MRI ([redacted]) on baseline imaging CRF BL07. Clinical tumor stage will be classified according to the TNM staging system. Clinical signs that result in tumor stage classification cT4 are:

- extension to chest wall ([redacted])
- skin involvement ([redacted])
- inflammation ([redacted]).

Clinical lymph node status will be determined using the nodal status of palpation ([redacted]) and breast ultrasound ([redacted]) on [redacted] in hierarchical order: preferably nodal status of US will be used followed by palpation. **Infraclavicular or supraclavicular affected lymph nodes ([redacted]) result in lymph node classification cN3.**

Clinical stage will be classified according to the right part of table 1 “ANATOMIC STAGE/PROGNOSTIC GROUPS”.

Pathologic stage

Pathologic tumor stage will be determined using the maximum diameter of the tumor ([redacted]) as documented on surgery CRF BL05b. If no tumor size is available the pathologic tumor stage will be determined on the basis of the TNM classification documented on the CRF [redacted].

Pathologic lymph node status will be determined using the total number of positive lymph nodes [redacted] and classified according to TNM staging system. If the total number of positive lymph nodes is not available the pathologic lymph

node status will be determined on the basis of the TNM classification documented on the [redacted] . If the lymph node status at surgery cannot be determined, it will be considered zero for the calculation of the CPS-EG score.

Pathologic stage will be classified according to the right part of table 1 “ANATOMIC STAGE/PROGNOSTIC GROUPS”.

ER status

ER status **will be locally assessed as the percentage of positive stained cells**

- **on core biopsies taken before start of neoadjuvant treatment and documented on the** [redacted]
- **on surgical specimen after neoadjuvant treatment and documented on the** [redacted]

If one of the percentages of positive stained cells is not available the estrogen receptor classification (negative/positive) will be used ([redacted] **).**

A negative ER status according to one of the assessments will be considered in the calculation of the CPS-EG score.

Tumor grading

Tumor grading **will be locally assessed on core biopsies taken before start of neoadjuvant treatment (documented on** [redacted] **) and on surgical specimen after neoadjuvant treatment documented on** [redacted] **). The highest grading according to one of the assessments will be considered in the calculation of the CPS-EG score.**

Points for the calculation of the CPS-EG score will be assigned according to the left part of table 1.

Rationale for using ER status and grading alternatively from the core biopsy prior to neo-adjuvant chemotherapy (NACT) or from surgery material after NACT for the CPS-EG score.

Agreement of ER and PgR expression between core needle biopsy performed prior to NACT and surgical sample after NACT have been analyzed extensively. In literature we can find at least six studies concluding that neoadjuvant chemotherapy does not affect the

ER status. In some studies no changes in ER receptor status were found (6, 7), whereas in others non-significant changes were reported. (8, 9, 10, 11, 12).

Concordance between histologic grading tested on core needle biopsy or on excised breast specimen is approximately 83%. In the discordant cases, the grade at core needle biopsy tended to be lower than that at surgical excision. Previous studies noted that this tendency of the histological grade was mainly due to underestimation of the mitotic count at core needle biopsy (Usami S, Jpn J Clin Oncol 2007).

As showing by those data CPS-EG score could be indifferently calculated using ER and grading assessment on both core needle biopsy or tumor sample.

5.2.5.3 Calculation with study protocol version C8 V3

Changes in the calculation of the CPS-EG score since study protocol version C8 V3 will be highlighted in bold print. For details of CRF versions please refer to data management plan.

Clinical stage

Clinical tumor stage will be determined using the maximum of all diameters from palpation [redacted], breast ultrasound [redacted], mammography [redacted], and breast MRI [redacted] on [redacted]. Clinical tumor stage will be classified according to the TNM staging system. Clinical signs that result in tumor stage classification cT4 are:

- extension to chest wall [redacted]
- skin involvement [redacted]
- inflammation [redacted]

Clinical lymph node status will be determined using the maximum nodal status of palpation [redacted], breast ultrasound [redacted], and number of positive sentinel nodes [redacted]. Infraclavicular or supraclavicular affected lymph nodes [redacted] result in lymph node classification cN3.

Clinical stage will be classified according to the right part of table 1 “ANATOMIC STAGE/PROGNOSTIC GROUPS”.

Pathologic stage

Pathologic tumor stage will be determined using the maximum diameter of the tumor ( ) as documented on . If no tumor size is available the pathologic tumor stage will be determined on the basis of the TNM classification documented on the CRF .

Pathologic lymph node status will be determined using the total number of positive lymph nodes ( ) and classified according to TNM staging system. If the total number of positive lymph nodes is not available the pathologic lymph node status will be determined on the basis of the TNM classification documented on the CRF . If the lymph node status at surgery cannot be determined, it will be considered zero for the calculation of the CPS-EG score.

Pathologic stage will be classified according to the right part of table 1 “ANATOMIC STAGE/PROGNOSTIC GROUPS”.

ER status

ER status will be locally assessed as the percentage of positive stained cells

- on core biopsies taken before start of neoadjuvant treatment and documented on the
- on surgical specimen after neoadjuvant treatment and documented on the

If one of the percentages of positive stained cells is not available the estrogen receptor classification (negative/positive) will be used (from).

A negative ER status according to one of the assessments will be considered in the calculation of the CPS-EG score.

Tumor grading

Tumor grading will be locally assessed on core biopsies taken before start of neoadjuvant treatment (documented on) and on surgical specimen after neoadjuvant treatment documented on . The highest grading according to one of the assessments will be considered in the calculation of the CPS-EG score.

Points for the calculation of the CPS-EG score will be assigned according to the left part of table 1.

5.3 Subgroups or covariates of interest

5.3.1 Subgroups

Time-to-event data (iDFS, iDFS excluding second non-breast cancers, OS, LRRFI and DDFS) will be analyzed using univariate Cox-proportional hazards models to calculate hazard ratios for subgroups based on the stratification factors, which were recorded in [redacted] during randomization (predefined in the protocol):

- age at first diagnosis (<=50 vs. >50 yrs.)
- Ki-67 value (>15% vs <=15%)
- histological lymph node status at surgery (ypN 0-1 vs ypN2-3)
- risk status (CPS-EG Score >=3 vs CPS-EG Score=2 and ypN+)

Additional exploratory subgroup analyses will be performed based on strata derived according to clean CRF data:

- CPS-EG Score (score 1/2, score 3, score 4/5)
- Patients with “worst” prognostic factors (age at first diagnosis <=40 years and CPS-EG Score >=3)
- Age at first diagnosis (<=65, >65 years)
- Ki-67 value (>25% vs <=25%)
- Geographical region (North America, Europe, Asia (Pacific))
- Baseline ECOG status (0, 1)
- Baseline menopausal status (premenopausal, postmenopausal)
- First endocrine therapy applied (tamoxifen, AI)
- Duration of chemotherapy (shorter <= 20 wks. vs. longer > 20 wks.)
- Type of surgery (mastectomy, breast conserving)
- Overall response to neoadjuvant chemotherapy (response (CR, PR) vs non-response (SD, PD))

5.3.2 Covariates

Time-to-event data (iDFS, iDFS excluding second non-breast cancers, OS, LRRFI and DDFS) will be analyzed using multivariate Cox-proportional hazards models with the treatment group and 2 sets of covariates. Stratification factors (strat. factor) and variables to determine the relapse risk (relapse risk factors) of the patient are included in the first set of covariates.

Set 1:

• age at first diagnosis (<=50 vs. >50 yrs.)	strat. factor
• Ki-67 value (>15% vs <=15%)	strat. factor
• global region of participating site (Asian vs non Asian)	strat. factor
• histological lymph node status at surgery (ypN 0-1 vs ypN2-3)	strat. factor/ relapse risk
• risk status (CPS-EG Score >=3 vs CPS-EG Score=2 and ypN+)	strat. factor/ relapse risk
• clinical tumor stage before neoadjuvant chemotherapy (cT1/2 vs cT3/4)	relapse risk
• pathological tumor stage after neoadjuvant chemotherapy (pTis/1/2 vs pT3/4)	relapse risk
• grade (grade 1, 2 vs. 3) before neoadjuvant chemotherapy	relapse risk

To investigate the impact of endocrine therapy on iDFS, a second set of covariates will be defined with endocrine therapy added to a reduced set of covariates.

Endocrine therapy will be dummy-coded with tamoxifen as the reference category into four indicator categories: tamoxifen alone, tamoxifen + LHRH, aromatase inhibitors (AI) + LHRH, AI alone.

Set 2:

• age at first diagnosis (<=50 vs. >50 yrs.)	strat. factor
• Ki-67 value (>15% vs <=15%)	strat. factor
• global region of participating site (Asian vs non Asian)	strat. factor
• risk status (CPS-EG Score >=3 vs CPS-EG Score=2 and ypN+)	strat. factor/ relapse risk
• endocrine therapy (dummy coded)	

6. ANALYSIS SUBSETS

Randomized patients consist of all patients who have given their written informed consent and for whom there is confirmation of successful allocation of a randomization number through the web-based data collection interface.

6.1 Data subsets

6.1.1 Intent-to-treat set

Intent to treat (ITT) efficacy population: all randomized patients (including those not starting treatment). This population provides the basis for the main efficacy analyses. Patients will be analyzed according to the treatment group they were randomized to, i.e. patients who took incorrect treatment will be reported under the study treatment assigned at randomization.

6.1.2 Per protocol population for efficacy

The per protocol (PP) population for efficacy is a subset of the ITT and includes patients without major protocol violations.

Prior to unblinding the patients' treatment allocation, the frequencies of the following deviations will be displayed:

- eligibility criteria,
- compliance with intake of study drugs, i.e. overdoses, never received study treatment, medication error (patient received medication she was not randomized to).

On the basis of the protocol deviations' frequencies a subgroup of the steering committee will define in a blinded review the major protocol violations which will lead to the exclusion of patients from the PP population for efficacy. The process of identification of major protocol violations will be documented.

The PP population serves supportive efficacy analyses. If the patient receives the wrong study treatment for the complete duration of the trial, patients will be analyzed according to the treatment actually received.

6.1.3 Safety analysis set

The safety analysis set includes all randomized patients who took at least once study medication. Patients will be analyzed according to the treatment actually received, i.e. a patient who was randomized to the palbociclib arm and received placebo for the entire treatment duration will be analyzed in the placebo arm and vice versa.

For patients in whom treatment was misallocated not for the entire treatment duration so that they received palbociclib and placebo during the trial, the treatment group allocation for safety analyses will be the palbociclib arm.

This is the only population for all safety analyses.

6.1.4 Evaluable subsets for QoL

QoL will be analyzed on the basis of the ITT population. Missing values will be reported descriptively.

6.2 Interim analyses sets

Two event-driven efficacy interim analyses (EIA) were planned according to study protocol. The first EIA was performed after 67 events were documented (details are given in section 7.1). The second EIA was performed after 200 events were documented, (details are given in section 7.1).

7. ADAPTIVE DESIGN

The study has an adaptive design with two interim efficacy analyses (EIA). The objectives of the first interim analysis were sample size re-estimation and the opportunity for early stopping of the trial due to futility.

Sample size re-estimation was planned to be performed only in one of the two interim analyses. Therefore this objective was dropped in the second interim analysis and the remaining objectives were the opportunity for early stopping of the trial due to futility and overwhelming efficacy, respectively.
The decision about early stopping of the trial due to futility or overwhelming efficacy was made on the basis of the primary efficacy parameter, iDFS.

The decision rules for the efficacy interim analyses are summarized in table 7-1.

Table 7-1: Decision rules for interim analyses

Analysis	Time/IF	Futility (HR)	Efficacy (p-value)	SSR*
1st EIA	April 2017 / 67 events	≥1.0	<0.0002	Yes
2nd EIA	April 2019 / 200 events	≥0.90	<0.0120	No
Final	290 events	N/A	<0.0463	N/A

* SSR – sample size re-estimation

Safety including any unexpected toxicity was assessed in both EIAs, based on the description given in section 10.5. Unexpected toxicities were described regarding suspected unexpected serious adverse reactions (SUSAR) and coded according to the MedDRA preferred term (MedDRA version XXX was used).

Details of the interim analyses' objectives are described in the following sections.

7.1 First efficacy interim analysis

The first EIA was performed in April 2017 after 67 events were documented. The EIA was preponed to allow for sample size increase during the enrollment period of the study, which was estimated to be terminated in August 2017.

The final first addendum to the general IDMC charter was signed on the 30th March 2017 where study specific procedures were laid down, i.e. to define decision rules for sample

size re-estimation and for early stopping of the trial due to futility. The addendum to the IDMC charter also contained an overview of the required information and templates of the tables and figures that the IDMC received. The report of the 1st EIA was finalized on the 21st April 2017 and sent to the IDMC on the 25th April 2017. In a telephone conference on 3rd May 2017 the IDMC reviewed the unblinded results of the 1st EIA and recommended to increase the sample size to 1250 patients and the number of events to 290. The recommendation was approved in the SC meeting on 11th May 2017, and implemented in protocol version G11.

7.1.1 Re-estimation of sample size

Sample size re-estimation was mainly based on the primary endpoint, iDFS. The secondary endpoint of iDFS excluding second primary of non-breast cancer was also be taken into consideration when performing sample size re-estimation.

Sample size re-estimation (SSR) was performed according to Cui, Hung and Wang, CHW, (13) who proposed “a valid inferential procedure that allows flexibility for adjusting sample size based on the updated estimate of treatment effect during the course of the trial” (ibid., p. 853).

The calculation of the total sample size N was described in the study protocol and was summarized in section 4.3 of this SAP. The decision whether to adapt the sample size or not was based on conditional power (CP) defined as "the probability that the final study result will be statistically significant, given the data observed thus far" (14).

Sample size adaptation was performed according to Mehta and Pocock (15), depending on the zone in which the CP of the estimated $\hat{\delta}_L$ at the Lth interim analysis (CP$\hat{\delta}_L$) will fall. The zones were defined as follows:

Unfavorable zone: CP$\hat{\delta}_L$ < CP_{min}
Promising zone: CP_{min} < CP$\hat{\delta}_L$ < 1-β
Favorable zone: 1-β < CP$\hat{\delta}$

If CP$\hat{\delta}_L$ falls into the promising zone, it is possible to increase the sample size to achieve a conditional power of an acceptable level. The threshold CP_{min} for determination of the promising zone was provided in the study specific addendum to the IDMC charter. This addendum was only reviewed by IDMC members and the unblinded statistician.

Based on the initially planned number of 255 events, the upper limit of events was set to a maximum number of 290 events (approximately 14% increase). To increase the sample size in a comparable ratio as the number of events the maximum sample size was set to 1250 patients.

Procedure for SSR in the first EIA (as described in the addendum to the IDMC charter):

- The hazard ratio and its 95%CI was calculated for the primary endpoint iDFS, and for the secondary endpoint of iDFS excluding second primary, respectively.
- The conditional power (CP) for both endpoints was computed based on their estimated treatment effects as though they were the true values of the treatment effect. Conditional power was derived from the power calculator within EAST 6.4.
- For a CP in the promising zone the increase of CP was calculated using the maximum number of events. For a CP out of the promising zone, the trial was continued to the originally planned next interim analysis with the original number of events and sample size.
- IDMC was asked to give their recommendation whether SSR might be contemplated and to what extent.
- The final decision about the SSR was made by the sponsor after receiving advice from the Steering Committee.
- In case of SSR the number of events was planned to be increased until target CP of 85% is achieved or until maximum number of events (= 290 events) is achieved, whichever is achieved first.
- If the estimated treatment effect from interim analysis was higher than the initially planned treatment effect, i.e. less events would be needed on the basis of the interim results, the original number of events were used.

Type I error after sample size increase was preserved by re-weighting the Wald statistics for the following analyses, i.e. by down weighting the contribution of the subsequent stage relative to the precedent stage (13).

7.1.2 Early stopping of the trial due to futility

O'Brien-Fleming type stopping boundaries based on the Lan-DeMets spending function were used. The main focus of the first EIA was to assess whether the treatment effect at this stage is futile. Futility criteria in terms of the HR and the Z-value are given in table 7-1. The nominal α level for the first efficacy interim analysis is provided, although there was no intention to stop the trial for efficacy at this time.

7.2 Second efficacy interim analysis

A second addendum to the general IDMC charter was signed on the 2nd April 2019 where study specific procedures are laid down and detailed. The addendum to the IDMC charter also contains an overview of the required information and templates of the tables and figures that the IDMC received.

The report of the 2nd EIA was finalized on the 29th April 2019 and sent to the IDMC. In a telephone conference on 7th May 2019 the IDMC reviewed the unblinded results of the 2nd EIA and recommended to continue the trial without any changes to the final analysis.

7.2.1 Early stopping of the trial

The efficacy and futility boundaries were generated simultaneously keeping the former unaffected by the definition of the latter to receive non-binding futility boundaries. Futility and efficacy criteria are given in table 7-1. O'Brien-Fleming type stopping boundaries based on the Lan-DeMets spending function were used.

Differences in the primary endpoint iDFS between treatment arms in the ITT population were tested with a 2-sided stratified log-rank test at the overall significance level 0.05. The factors used in the stratified log-rank test were histological lymph node status at surgery (ypN 0-1 vs ypN2-3), age at first diagnosis (<=50 vs >50 yrs.), Ki-67 (>15% vs <=15%), and risk status (CPS-EG Score >=3 vs CPS-EG Score=2 and ypN+) as recorded during randomization.

Due to sample size increase the type I error in testing the primary endpoint was preserved using the CHW method.

7.3 Final Analysis

In the final analysis the primary endpoint iDFS will be tested in the same way as in the second EIA (described in the previous section 7.2.1). According to the design parameters the nominal significance level for the final analysis on iDFS is 0.0463 (Z-statistic=-1.993) calculated using the Lan-DeMets (O'Brien-Fleming) α-spending function.

To maintain the type I error after sample size increase, statistical significance will be determined using a weighted statistic based on the CHW method (13) which is implemented in the CHW interim monitoring in EAST version 6.5 (Cytel Inc.). The parameter from EAST will be derived, exported, and archived.

7.4 Interim Overall Survival Analysis

At the time of the final iDFS analysis (with 290 documented iDFS events) an interim OS analysis will be conducted. OS was defined in section 4.1.

7.5 Handling of blinding and unblinding

Handling of blinding and unblinding was performed according to [redacted]. The study statistician prepared SAS programs to perform the analyses using a dummy variable representing the treatment code to remain blinded to the treatment allocation of the patients. The dummy treatment variable was created using the rantbl function in SAS to create a dichotomous variable; stratification was not incorporated in the dummy variable.

Final SAS programs for the interim analyses were handed over to the unblinded statistician who replaced the dummy treatment code by the actual treatment allocation of the patients. The study statistician does not have access to the unblinded analyses or any unblinded results prior to database lock.

The IDMC received reports in two parts: an "open" session, which presented data only in aggregate and focuses on trial conduct issues such as accrual and dropout rates, timeliness of data submission, eligibility rates and reasons for ineligibility; and a "closed" session, in which the comparative outcome data were presented." (FDA guideline (16). p. 11). In the “closed” session of the IDMC meeting the unblinded statistician presented the study data to IDMC members while all people other than the unblinded statistician and IDMC members have to be absent.

The IDMC gave advice concerning early stopping, modification on sample size, and continuing the study as planned after the meeting in writing (detailed procedures of the IDMC meetings are given in the general IDMC charter and in its study specific addendum for the PENELOPE^{B} trial).

8. MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

Due to a very small subgroup of luminal B tumors in the patients' cohort the respective secondary endpoint cannot be analyzed and further exploratory analysis will be performed which will be described in a separate biomarker SAP.

The following Amendment(s) were implemented since the first study protocol A (Version 6-23.09.2013)

Country	Amendment No.	Date of Amendment
Global	**Protocol G (Version 11 – 04th May 2017)** **Summary of Changes** Specification of potential outcomes of the efficacy interim analysis (Futility, sample size re-estimation up to a new total number of 1250 patients, efficacy) have been added to statistical section in synopsis, section 11.1 and 15 ff. Rationale: Adaptive design of the trial does allow adjustment of patient number based on outcome of 1st efficacy interim analysis. **Administrative and corrective changes:** 1. Procedure for unblinding in case of progressive disease specified in section 12.2	04th May 2017
Global	**Protocol E (Version 10 – 12 April 2016** **Summary of Changes:** **Substantial** Inclusion Criterion #2: Specification for bilaterial breast cancer added Inclusion Criterion #5: Specification which tissue can be used for central testing added Inclusion Criterion #6: Specification for bilaterial breast cancer added Exclusion Criterion #16: Study entry time was specified as date of randomization. Endocrine treatment options updated. (section 12.8.1) Safety Monitoring Frequency adjusted to IDMC recommendation in section 15.5.7 Specification of relevant overdose definition in section 17.5.2 and removal notification requirement for non-relevant overdose **Administrative and corrective changes:** Alignment of table "schedule of activities" and tables "assessments and schedules" with current protocol Addition of additional time points for PK-DDI samples in case a sample was missed for a patient.	12 April 2016
Global	**Protocol D (Version 9 - 09 February 2015)** **Summary of Changes:** **Substantial** Inclusion criterion #5: Clarification for testing in case of bilateral breast	09 Feb 2015

Country	Amendment No.	Date of Amendment
	cancer Inclusion criterion #6: Centrally testing possibility extended to core biopsy Inclusion criterion #12: allows now patients with a CPS-EG Score of 2 if ypN+ to participate Addition of new stratification criterion: CPS-EG score 3 vs 2&ypN+ Update of section 15.5 ff Statistical Analysis due to change of inclusion criterion #12 and increase of study population **Administrative and corrective changes:** Specification of PK sample set requirements in section 13.3 Update of Section 7 rationale of the study	
Global	**Protocol C (Version 8-06.08.2014)** **Substantial Amendment Summary of Changes:** Prognostic Marker Inclusion Criterion #12 with score CPS-EG, allows now the use of surgical biopsy: … using local estrogen receptor status and grade assessed on core biopsies taken before start of neoadjuvant treatment either / or surgical biopsy.	06.Aug.2014

9. DATA SCREENING AND ACCEPTANCE

9.1 General principles

The documentation and verification procedure will be performed according to [redacted] .

9.2 Database lock

The database lock for the final analysis is planned after 290 iDFS events are documented and confirmed [redacted]. Confirmation of 290 events was documented on the 24th August 2020. DB lock will be performed on the 25th September 2020.

9.3 Data handling and electronic transfer of data

Clinical trial data will be exported from [redacted] into [redacted]. Additional / non-CRF data will be used for the final analysis, which are listed in [redacted]:

All data will be imported in SAS® (Statistical Analysis Software) Version 9.4 and SAS Enterprise Guide Version 7.1 on Microsoft Windows 10 Enterprise.

[redacted]

9.4 Handling of missing and incomplete data, drop-outs

Missing statistics, e.g., when they cannot be calculated, should be presented as ‘n.a.”. For example, if N=1, the measure of variability (SD) cannot be computed and should be presented as ‘n.a.’.

9.4.1 Missing data in CPS-EG score

If no LN status could be determined for calculation of CPS-EG score, LN will be assumed to be 0.

9.4.2 Missing dates

In general, missing or incomplete dates will be imputed as follows: if day is missing, the first day of the month will be assigned. If the day of the month and the month are missing for any date used in a calculation, January 1st will be used to replace the missing date. These rules are used unless the calculations result in negative time durations (e.g., date

of onset cannot be prior to day 1 date). In these cases, the dates resulting in 0 day duration will be used.

9.4.3 Missing Data in Time-To-Event Endpoints

For all time to event endpoints if day of event is missing, the first day of the month will be assigned. However, if the month and the year of the event is similar to the month and year when the study drug was given, then the date of the application of study drug will be assigned.

Drop-outs without treatment are defined as those patients who have been randomized in the study but withdrew their consent or are withdrawn by the investigator from the study immediately thereafter but prior to first application of study medication. The reasons are collected and reported in the Consort Statement. These patients will not be replaced but belong to the ITT efficacy population.

In case of missing disease assessments before the last disease assessment, or before death iDFS will be calculated on the basis of the given disease assessments and not the scheduled disease assessments. Therefore, iDFS will be reported according to the actual reporting time and not the scheduled time.

In case that the imputation of the start date of new anticancer treatment results in a start date which is before randomization the imputation will be corrected to the day after randomization (date of randomization + 1).

All other missing values will be reported explicitly.

9.4.4 Date of Last Contact Derivation

In the analysis of overall survival the date of last contact will be derived for patients not known to have died at the analysis cutoff using the latest complete date (after imputation) among but not limited to the following:

- Date of follow up [redacted]
- Date of withdrawal [redacted]
- Date of last contact when patient status is “lost to follow up” [redacted]
- Date of relapse [redacted]
- Date of assessments (e.g., laboratory, physical examination, ECG, disease assessments, QoL questionnaires)
- Date of first endocrine therapy [redacted] and end date of endocrine treatment and date of change on endocrine treatment [redacted]
- Date of first intake of study drug in given cycles [redacted]
- Date of last known treatment intake [redacted] and date of last known endocrine treatment [redacted]

- Start and end dates of follow-up anticancer therapies administered [redacted]
- Randomization date

Only dates associated with actual examinations of the patient will be used in the derivation. Assessment dates after the cut-off date will not be applied to derive the last contact date.

9.4.5 Death Date

Missing or partial death dates will be imputed based on the last contact date:

- If the date is missing it will be imputed as the day after the date of last contact
- If the day or both day and month is missing, death will be imputed to the maximum of the full day after the date of last contact and the following:
  - Missing day: 1st day of the month and year of death
  - Missing day and month: January 1st of the year of death

9.4.6 Missing data in safety analysis

Missing CTC grades will be imputed by the worst (non-fatal) grade possible for the respective AE, i.e. CTC 2 for alopecia or CTC grade 3 for dysgeusia etc.

AEs with missing relationship to study medication will be considered related.

9.5 Evaluable subject subset(s)

Patients will be included in the Safety Population if they started treatment. Information about the start of the study treatment will be available for all patients. Therefore, imputation will not be performed.

9.6 Outliers

Most variables in the analysis are time-to-event or categorical so the outlier problems do not apply to them.

Before converting hematology and other lab values in CTC-grades, outliers will be examined and queries will be sent out by the [redacted].

9.7 Distributional characteristics

No tests for distribution will be performed, non-parametrical tests will be used for all continuous baseline variables.

9.8 Testing/validation plan

Analysis programs will be validated according to the [redacted].

All data, except for the HTG data, will be analyzed using SAS® (Statistical Analysis Software) Version 9.4 or higher with SAS Enterprise Guide Version 7.1 on Microsoft Windows 10 Enterprise. HTG data will be analyzed using an R-software package. Details are available in the biomarker SAP regarding the luminal B analyses. SAS programs will be clearly documented (using commenting facilities within programs) and stored. All SAS-logs will be stored.

Estimates based on the CHW method will be derived from the CHW interim monitoring implemented in EAST version 6.5 (Cytel Inc.). The EAST workbook will be stored.

10. STATISTICAL METHODS OF ANALYSIS

10.1 General principles

The overall significance level for the primary endpoint is set to two-sided alpha = 0.05. There will be two interim efficacy analyses and a final analysis. Type I error inflation will be controlled by using the Lan-DeMets spending function which results in stopping boundaries comparable to O'Brien-Fleming stopping boundaries.

For all other tests alpha is also set to 0.05 (two-sided). Adjustment for multiple testing in the other tests is not planned. These statistical tests are to be considered descriptive.

All multivariate models will be fit as full models without variable selection.

All percentage values will be valid percent, i.e. missing data will be excluded from the calculations.

10.1.1 Analyses for Binary Data

In general, binary endpoints will be summarized as number in percentages in each group and overall. Fisher's exact test for binary variables will be used to assess the comparability of two treatment arms (usually used to assess baseline comparability).

10.1.2 Analyses for Categorical Data

Categorical variables will be summarized by frequency counts and percentages. All percentage values will be valid percent. The Pearson Chi^2-test for categorical parameters with more than 2 categories will be used to assess the comparability of two treatment arms.

10.1.3 Analyses for Continuous Data

Continuous parameters will be summarized as mean, standard deviation, median, minimum and maximum. The Mann-Whitney test (for continuous parameters) will be used to assess the comparability of two randomized treatment arms.

10.1.4 Analyses for Time-to-Event Data

The time-to-event data include iDFS, iDFS excluding second primary invasive non-breast cancers, LRRFI, OS, DDFS, and iDFS in patients with luminal-B tumors.

Time-to-event variables (except LRRFI) will be presented using the Kaplan-Meier method and displayed graphically when appropriate. The survival probability at the specific time point will be estimated using the Kaplan-Meier method and a 2-sided 95% CI will be calculated using the log-log transformation with back transformation to a CI on the untransformed scale.

LRRFI will be analyzed using the Fine-Gray model (21), as described in section 10.8.1.

10.1.5 Standard Derivations and Reporting Conventions

The following conversion factors will be used to convert days into weeks, months or years: 1 week = 7 days, 1 month = 30.4 days, 1 year = 365.25 days.

10.2 Subject accountability

The number of patients in each of the following categories will be reported:

- Screened patients
- Screening failures and reasons
- Randomized patients
- Randomized but not treated patients with reason for not starting treatment
- Randomized and treated patients in the following subgroups:
    - histological lymph node status at surgery (ypN 0-1 vs ypN2-3)
    - age at first diagnosis (<=50 vs >50 yrs.)
    - Ki-67 value (>15% vs <=15%)
    - global region of participating site (Asian vs non Asian)
    - risk status (CPS-EG Score >=3 vs CPS-EG Score=2 and ypN+).
- Patients who completed the study treatment period as per protocol
- Patients who discontinued study treatment by main reason for permanent treatment discontinuation
- Status at last study contact – will be reported with the time-to-event outcome analysis

It will be checked for each patient enrolled whether she violates any of the inclusion or exclusion criteria.

Median follow-up time will be estimated with the reverse Kaplan-Meier method and the completeness of the follow-up will be assessed according to Clark and colleagues (18).

10.3 Protocol deviations

Patients with deviation from the eligibility criteria will be listed with site ID, treatment arm, and protocol version under which the patient was randomized.

Other important protocol deviations will be reported according to the following categories (subcategories)

- Safety Reporting (SAE not reported within 24 hours to sponsor)
- Concomitant treatment (Administration of prohibited concomitant medications during active treatment)
- Informed Consent (Missing complete ICD or ICD not obtained, Missed date by patient, Missing patient's signature, Randomization/specific protocol procedures initiated prior to signing/dating consent Investigational Product)
- Administration/Study Drug (Study drug not dispensed by site via IWRS, Received study drug not as randomized, relevant overdose)
- Randomization Errors (Patient randomized under incorrect stratification, Patient randomized twice)
- Withdrawal Criteria (Patient experienced unacceptable toxicity requiring permanent treatment discontinuation but was not discontinued from active treatment)

Frequency (number and percentage) of patients with concordant strata as documented during the randomization and based on clean data will be presented.

Other than the above listed irregularities within the randomization process or any other occurrences related to the hand-over of randomization codes will be listed.

10.4 Demographic and baseline characteristics

The demographic and baseline characteristics will be reported descriptively in each setting per treatment arm and overall.

Categorical and ordinal data will be summarized using the number and percentage of patients in each treatment group. Continuous data will be summarized using the number of available data, mean, standard deviation (SD), median, minimum, and maximum for each treatment group.

The Pearson Chi^2-test (for categorical parameters with more than 2 categories), Fisher's exact test (for binary parameters), and the Mann-Whitney test (for continuous parameters) will be used to assess the comparability of two randomized treatment arms in each setting. These statistical tests are to be considered descriptive, no adjustment for multiple testing is planned.

The following demographic and baseline characteristics will be displayed:

- Age at first diagnosis (prior to neo-adjuvant chemotherapy), race
- Height, weight, BMI, BMI classes
- ECOG performance status
- Menopausal status
- Tumor size, tumor focality
- Tumor size assessed at surgery (after neo-adjuvant chemotherapy), histological tumor stage, lymph node status
- CPS-EG score
- Hormone receptor status assessed locally at the centers / at central pathology;
- HER2 status assessed locally at the centers / at central pathology;
- Tumor grading, local (calculated as the maximum of the assessment on core biopsy and on surgical tissue),
- histological tumor type
- Ki-67 assessed at central pathology;
- Cardiac examination at baseline
- General medical history
- Concomitant medication
- Cancer pretreatment
  - Previous neoadjuvant chemotherapy and duration
  - First Endocrine therapy applied
  - Radiotherapy
- Overall clinical response after neoadjuvant chemotherapy
- Time between date of histological confirmation of breast cancer and start of neoadjuvant chemotherapy
- Time between end of neoadjuvant chemotherapy and first breast surgery
- Time between last breast surgery / end of radiotherapy and start of study treatment
- First endocrine therapy and concomitant start of endocrine therapy and Palbociclib/Placebo by treatment arm and by menopausal status

10.5 Safety analyses

All safety analyses will be performed using the evaluable subset for safety.

The most current version of the MedDRA classification system available prior to database lock will be used for AE coding.

Analysis will be performed using any AE that occurred after first dose of study medication ignoring baseline events.

AE data will be aggregated based on the maximum grade that the patient experienced within each MedDRA preferred term (PT) during the reporting period. If a patient experiences more than one AE within a PT for the same summary period, only the AE with the worst NCI CTCAE v.4.0 severity grade will be included in the summaries of severity ('patients-based'). Missing CTC grades will be imputed by the worst non-fatal grade possible for the respective AE, i.e. CTC 2 for alopecia or CTC grade 3 for dysgeusia etc. Safety analyses will be performed based on both all causality AEs (irrespective of relationship) and treatment-related AEs. If the relationship is missing the AE will be considered related.

Predefined and free-text AEs are documented with CTC grades (NCI-CTCAE version 4.0, NYHA for cardiac events) on the CRFs. The hematological parameters will be converted into CTC grades according to the [redacted]. Adverse Events of Special Interest (AESIs) are documented as being present or absent without CTC grades.

If an AE is reported as free text but belongs to one of the predefined categories, it will be included in the corresponding predefined AE category (including multiple infections as described in section 5.1.4).
Unless otherwise specified, AEs will be summarized by frequency and percentage of patients within the AE category of interest, by treatment arm, and overall, in the order given on [redacted].

SAEs and AEs will be completely reconciled. Therefore, all hematological and non-hematological SAEs will be considered in the AE summary tables. In the AE summary tables the grade of the documented AE (only severity grade 1-4 available on [redacted] will be given. This might result in discrepancies in case of fatal serious AEs, which will be displayed with CTC grade 5 in the 'Listing of SAEs' whereas the corresponding AE will be displayed with the CTC grade as documented on the CRF [redacted], respectively.

A summary table of the most common serious adverse events by preferred term will be displayed. “Most common” is defined as occurring in at least 2 patients in either treatment arm.

Anemia, leucopenia, neutropenia, febrile neutropenia, and thrombocytopenia will be merged into ‘Patients with hematological AE’. All other AEs will be merged into ‘Patients with non-hematological AE. All AEs will be aggregated into ‘Patients with Adverse Event. AEs reported as free text (AE 02 - Adverse Events) will be summarized in ‘Patients with other AE’. SAEs will be aggregated as ‘Patients with Serious AE, ’ . ‘Patients with hematological SAE‘, and ‘Patients with non-hematological SAE’. This results in a “Summary of Adverse Events” table including the above mentioned categories.

In addition, a summary table of the most common adverse events by preferred term will be given based on any grade AEs. “Most common” is defined as occurring in at least 10% of the patients in either treatment arm.

Free text AEs occurred in at least 5% of patients will be reported according to MedDRA preferred terms, less frequent AEs will be included into “other AE SOC XX”.

AESIs will be displayed according to the template table 14-21 with treatment modification (1 = no action taken / 2 = Dose delay / 3 = Dose reduction / 4 = Interruption / 5 = Permanent discontinuation of treatment). AESIs will not be included in the overall AE tables due to documentation without grades.

The occurrence of AE categories will be displayed as

- number and percentage of grades 1-4 (any grade) per treatment group and overall,
- number and percentage of grades 3-4 (high grade) per treatment group and overall,
- number and percentage of patients per grade (none, grade 1, grade 2, grade 3, grade 4) per treatment group and overall.

Fisher’s exact test will be performed to explore differences in the incidence of any grade and high grade AEs, respectively, between treatment groups. These p-values are descriptive in nature and no adjustment for alpha inflation will be performed.

Any death of patients either under therapy (from start of treatment until and including 30 days after last dose) or in the follow-up period will be listed together with the treatment arm, cycle/follow-up number, and primary death cause.

Death occurring from start of treatment until and including 30 days after last dose will be displayed for each treatment arm and overall, grouped (according to DRF) as (single choice)

- Tumor related
- Therapy related
- Not therapy and not tumor related
- Unknown.

10.6 Compliance analyses

The extent of study treatment exposure and compliance will be assessed and summarized by actual treatment received within the *All randomized and treated patients population*.

10.6.1 Investigational product

The duration of palbociclib/placebo exposure and the total number of cycles per patient will be displayed for each treatment arm and overall.

The cumulative dose, the ATDI, the RTDI, and the RTD will be given per patient for each treatment arm and overall.

Incidence of dose delays, dose reductions and interruptions of palbociclib/placebo as well as duration of dose delays and interruptions will be reported per patient for each treatment arm and overall.

Reasons for dose delays, dose reductions and interruptions (including omitted cycles) will be grouped according to CRF as (multiple choice):

- Organizational
- AE:
    - Hematological toxicity
    - (Other) non-hematological toxicity
    - AE not related to study medication
- Other

Based on blinded review of data the following additional reasons will be extracted for dose delays and interruptions from the category “other”:

- Patient's non-compliance,
- Surgical or Non-surgical medical procedure.
- And Disease Progression only for interruption.

Treatment interruption and premature treatment discontinuation

For patients who discontinue the study treatment prematurely within a cycle, all remaining days without drug intake are indicated by the center as “no intake” on the [redacted], which will be reported as treatment interruptions. To avoid that the documentation of a premature treatment discontinuation also has impact on other measure of the compliance analysis, the days without drug intake indicated by the center as “no intake” on Treat 01/P-P will not be taken into account.

Frequency (number and percentage) of patients who completed 13 cycles of palbociclib/placebo and who completed at least 7 cycles of palbociclib/placebo, respectively, will be given for each treatment arm and overall.

The incidence and reasons of premature end of treatment will be reported for each treatment arm and overall.

Reasons for premature end of treatment will be grouped as (single choice)

- Disease recurrence/Second primary invasive non-breast cancer
- Death
- Adverse event:
    - Hematological toxicity
    - (Other) non-hematological toxicity
    - AE not related to study medication
- Patient decision
- Investigator decision

10.6.2 Non-investigational product

The incidence and reasons of changes in the endocrine treatment and the incidence and reasons of delays in the endocrine treatment will be reported for each treatment arm and overall.

Reasons for changes and dose delays in the endocrine treatment will be grouped according to CRF as (multiple choice):

- Organizational (only for delays and interruptions, up to 3 days)
- AE:
    - Hematological toxicity
    - (Other) non-hematological toxicity
    - AE not related to study medication
- Other

The incidence and reasons of premature end of endocrine treatment will be reported for each treatment arm and overall.

by menopausal status

10.7 Primary efficacy analyses

Final analysis of the primary endpoint will be conducted when 290 events in iDFS have been documented in the database and have been confirmed.

The primary analysis will be the comparison of survivor functions between treatment arms in the primary efficacy endpoint iDFS in the ITT population.

10.7.1 Estimation of outcomes

The Kaplan-Meier curves of iDFS, Kaplan-Meier estimates of 3-, and 5-year probability of iDFS and the corresponding 95% CI will be provided.

Frequency (number and percentage) of patients with each type of iDFS event and censoring reasons will be presented by treatment arm.

Duration of follow-up in the treatment arms will be summarized using the inverse Kaplan-Meier method, reversing the OS censoring and event indicators.

10.7.2 Hypotheses to be tested

Null hypothesis: survivor functions of the two treatment arms are equal;
Alternative hypothesis: survivor functions of the two treatment arms are not equal.

A two-sided stratified log-rank test at the overall significance level of 0.05 will be performed. The stratification factors are: histological lymph node status at surgery (ypN 0-1 vs ypN2-3), age at first diagnosis (<=50 vs >50 yrs.), Ki-67 (>15% vs <=15%) and risk status (CPS-EG score >= 3 vs. CPS-EG score = 2 and ypN+). In the primary analysis the stratification factors will be used as randomized, i.e. with the values collected at the time of randomization. Based on the estimated treatment effect and its standard error the weighted statistic (CHW) and the repeated CI95% of the HR will be derived from the CHW Interim Monitoring dashboard in EAST taking into account the adaptive sample size re-estimation and group-sequential nature of the design.

This study will be considered a positive trial if the two-sided p-value from the weighted log-rank test based on the CHW method is significant at the significance level of 0.0463 at the time of the final analysis. In addition, the unweighted stratified log-rank test will be presented to support the main analysis of the primary endpoint. The analysis of the key secondary endpoint will be performed to test the null hypothesis that the survivor functions of iDFS excluding second primary invasive non-breast cancers are the same for the two

treatment arms vs. the alternative hypothesis that the survivor functions are different for the two treatment arms.

10.7.3 Subgroup analysis

The primary endpoint will be analyzed in the subgroups listed in 5.3.1. There will be no adjustment for multiple comparisons in the analyses in subgroups. The results will be presented in tables and graphically as a forest plot.

10.7.4 Regression analysis

Univariate and multivariate Cox-proportional hazards model will be used for iDFS to report hazard ratios with 95% CI and to adjust for the covariates defined in 5.3.2. The results will be presented in tables and graphically as a forest plot.

10.7.5 Sensitivity/Supportive Analyses

The following sensitivity analyses / supportive analyses will be performed to explore the robustness of the primary analysis' results:

A. Analysis based on the per protocol analysis set for efficacy,
B. Unstratified analysis,
C. Analysis using strata derived according to clean CRF data instead of those recorded in [redacted] during randomization,
D. Analysis in which patients receiving new anticancer therapy prior to occurrence of an iDFS event will be censored at the date of last disease assessment where the patient was "alive and tumor-free/without new signs of disease" prior to new anticancer therapy.
E. Analysis in which patients who start a new anticancer therapy prior to disease relapse or occurrence of a second primary invasive non-breast cancers or death will be considered to have had an event on the start date of new anticancer therapy.
F. Analysis in which patients who are ER-positive at surgery based on centrally assessed tissue from post-neoadjuvant residual invasive disease.

These analyses are regarded as purely exploratory.

10.8 Secondary efficacy analyses

Final analysis on the secondary efficacy endpoints (except for OS) will be conducted when 290 events in iDFS have been observed and documented in the database. At the time of the final iDFS analysis an interim OS analysis will be conducted (see section 7.2).

At the time of interim iDFS analyses iDFS excluding second primary invasive non-breast cancers was hierarchically tested for significance, provided the primary endpoint, iDFS,

was statistically significant. As iDFS was not significant, iDFS excluding second primary invasive non-breast cancers was not statistically evaluated.

At the time of final iDFS analyses, the hierarchical testing will be performed in the following sequence: primary endpoint iDFS, iDFS excluding second primary invasive non-breast cancers and OS. The next level of testing will only be provided if the prior level is significant. If a test is not significant, the next test will not be statistically evaluated.

10.8.1 Estimation of outcomes

The Kaplan-Meier curves of the time-to-event data, Kaplan-Meier estimates of 3-, and 5-year probability of the time-to-event data and the corresponding 95% CI will be provided. The time-to-event data are: iDFS excluding second non-breast cancers, OS, DDFS, iDFS per treatment group in patients with luminal-B tumors.

LRRFI including competing risks will be analyzed using the Fine-Gray model (21), both univariately (the cumulative incidence function will be presented graphically per treatment arm) and multivariately to report HR for treatment and to adjust for the stratification factors.

10.8.2 Hypotheses to be tested

Null hypothesis: survivor functions of the two treatment arms are equal;
Alternative hypothesis: survivor functions of the two treatment arms are not equal.

A two-sided stratified log-rank test at the overall significance level 0.05 will be performed to compare time-to-event efficacy outcomes between treatment arms, the same stratification factors will be used as for the primary endpoint.

10.8.3 Subgroup analysis

The secondary time-to event endpoints will be analyzed in the stratified subgroups. There will be no adjustment for multiple comparisons in the analyses in subgroups. The results will be presented in table and graphically as a forest plot.

10.8.4 Regression analysis

Univariate and multivariate Cox-proportional hazards model will be used for time-to-event data to report hazard ratios with 95% CI and to adjust for the covariates defined in 5.3.2. The results will be presented in table and graphically as a forest plot.

10.9 Quality-of-life analyses

Quality of life analyses (QoL) will be performed using the evaluable subset for QoL.

The scales (based on multiple items) of all QoL instruments will be computed according to the respective scoring manuals. Summary statistics (mean, standard deviation, median, and range) of observed values and calculated scores and their changes from baseline to each assessment time point will be summarized for the two treatment arms measured as Cohens effect size. For the EORTC QLQ C30 questionnaire standard values are available, which will be displayed in the descriptive tables.

Repeated-measures mixed-effects models will be used with main effect terms “treatment” and “time”, the interaction term “treatment-by-time”, and baseline values as covariate. Time is considered to be continuous. Treatment effect changes of the QoL scales over time and treatment-by-time interaction will be explored.

Repeated-measures mixed-effects models will be fit for questionnaires / scales consisting of more than one item:

- EORTC QLQ-C30: Global health status/QoL, Functional scales: Physical Functioning, Role functioning, Emotional functioning, Cognitive functioning, Social functioning; Symptom scales: Fatigue, Nausea and Vomiting, Pain.
- EORTC QLQ-BR23: Functional scales: Body image; Symptom scales: Systemic therapy side effects, Breast symptoms, Arm symptoms; Sexual functioning and Sexual enjoyment are often missing and will also not be considered.
- QLQ-FA13: Physical fatigue, Emotional fatigue, Cognitive fatigue.
- EQ-5D-3L: Index score

10.10 Follow-Up Therapy

Follow-up anti-cancer therapy will be summarized by treatment type (radiation, chemotherapy, hormone therapy, targeted therapy, surgery, and other) and treatment line as patients with number of regimens (0, 1, 2, ≥3).

11. LITERATURE CITATIONS / REFERENCES

1. Mittendorf EA, Jeruss JS, Tucker SL, et al. Validation of a novel staging system for disease-specific survival in patients with breast cancer treated with neoadjuvant chemotherapy. J Clin Oncol. 29(15):1956-62, 2011

2. Jeruss JS, Mittendorf EA, Tucker SL, et al. Combined use of clinical and pathologic staging variables to define outcomes for breast cancer patients treated with neoadjuvant therapy. J Clin Oncol. 26(2):246-52, 2008

3. Marmé F, Lederer B, Blohmer JU, Costa SD, Denkert C, Eidtmann H, Gerber B, Hanusch C, Hilfrich J, Huober J, Jackisch C, Kümmel S, Loibl S, Paepke S, Untch M, von Minckwitz G, Schneeweiss A. Utility of the CPS+EG staging system in hormone receptor-positive, human epidermal growth factor receptor 2-negative breast cancer treated with neoadjuvant chemotherapy. Eur J Cancer. 2015 Dec 13;53:65-74.

4. Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, Sparano JA, Hunsberger S, Enos RA, Gelber RD, Zujewski JA (2007). Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. J Clin Oncol. May 20;25(15):2127-32.

5. Weis J, Arraras JI, Conroy T, Efficace F, Fleissner C, Görög A, Hammerlid E, Holzner B, Jones L, Lanceley A, Singer S, Wirtz M, Flechtner H, Bottomley A (2013). Development of an EORTC quality of life phase III module measuring cancer-related fatigue (EORTC QLQ-FA13). Psycho-Oncology, 22(5), 1002-1007

6. Arens N, Bleyl U, Hildenbrand R (2005). HER2/neu, p53, Ki67, and hormone receptors do not change during neoadjuvant chemotherapy in breast cancer. Virchows Arch. 2005 May;446(5):489-96.

7. Miglietta L, Vanella P, Canobbio L, Parodi MA, Guglielmini P, Boccardo F (2009). Clinical and pathological response to primary chemotherapy in patients with locally advanced breast cancer grouped according to hormonal receptors, Her2 status, grading and Ki-67 proliferation index. Anticancer Res. May;29(5):1621-5.

8. Lee SH, Chung MA, Quddus MR, Steinhoff MM, Cady B (2003). The effect of neoadjuvant chemotherapy on estrogen and progesterone receptor expression and hormone receptor status in breast cancer. Am J Surg. Oct;186(4):348-50.

9. Faneyte IF, Schrama JG, Peterse JL, Remijnse PL, Rodenhuis S, van de Vijver MJ (2003). Breast cancer response to neoadjuvant chemotherapy: predictive markers and relation with outcome. Br J Cancer. Feb 10;88(3):406-12.

10. Bottini A, Berruti A, Bersiga A, Brunelli A, Brizzi MP, Marco BD, Cirillo F, Bolsi G, Bertoli G, Alquati P, Dogliotti L (1996). Effect of neoadjuvant chemotherapy on Ki67 labelling index, c-erbB-2 expression and steroid hormone receptor status in human breast tumours. Anticancer Res. Sep-Oct;16(5B):3105-10.

11. Varga Z, Caduff R, Pestalozzi B (2005). Stability of the HER2 gene after primary chemotherapy in advanced breast cancer. Virchows Arch. Feb;446(2):136-41. Epub 2005 Jan 14.

12. Adams AL, Eltoum I, Krontiras H, Wang W, Chhieng DC (2008). The effect of neoadjuvant chemotherapy on histologic grade, hormone receptor status, and HER2/neu status in breast carcinoma. Breast J. 2008 Mar-Apr;14(2):141-6.14

13. Cui L, Hung HM, Wang SJ (1999). Modification of sample size in group sequential clinical trials. Biometrics Sep;55(3):853-7.

14. Lachin JM (2005) A review of methods for futility stopping based on conditional power. Stat Med Sep 30;24(18):2747-64.

15. Mehta CR, Pocock SJ (2011). Adaptive Increase in Sample Size when Interim Results are Promising: A Practical Guide with Examples. Stat Med 30; 3267-84.

16. FDA (2006) Guidance for Clinical Trial Sponsors. Establishment and Operation of Clinical Trial Data Monitoring Committees. URL: http://www.fda.gov/RegulatoryInformation/Guidances/ucm127069.htm (last visited: 05.06.2014).

17. GBG SOP *Creation, validation and documentation of statistical analysis programs* Version 3.0

18. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up, The Lancet, Vol. 359 April 13, 2002 ICH E9, Statistical Principles for Clinical Trials.

19. Sinn HP, Schmid H, Junkermann H, et al. Histologische Regression des Mammakarzinoms nach primärer (neoadjuvanter) Chemotherapie. Geburtshilfe Frauenheilkd 54:552-8, 1994

20. GBG Laboratory Value Guideline, Version 8.0

21. Jason P. Fine and Robert J. Gray. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association Vol. 94, No. 446 (Jun., 1999), pp. 496-509.