Clinical Development

LEE011 (Ribociclib)

CLEE011E2301 (MONALEESA-7) / NCT02278120

A Phase III randomized, double-blind, placebo-controlled study of LEE011 or placebo in combination with tamoxifen and goserelin or a non-steroidal aromatase inhibitor (NSAI) and goserelin for the treatment of premenopausal women with hormone receptor positive, HER2-negative, advanced breast cancer

Statistical Analysis Plan (SAP)

Author: [redacted] Trial Statistician

Document type: SAP Documentation

Document status: Final Amendment 2

Release date: 27-Sep-2017

Number of pages: 78

Property of Novartis
Confidential
May not be used, divulged, published or otherwise disclosed
without the consent of Novartis

Document history – LEE011E2301 Statistical Analysis Plan (SAP).

Version	Date	Changes
1.0	18-Feb-2015	First approved version
2.0	5-Jun-2017	Amendment 1 Amendment rationale: to incorporate changes to the analysis plan due to protocol amendments and implement changes to align with Novartis guidelines and program standards Main changes: • Change document name to Statistical Analysis Plan (SAP) for alignment with updated process • Remove interim futility and efficacy analyses for PFS to reflect protocol amendments • Update analyses based on BIRC assessment of imaging data to reflect change to an audit-based assessment • [redacted] • Update definition of Safety Set to reflect the latest Novartis Oncology standard definition Specific other changes • Clarified that Asian vs. non-Asian analyses are based on race, not ethnicity • Change “LEE011” to “ribociclib” in multiple places throughout the document Section 1 • Update introduction to refer to the protocol version the SAP is based on Section 2.1 • Update last contact date derivation details based on the latest SAP template and the data panels available in the study Section 2.3 • Update definition of per protocol set to include requirement of at least one dose and remove requirement of evaluable for efficacy • Update list of protocol deviations leading to exclusion from per protocol set • [redacted] Section 2.4 • Update section title to refer to therapies since it’s not limited to medications • Remove detailed list of prohibited medications and replace with reference to protocol • Expand on description of specific impact of concomitant therapies on analyses and clarify here and elsewhere that palliative radiation and continue of combination therapy alone after end of study treatment will not be considered as new antineoplastic therapies Section 2.5 • Clarify definition of PD too late as reason for unknown BOR • Update calculation of missing tumor assessments to align with Novartis RECIST macro • Remove detailed algorithm for waterfall plot construction

Version	Date	Changes
		Section 3.2 • Clarify use of IRT data versus CRF data for variables related to stratification factors • Include definitions for disease-free interval and endocrine sensitivity • Add clarification for biopsy to be considered separately from other prior antineoplastic surgeries • Add further details for summary of prior antineoplastic therapies Section 3.5 • Add summary in disposition table for number of patients entering survival follow-up Section 3.6 • Clarify last date of exposure to goserelin for patients who died or were lost to follow-up • Update dose intensity calculations to use duration of exposure to treatment component rather than duration of exposure to study treatment • Adjust DI calculation for goserelin to consider the dosing schedule (once every 28 days) • Remove calculation of RDI on a per-patient basis across treatment components • Add further details for definitions of dose interruption, delay and reduction • Add specific definitions of partial discontinuation of ribociclib/placebo and goserelin Section 3.7 • Separate concomitant and post-treatment therapies • Add summary for first and second new systemic antineoplastic medication after discontinuation of study treatment Section 3.8 • Clarify how the different sources of overall lesion response are used • Specify the primary analysis will be based on FAS including all data up to the cutoff date • Add table of censoring rules to section 3.8.1.1 • Clarify that allocation of patients to NSAI/tamoxifen subgroups for supportive efficacy analyses are based on IRT stratification data • Increase the number of time points at which Kaplan-Meier estimates of survival probabilities are produced • Add details of audit-based BIRC assessment of PFS • Clarify details of multivariate Cox regression models for PFS and OS • Remove some of the minor supportive analyses for PFS, namely summary of difference in days to progression per local and central review, summary of missing tumor assessments, and time to first and second tumor assessment • [redacted] • Clarify summary of time to response for responders is a descriptive summary • Clarify parts of the definition of time to definitive deterioration in ECOG PS

Version	Date	Changes
		• For EQ-5D-5L PRO, change “index score” to “visual analog scale” for consistency with the protocol • Clarify the repeated measures analyses of PROs Section 3.9 • Clarify conventions for AE reporting • Update list of AE tables to be produced • Change “Clinically notable AEs” to “AEs of special interest” to match standard definition, and clarify/update description of these • Include some additional details regarding laboratory parameter and CTC grade derivations • Update list of liver function test criteria to be summarized and clarified that combined categories need not be concurrent • Include box plot of laboratory parameters • Add analyses of time to first occurrence of laboratory toxicities • Add description of analysis to be performed for urinalysis • Clarify how averaging of unscheduled ECGs is performed • Clarify notable ECG values to be summarized • Add analysis of time to grade 2 or worse QT prolongation • [redacted] Section 3.10 • [redacted] • [redacted] • Broaden definition of evaluable ribociclib/LEQ803 concentration to ribociclib doses of 200 mg, 400 mg, and 600 mg [redacted] ribociclib dose • [redacted] • [redacted] • Added linear model to summarize Ctrough • [redacted] • [redacted] • Remove PK-ECG model and add graphical summaries of PK-ECG and PK-liver function • [redacted] • [redacted] Section 3.11 • [redacted]

Version	Date	Changes
		• Clarify details of analyses to be performed Section 3.13 • [redacted] Section 3.14 • Update interim analysis descriptions due to protocol amendments [redacted] Section 3.16 – 3.18 • Update sample size calculation to reflect changes in protocol amendments Section 4 • Update analysis descriptions for clarity and remove extraneous details and SAS code from analysis descriptions • Clarified analysis for duration of follow-up

3.0	27-Sep-2017	Amendment 2
		Amendment rationale: to clarify some analysis conventions and terminology used
		
		List of changes:
		• Section 2.1.7: Clarify the definition of baseline as it relates to RECIST-based endpoints, to consider baseline valid if within seven days of treatment start date.
		• Section 2.3.2: Remove one redundant sentence from the definition of the Safety set to avoid potential confusion
		• Section 2.3: Include a section on withdrawal of informed consent to ensure data after withdrawal of consent are not included in analyses.
		• Section 3.2.2:
		  • Add number of metastatic sites summary to diagnosis and extent of cancer summary
		  • Simplify de novo/DFI definitions
		  • Change endocrine resistant/sensitive terminology to more generic wording
		• Section 3.8.1.5:
		  • Specify that NSAI/tamoxifen subgroup analyses will be based on assigned combination partner per CRF, for consistency with other subgroup analyses. The analysis according to IRT stratification factor are part of a separate analysis.
		  • Change endocrine resistant/sensitive terminology as described for Section 3.2.2
		• Section 3.8.2.2: Change endocrine resistant/sensitive terminology as described for Section 3.2.2
		• Section 3.11.6.2: Remove reference to “each time point” since only baseline data are included
		• [redacted]

Version	Date	Changes
		• Section 3.13: Delete bullet point indicating PFS [redacted] may be identical if adequate tumor assessments continue after initiation of next-line therapy, to avoid potential confusion [redacted] Remove no baseline assessment as censoring reason. • [redacted] • Section 4.2.5: • Increase the number of samples for bootstrap correlation estimate. Specify window for comparison of timing of response assessments.

Table of contents

Table of contents 7
List of tables 11
List of figures 11
1 Introduction 14
1.1 Study Design 14
1.2 Objectives 15
2 Definitions and general methodology 17
2.1 Definitions 17
2.1.1 Study drug and study treatment 17
2.1.2 Date of first administration of study drug 17
2.1.3 Date of last administration of study drug 17
2.1.4 Date of first administration of study treatment 17
2.1.5 Date of last administration of study treatment 18
2.1.6 Study day 18
2.1.7 Baseline 18
2.1.8 On-treatment assessment/event 19
2.1.9 Last contact date 19
2.2 Data included in the analysis 20
2.3 Analysis sets 20
2.3.1 Full analysis set (FAS) 20
2.3.2 Safety set 20
2.3.3 Per protocol set 21
[redacted] 21
2.3.5 Patient classification 21
2.4 Concomitant therapies with specific impact on the analysis 22
2.5 Implementation of RECIST 23
2.5.1 Overall lesions response for patients with only non-measurable lesions at baseline 23
2.5.2 Disease progression 23
2.5.3 Best overall response (BOR) 23
2.5.4 Change in imaging modality 24
2.5.5 Determination of missing adequate assessments 25
2.5.6 No baseline tumor assessments 26
2.5.7 Construction of waterfall graphs 27

3 Statistical methods used in reporting 28
3.1 Enrollment status 28
3.2 Background and demographic characteristics 28
3.2.1 Basic demographic and background data 28
3.2.2 Diagnosis and extent of cancer 28
3.2.3 Medical history 29
3.2.4 Prior anti-neoplastic therapy 29
3.2.5 Other 29
3.3 Protocol deviation summaries 29
3.4 Groupings for analysis 30
3.5 Patient disposition 30
3.6 Study treatment 30
3.6.1 Duration of study treatment exposure 30
3.6.2 Cumulative dose and average daily dose 31
3.6.3 Dose intensity and relative dose intensity 31
3.6.4 Dose reductions, interruptions and delays 32
3.6.5 Discontinuation of study treatment components 33
3.7 Concomitant and post treatment therapy 33
3.8 Efficacy evaluation 34
3.8.1 Primary efficacy 35
3.8.2 Secondary efficacy analyses 40
3.9 Safety evaluation 47
3.9.1 Adverse events (AEs) 47
3.9.2 Laboratory data 50
3.9.3 Vital signs 52
3.9.4 ECG 53
3.9.5 Cardiac imaging (MUGA / ECHO) 54
3.9.6 Other safety data 54
[redacted] 54
[redacted] 54
[redacted] 56
[redacted] 57
[redacted] 58
[redacted] 59
[redacted] 59

[redacted] 60
[redacted] 60
[redacted] 60
[redacted] 61
[redacted] 61
[redacted] 62
[redacted] 62
[redacted] 63
[redacted] 63
[redacted] 63
[redacted] 63
[redacted] 63
[redacted] 64
3.14 Interim analyses 64
3.14.1 Primary Endpoint: PFS 64
3.14.2 Key Secondary Endpoint: OS 65
3.14.3 Confidentiality of Interim OS results 65
[redacted] 65
[redacted] 66
[redacted] 67
3.16 Sample size calculation 68
3.17 Power for analysis of key secondary variables 69
[redacted] 69
[redacted] 69
4 Details of the statistical analysis 70
4.1 Baseline comparability 70
4.2 Time-to-event analyses 70
4.2.1 Analysis of time-to-event data with ties 70
4.2.2 Hazard ratio 70
4.2.3 Hypothesis and test statistic 71
4.2.4 Kaplan-Meier estimates 71
4.2.5 Audit-based BIRC assessment of PFS 71
4.3 Group sequential design used in Phase III studies 72
4.3.1 Alpha-spending function 73
4.3.2 Methodology 73
4.3.3 Calculation of stopping boundaries 74

4.4 Duration of follow-up 74
4.5 Confidence intervals for response rate and clinical benefit rate 74
5 References 75
Appendix 1 CTC grades for laboratory values in Novartis Oncology 77

List of tables

Document history – LEE011E2301 Statistical Analysis Plan (SAP). 2
Table 1-1 Study objectives 15
Table 2-1 Last contact date data sources 19
Table 2-2 Patient classification based on protocol deviations and non-protocol deviation criteria 21
Table 2-3 Schedule for tumor assessment and time windows 25
Table 2-4 Assessments considered for calculation of best percentage change for waterfall graphs 27
Table 3-1 Planned dose intensity 32
Table 3-2 Sources for overall lesion response 34
Table 3-3 Outcome and event/censor dates for PFS analysis 35
Table 3-4 ECOG Performance Scale 43
Table 3-5 Time windows for ECOG PS assessments 43
Table 3-6 Time windows for patient reported outcomes 44
Table 3-7 AESI groupings 50
Table 3-8 Clinically notable ECG values 53
[redacted] [redacted] 60
[redacted] [redacted] 64
Table 3-11 Timelines for interim and final analyses 65
Table 4-1 Local versus central disease progression assessments 72

List of figures

Figure 1-1 Study design 15

List of abbreviations

AE	Adverse event
ATC	Anatomical Therapeutic Chemical
AUC	Area Under the Curve
BOR	Best overall response
CI	Confidence Interval
CBR	Clinical benefit rate
CR	Complete response
CRF	Case Report Form
CRO	Contract Research Organization
CRS	Case retrieval strategy
CSR	Clinical study report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
ΔQTcF	Change from baseline in QTcF
DAR	Dosage administration Record
DI	Dose Intensity
DMC	Data Monitoring Committee
DOR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOT	End of treatment
FAS	Full analysis set
HER2	Human epidermal growth factor receptor 2
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
IRT	Interactive Response Technology
LFT	Liver Function Test
LLOQ	Lower Limit of Quantification
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
NCI	National Cancer Institute
NMQ	Novartis MedDRA queries
NSAI	Non-steroidal aromatase inhibitor
ORR	Overall response rate
OS	Overall survival
PAS	Pharmacokinetic analysis set
PD	Progressive disease

PDI	Planned dose intensity
PDS	Programming Datasets Specifications
PFS	Progression-free survival
PhRMA	Pharmaceutical Research and Manufacturers of America
PK	Pharmacokinetic
PPS	Per Protocol Set
PR	Partial response
PRO	Patient Reported Outcome
PS	Performance Status
PT	Preferred term
QTcF	QT interval corrected by Fridericia method
RAP	Report and Analysis Process
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria In Solid Tumors
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SD	Standard deviation
SEC	Safety Event Categories
SMQ	Standardized MedDRA queries
SOC	System organ class
TA	Tumor assessment
TBIL	Total Bilirubin
TEAE	Treatment-emergent adverse event
TFLs	Tables, Figures, Listings
UNK	Unknown

1 Introduction

This document describes the detailed statistical methodology to be used for the Clinical Study Report (CSR) for the primary PFS analysis of study CLEE011E2301, a phase III, randomized, double-blind, placebo-controlled study of LEE011 (ribociclib) or placebo in combination with tamoxifen and goserelin or a non-steroidal aromatase inhibitor (NSAI) and goserelin for the treatment of premenopausal women with hormone receptor positive, HER2-negative, advanced breast cancer.

The content of this SAP is based on protocol amendment 4 (clee011e2301-v04--protocol). All decisions regarding final analysis, as defined in the SAP document, have been made prior to database lock and unblinding of the study data.

CSR deliverables (shells for tables, figures, listings) and further programming specifications are described in the Tables, Figures & Listings (TFL) shells and Programming Datasets Specifications (PDS), respectively.

1.1 Study Design

This is a phase III randomized, double-blind, placebo-controlled global study comparing the combination of tamoxifen or a NSAI (letrozole or anastrozole) + goserelin + LEE011 to tamoxifen or a NSAI (letrozole or anastrozole) + goserelin + placebo in premenopausal women with HR+, HER2-negative advanced breast cancer.

Approximately 660 patients will be randomly assigned to one of the below treatment arms in 1:1 ratio:

a. [Tamoxifen or a NSAI (letrozole or anastrozole)] + goserelin + LEE011
   OR,
b. [Tamoxifen or a NSAI (letrozole or anastrozole)] + goserelin + placebo.

Randomization will be stratified by the presence of liver and/or lung metastases (yes versus no), prior chemotherapy for advanced disease (yes versus no), and endocrine combination partner (tamoxifen and goserelin versus NSAI (letrozole or anastrozole) and goserelin).

Treatment with tamoxifen or NSAI in this study will be based on the patient's prior (neo) adjuvant therapy for breast cancer as outlined in the Clinical Study Protocol. The choice of NSAI (letrozole or anastrozole) is left to the discretion of the investigator. Once assigned, changing the dose or type of endocrine therapy is not permitted.

PFS, as assessed by the local radiologists/investigators and using RECIST 1.1 criteria will be the primary endpoint. PFS as assessed through audit-based blinded independent central review will be used for supportive evidence of the primary efficacy endpoint.

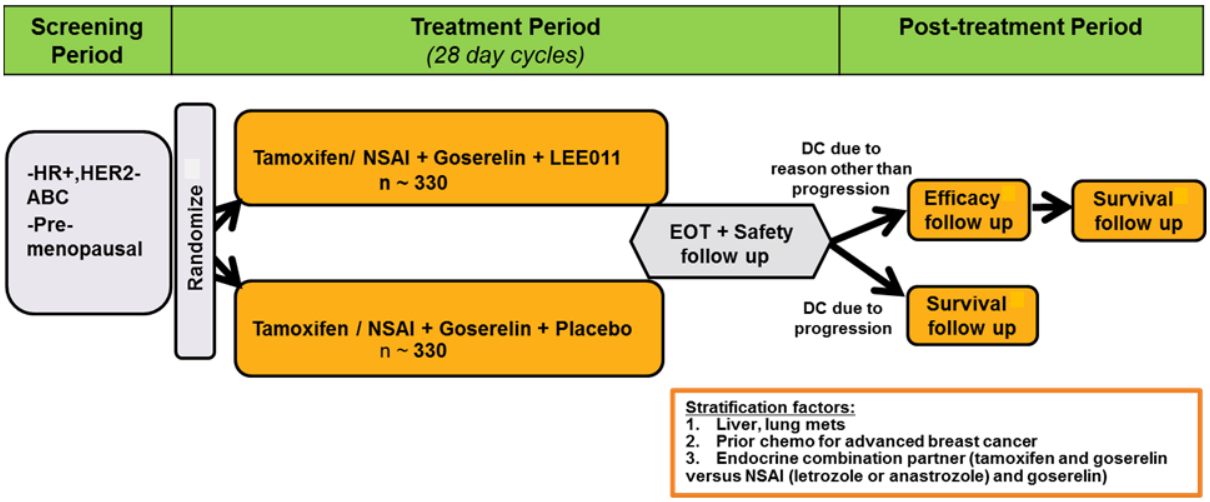
An independent Data Monitoring Committee (DMC) will monitor unblinded safety and PK data during the trial. Separate DMC RAP documents specify the analyses to be performed for the DMC reviews.

Overall survival is a key secondary endpoint in this study and will be analyzed provided the primary endpoint PFS is statistically significant. A Lan-DeMets alpha spending function with O'Brien Fleming type stopping boundary (as implemented in East 6.3) will be used to maintain

the overall type-I error rate for OS. The interim OS analyses will be performed by Novartis. Further details regarding the group sequential design are provided in Section 3.14.

The study design is summarized in Figure 1-1.

Figure 1-1 Study design



1.2 Objectives

The study objectives and corresponding endpoints as specified in the protocol are provided in Table 1-1.

Table 1-1 Study objectives

Objective	Endpoint	Analysis
Primary		
To determine whether treatment with tamoxifen or a NSAI + goserelin + LEE011 prolongs PFS compared to treatment with tamoxifen or a NSAI + goserelin + placebo in premenopausal women with HR+, HER2-negative advanced breast cancer	PFS per local assessment and RECIST 1.1	Section 3.8.1
Key secondary		
To determine whether treatment with tamoxifen or a NSAI + goserelin + LEE011 prolongs OS compared to treatment with tamoxifen or a NSAI + goserelin + placebo in premenopausal women with HR+, HER2-negative advanced breast cancer.	OS	Section 3.8.2.1
Other secondary		

Objective	Endpoint	Analysis
To evaluate the safety and tolerability of LEE011 in combination with tamoxifen + goserelin or a NSAI + goserelin.	Frequency/ severity of adverse events, lab abnormalities	Section 3.9
To evaluate the two treatment arms with respect to response rate and clinical benefit rate (CBR).	ORR as defined by RECIST 1.1. CBR, defined as percentage of patients with CR, PR per RECIST 1.1 or SD lasting 24 weeks or longer	Section 3.8.2.2 and 3.8.2.3
To describe time to response (TTR) and duration of response (DOR) in each treatment arm.	Time to response and duration of response per RECIST 1.1	Section 3.8.2.4 and 3.8.2.5
To evaluate the two treatment arms with respect to time to deterioration of ECOG PS.	Time to definitive deterioration of the ECOG PS from baseline	Section 3.8.2.6
To evaluate patient reported outcomes for health-related quality of life in the two treatment arms.	Time to 10% deterioration in the global health status/QOL scale score of the EORTC QLQ-C30 Change from baseline in the global health status/QOL scale score of the EORTC QLQ-C30	Section 3.8.2.7
[redacted]		

[redacted]

2 Definitions and general methodology

2.1 Definitions

2.1.1 Study drug and study treatment

Study drug is defined as ribociclib or matching placebo.

Study treatment is defined as ribociclib + (tamoxifen or a NSAI) + goserelin, or matching placebo + (tamoxifen or a NSAI) + goserelin.

2.1.2 Date of first administration of study drug

The date of first administration of study drug is defined as the first date when a nonzero dose of study drug is administered and recorded on the dose administration DAR CRF. The date of first administration of study drug will also be referred to as start of study drug. Similar definitions apply for the other components of study treatment.

2.1.3 Date of last administration of study drug

The date of last administration of study drug is defined as the last date when a nonzero dose of study drug is administered and recorded on the DAR CRF date. Similar definitions apply for the other components of study treatment.

2.1.4 Date of first administration of study treatment

The date of first administration of study treatment is defined as the first date when a nonzero dose of any component of study treatment is administered and recorded on the DAR CRF. The

date of first administration of study treatment will also be referred to as the start of study treatment.

2.1.5 Date of last administration of study treatment

The date of last administration of study treatment is defined as the last date when a nonzero dose of any component of study treatment was administered and recorded on the DAR CRF.

2.1.6 Study day

The study day will be calculated as:

- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date + 1 if event is on or after the reference start date;
- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date if event precedes the reference start date.

The reference start date for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption etc.) is the start of study treatment. (Note: if an adverse event starts before the start of study treatment the study day displayed on the listing will be negative).

The reference start date for all other, non-safety assessments (i.e., tumor assessment, death, disease progression, tumor response, ECOG performance status, and patient reported outcomes (PRO)) is the date of randomization. In other words, all efficacy time-to-event variables (e.g. progression-free survival, overall survival, time to response) will be calculated from date of randomization. (Example: if randomization date is 15DEC2014, start of study drug is on 18DEC2014, and the date of death is 28DEC2014 then the study day when the death occurred is 14).

The study day will be displayed in data listings.

2.1.7 Baseline

For efficacy evaluations, the last available assessment on or before the date of randomization will be used as the “baseline” value or “baseline” assessment. In the context of baseline definition, the efficacy evaluations also include ECOG performance status and patient-reported outcomes. For RECIST-based endpoints (e.g. PFS, best overall response), a window of 7 days after the start of study treatment will be allowed, i.e. the baseline will also be considered valid if it occurs within 7 days of the treatment start date.

For safety evaluations (e.g. laboratory assessments and ECG), the last available assessment before or at date of start of study treatment will be used as the ‘baseline’ assessment. Assessments specified to be collected post-dose on the first date of treatment are not considered as baseline values.

If patients have no value as defined above, the baseline results will be considered missing.

2.1.8 On-treatment assessment/event

Safety summaries and selected summaries of deaths will summarize only on-treatment assessments/events. An on-treatment assessment/event is defined as any assessment/event in the following time interval:

[date of first administration of study treatment, date of last administration of study treatment + 30 days], i.e. including the lower and upper limits. (Note: However, the calculation of study treatment duration will use different rules as specified in Section 3.6.1).

An AE started in the screening phase and ongoing in the on-treatment phase will not be considered an on-treatment AE unless it has worsened.

If the last date of study treatment is missing, any assessment/event occurring after the start of study treatment will be considered as on-treatment.

Data listings will include all assessments/events, flagging those which are not on-treatment.

Note: The date of first administration of study treatment and the date of last administration of study treatment are defined in Sections 2.1.4 and 2.1.5, respectively.

2.1.9 Last contact date

For patients not known to have died at the analysis cut-off the last contact date will be derived using the last complete date among the following:

Table 2-1 Last contact date data sources

Source data	Conditions
Date of Randomization	No Condition
Last date patient was known to be alive from Survival Follow-up page	Patient status is reported to be alive, lost to follow-up or unknown.
Start/End dates from further antineoplastic therapy	Non-missing medication/procedure term.
Start/End dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
Date of discontinuation/study phase completion from end of treatment or end of post-treatment follow-up disposition page	No condition.
Date of ECG assessment	At least one non-missing parameter value
Date of PRO assessment	At least one non-missing answer to questionnaire
Tumor (RECIST) assessment date	For non-target lesion: non-missing lesion status For target lesion: non-missing lesion diameter For new lesion: “Is there a new lesion?” = yes
Laboratory collection dates	At least one non-missing parameter value
[redacted]	[redacted]
Vital signs date	At least one non-missing parameter value

Source data	Conditions
Performance Status date	Non-missing performance status
Start/End dates of AE	Non-missing verbatim term
Concomitant medication date	Non-missing medication
[redacted]	[redacted]
Body fluid/Tissue Collection/Results	Non-missing result (positive/negative tumor cells)
[redacted]	[redacted]
Cardiac imaging date	Non-missing LVEF or overall interpretation

The last contact date is defined as the latest complete date from the above list on or before the data cut-off date. The cut-off date will not be used for last contact date, unless the patient was seen or contacted on that date. No date post cut-off date will be used. Completely imputed dates (e.g. the analysis cut-off date programmatically imputed to replace the missing end date of a dose administration record) will not be used to derive the last contact date. For patients who died, the last contact date is defined as the date of death. Partial date imputation is allowed for event (death)/censoring from the 'Survival information' eCRF.

The last contact date will be used for censoring of patients in the analysis of overall survival.

2.2 Data included in the analysis

The primary PFS endpoint will be carried out after approximately 329 PFS events have been documented based on local assessment, and up to 3 analyses for OS may be performed as described in Section 3.8.2.1.

For each of the analyses, cutoff dates will be established and statistical analyses will be performed using all data collected in the database up to the data cutoff date. Any data collected beyond the cutoff date will not be included in the analysis and will not be used for any derivations.

2.3 Analysis sets

2.3.1 Full analysis set (FAS)

The Full Analysis Set (FAS) consists of all randomized patients. Following the intent-to-treat principle, patients will be analyzed according to the treatment and stratum they were assigned to at randomization. The FAS will be the primary analysis set for efficacy analyses.

2.3.2 Safety set

The Safety Set consists of all patients who received at least one dose of any component of study treatment. Patients will be analyzed according to the treatment actually received. Treatment

actually received refers to the treatment the patient was randomized to, unless the alternative treatment was received throughout the trial.

2.3.3 Per protocol set

The Per-Protocol Set (PPS) includes the subset of the patients in the FAS without any major protocol deviation who took at least one dose of study treatment. Patients with any of the following protocol deviations will be excluded from the PPS.

- Written informed consent not obtained;
- Patient without HR+ and HER2- advanced breast cancer at baseline;
- Patient is postmenopausal;
- Patient received prior CDK4/6 inhibitor;
- Baseline ECOG performance status>1;
- Neither measurable disease nor predominantly lytic bone lesion at baseline;
- Patient received any prior hormonal anti-cancer therapy for advanced breast cancer, except for ≤ 14 days of tamoxifen or a NSAI ± goserelin or ≤ 28 days of goserelin prior to randomization;
- Patient received >1 line of chemotherapy for advanced breast cancer;
- Patient received different treatment throughout the study than the one randomized to;
- Inadequate documentation available to support patient eligibility;
- Patient has inflammatory breast cancer;
- Patient has CNS metastases.

A sensitivity analysis of the primary endpoint (i.e., PFS) will be performed using PPS if the primary efficacy analysis is significant and the FAS and PPS differ.

[redacted]

2.3.5 Patient classification

Patients may be excluded from the analysis sets defined above based on the protocol deviations entered in the database and/or on specific subject classification rules as defined in Table 2-2.

Table 2-2 Patient classification based on protocol deviations and non-protocol deviation criteria

Analysis set	Protocol deviations leading to exclusion	Non-protocol deviation criteria leading to exclusion
FAS	No written informed consent	NA
Safety set	No written informed consent	No dose of study treatment

Per Protocol set	Any major protocol deviation as listed in definition of per protocol set	No dose of study treatment

Withdrawal of informed consent

Any data collected in the clinical database after a patient withdraws informed consent from all further participation in the trial, will not be included in the analysis. The date on which a patient withdraws full consent is recorded in the eCRF.

Death events may be used in the analysis if captured from public records (registers), local law and patient informed consent permiting.

[redacted]

2.4 Concomitant therapies with specific impact on the analysis

According to the study protocol, the following medications are prohibited during combined ribociclib, tamoxifen or a NSAI, and goserelin treatment in this study:

- Strong inhibitors or inducers of CYP3A4/5
- Strong inhibitors of CYP2D6 (that could affect tamoxifen exposure)
- Substrates of CYP3A4/5 with a narrow therapeutic index
- Medications with a known risk of QT prolongation
- Warfarin and coumarin derivatives
- Other investigational and antineoplastic therapies (with the exception of palliative radiotherapy)
- Herbal medications/preparations

These substances are listed in Table 14-1 in the study protocol. A corresponding list for programming purposes will be saved in a separate document. If there is an update to the list of prohibited medications (e.g., in a protocol amendment), the most up-to-date list shall be used for the Clinical Study Report.

However, some patients may take these substances during the treatment period so these concomitant medications will be selected via programming and tabulated and listed in the Clinical Study Report.

Treatment with the prohibited substances mentioned above will be identified in the database as protocol deviations.

With the exception of palliative radiotherapy, administration of anti-neoplastic therapies (apart from study treatment) is not allowed during study treatment. Patients who take anti-neoplastic therapies after randomization but before end of treatment will be identified through data review. Tumor assessments (TAs) made after the start of anti-neoplastic therapies (whether on study treatment or afterwards) will not be included in the efficacy analyses based on best overall response (i.e. Overall Response Rate (ORR), Clinical Benefit Rate (CBR), time to response, and duration of response (Section 3.8.2.2 to 3.8.2.3). In addition, a sensitivity analysis will be

performed for the primary efficacy PFS analysis in which PFS will be censored at the last tumor assessment before the start of new anti-neoplastic therapy. Further details are provided in Section 3.8.1. Time to event analyses based on ECOG performance status and patient reported outcomes will not consider assessments after the start of new anti-neoplastic therapy. Further details are provided in Section 3.8.2.6 and 3.8.2.7. For these analyses, the following will not be considered as new antineoplastic therapies: (i) palliative radiotherapy; (ii) continuation of combination partner therapy alone after end of study treatment.

2.5 Implementation of RECIST

Response and progression evaluation will be performed according to the Novartis RECIST guideline (as described in detail in Appendix 3 of the Clinical Study Protocol), which is based on the RECIST 1.1 guidelines (Eisenhauer et al 2009). The text below gives instructions and rules to provide details needed for programming.

2.5.1 Overall lesions response for patients with only non-measurable lesions at baseline

Patients without measurable disease per RECIST 1.1 are eligible to enter the study if they have at least one predominantly lytic bone lesion. For patients with non-measurable lesions only at baseline, the overall lesion response will be based solely on non-target lesion response or an occurrence of a new lesion. Non-measurable lesions will be entered as non-target lesions. Therefore, the best overall response is determined from non-target lesion response and presence of new lesions (refer to Table 3-1 in Section 3.2.9 of RECIST Novartis guidelines as described in detail in Appendix 3 of the Clinical Study Protocol).

Note: Pathologic fracture, new compression fracture, or complications of bone metastases will not be considered as evidence of disease progression, unless there is unequivocal progression of existing non-target lesions or a new lesion.

2.5.2 Disease progression

Progressive disease should only be assigned if it is confirmed by an assessment method as per RECIST 1.1 guidelines (e.g. radiologic scan, photos for skin lesions, etc). If a new lesion is detected using an objective assessment method other than radiologic scan then it should also be entered on the 'New lesion' RECIST CRF with appropriate method. Discontinuation due to disease progression or death due to study indication, without corresponding supportive data in the RECIST CRF (as defined above), will not be considered as progressive disease in the calculation of best overall response and in the analysis of PFS.

2.5.3 Best overall response (BOR)

The best overall tumor response will be assessed per RECIST 1.1 criteria. The definitions and the details on the derivation are given in Appendix 3 of the Clinical Study Protocol.

Only tumor assessments performed before the start of any further anti-neoplastic therapies (i.e. any additional anti-neoplastic therapy or surgery) and within 30 days after the last administration of study treatment will be considered in the assessment of best overall response. Further anti-neoplastic therapies will be identified from the data collected on 'Anti-neoplastic

therapies since discontinuation of study treatment' CRFs. Palliative radiotherapy is the only setting of radiotherapy allowed during the study. Therefore, palliative radiotherapy will not be considered as an anti-neoplastic therapy for assessment of BOR. Continuation of combination partner therapy alone after end of study treatment will also not be considered as a new anti-neoplastic therapy.

The standard definition of a best overall response evaluation of 'stable disease', 'disease progression' or 'unknown' given in the Appendix 3 of the Clinical Study Protocol will be used for this study. Best overall response for each patient is determined from the sequence of overall (lesion) responses (as reported by the investigator for local BOR, and as reported by BIRC for central BOR) according to the following rules:

- CR = at least two determinations of CR at least 4 weeks apart before progression.
- PR = at least two determinations of PR or better at least 4 weeks apart before progression (and not qualifying for a CR).
- SD = at least one SD assessment (or better) > 6 weeks after randomization (and not qualifying for CR or PR).
- Non-CR/non-PD = at least one non-CR/non-PD assessment (or better) > 6 weeks after randomization date (and not qualifying for CR). This applies only for patients with non-measurable disease alone at baseline.
- PD = progression ≤ 12 weeks after randomization (and not qualifying for CR, PR, SD or Non-CR/non-PD).
- UNK = all other cases (i.e. not qualifying for confirmed CR or PR and without SD or Non-CR/Non-PD after more than 6 weeks or progression within the first 12 weeks).

Patients with best overall response "unknown" will be summarized by reason for having unknown status. The following reasons will be used:

- No valid post-baseline assessment
- All post-baseline assessments have overall response UNK
- New anti-neoplastic therapy started before first post-baseline assessment
- SD too early (≤6 weeks after randomization)
- PD too late (>12 weeks after randomization and not qualifying for CR, PR, SD or Non-CR/Non-PD)

Special (and rare) cases where BOR is unknown due to both early SD and late PD will be classified as "SD too early".

2.5.4 Change in imaging modality

Per RECIST 1.1, a change in methodology can be defined as either a change in contrast use (e.g. keeping the same technique, like CT, but switching from with to without contrast use or vice-versa, regardless of the justification for the change) or a major change in technique (e.g. from CT to MRI, or vice-versa), or a change in any other imaging modality. A change from conventional to spiral CT or vice versa will not constitute a major "change in method" for the purposes of response assessment. A change in methodology will result by default in an UNK (unknown) overall lesion response based on the Novartis calculation. However, a response from

the investigator or the central blinded reviewer that differs from the Novartis calculated UNK is acceptable if a definitive response assessment can be justified based on the available information.

Potential discrepancies between the modality used and overall lesion response (e.g. change in modality but response is different from 'Unknown') will be queried during the data validation process.

2.5.5 Determination of missing adequate assessments

The term 'missing adequate assessment' refers to assessments that are not done or for which the overall lesion response is 'Unknown'. 'Missing adequate assessment' will also be referred to as 'missing assessment'.

As detailed in Section 3.2.10 and in Appendix 3 of the Clinical Study Protocol, the PFS censoring and event date options depend on the presence and the number of missing tumor assessments. An event occurring after two or more missing assessments is censored at the last adequate tumor assessment.

An exact rule to determine whether there are no, one or two missing TAs is therefore needed. This rule is based on the interval between the last adequate tumor assessment (LATA) date and the event date. The scheduled date of tumor assessments (in weeks from randomization), protocol specified window for tumor assessments, and the thresholds for LATA to belong to a visit can be found in the following table.

Table 2-3 Schedule for tumor assessment and time windows

Assessment schedule		Scheduled date – 1 week	Scheduled date (weeks from randomization)	Scheduled date +1 week	Threshold*
Every 8 weeks for the first 18 months	Baseline	0	0	0	0
	C3D1	7	8	9	12
	C5D1	15	16	17	20
	C7D1	23	24	25	28
	C9D1	31	32	33	36
	C11D1	39	40	41	44
	C13D1	47	48	49	52
	C15D1	55	56	57	**60**
	C17D1	63	64	65	**68**
	C19D1	71	72	73	78
Every 12 weeks after 18 months	C22D1	83	84	85	90
	C25D1	95	96	97	102
	C28D1	107	108	109	114
	C31D1	119	120	121	126

* The mid-point between current and next visit (except for baseline) and the upper limit for LATA to be matched to a certain scheduled assessment, e.g. if LATA is at week 13, this is after threshold for C3D1 and before that for C5D1, so the matching scheduled assessment is C5D1.

To calculate the number of missing tumor assessments, the LATA before an event is matched with a scheduled tumor assessment using the time window in Table 2-3 (essentially whichever scheduled assessment it is closest to). Two thresholds, D1 and D2 are calculated for that scheduled assessment based on the protocol-specified schedule and windows

- An event after LATA+D1 will be considered as having >=1 missing assessment
- An event after LATA+D2 will be considered as having >=2 missing assessments

Since there is a change of schedule for tumor assessments at 18 months, D1 and D2 are defined differently depending on when LATA happens.

Rule 1: if LATA happens within 60 weeks from randomization (the matched scheduled tumor assessment is C15D1 or before)

- D1=8+2=10 weeks
- D2=2*8+2=18 weeks

Rule 2: if LATA happens after 60 weeks but within 68 weeks from randomization (the matched scheduled tumor assessment is C17D1)

- D1=8+2=10 weeks
- D2=8+12+2=22 weeks

Rule 3: if LATA happens after 68 weeks from randomization (the matched scheduled tumor assessment is C19D1 or later)

- D1=12+2=14 weeks

D2=2*12+2=26 weeks.

Therefore, using the D2 definition above, the censoring of an event occurring after ≥2 missing TAs (in the primary PFS analysis) can be refined as follows: if the distance between the last adequate TA date and the PFS event date is larger than D2 then the patient will be censored and the censoring reason will be 'Event documented after two or more missing tumor assessments'.

The same definition of D2 will be used to determine the PFS censoring reason. If the distance between the last adequate tumor assessment date and the earliest of the following dates:

1. Analysis cut-off date
2. Date of consent withdrawal
3. Date of loss to follow-up

is less than or equal to D2 then the censoring reason will be 1. 'Ongoing without event', 2. 'Withdrew consent', or 3. 'Lost to follow-up', respectively. However, if this distance is larger than D2 then the censoring reason will be 'Adequate assessment no longer available'. If an event is documented after two missing assessments, then the censoring reason will be 'Event documented after two or more missing tumor assessments'.

2.5.6 No baseline tumor assessments

As specified in Table 3-2 in Section 3.2.10 of the Appendix 3 of the Clinical Study Protocol, since the timing of disease progression cannot be determined for patients with missing baseline tumor assessment, these patients are censored in the PFS analysis at the date of randomization.

This rule however only applies to the disease progression component of the PFS assessment, and not to the survival component.

Patients without baseline tumor assessments who die within D2 distance (see Section 2.5.5 for definition) of randomization will be counted as having an event in the primary analysis of PFS at the date of death (Note: all deaths will be counted in the overall survival analysis regardless of presence or absence of the baseline tumor assessment).

These patients will be excluded from the PPS for the analysis of PFS as defined in Section 2.3

2.5.7 Construction of waterfall graphs

Waterfall graphs will be used to depict the anti-tumor activity. These plots will display the best percentage change from baseline in the sum of diameters of all target lesions for each patient. Only patients with measurable disease at baseline will be included in the waterfall graphs.

Special consideration is needed for assessments where the target lesion response is CR, PR or SD, but the appearance of a new lesion or a worsening of non-target lesions results in an overall lesion response of PD. As a conservative approach, such assessments will not be considered for display as bars in the graph, since the percentage change in the sum of diameters of target lesions reflects the non-PD target lesion response, but the overall lesion response is PD. A patient with only such assessments will be represented by a special symbol (e.g. ✶) in the waterfall graph.

Assessments with "unknown" target lesion response and assessments with unknown overall response will be excluded from the waterfall plots. Patients without any valid assessments will be completely excluded from the graphs.

The total number of patients displayed in the graph will be shown and this number will be used as the denominator for calculating the percentages of patients with tumor shrinkage and tumor growth. Footnote will explain the reason for excluding some patients (due to absence of any valid assessment).

All possible assessment scenarios are described in Table 2-4.

Table 2-4 Assessments considered for calculation of best percentage change for waterfall graphs

Case	Target response	Overall lesion response	Calculate % change from baseline in sum of diameters?
1	UNK	Any	No, exclude assessment
2	Any	UNK	No, exclude assessment
3	CR/PR/SD	PD	No, flag assessment with ✶
4	PD	PD	Yes
5	CR/PR/SD	CR/PR/SD	Yes

3 Statistical methods used in reporting

3.1 Enrollment status

Enrollment by country and center will be summarized for all screened patients and also by treatment arm using the FAS. The reasons for screen failure will also be summarized.

3.2 Background and demographic characteristics

The FAS will be used for all baseline disease characteristics and demographic summaries and data listings.

3.2.1 Basic demographic and background data

Demographic and background disease characteristics data will be listed in detail. Qualitative data (e.g. race, ECOG performance status, etc.) will be summarized by means of contingency tables by treatment arm and quantitative data (e.g. age, body weight, etc) will be summarized by appropriate descriptive statistics (mean, standard deviation, median, minimum, and maximum) by treatment arm.

Discrepancies between randomization stratification information (obtained from the Interactive Response Technology (IRT) system) and stratum information based on data collected on CRFs will be tabulated and listed.

Unless otherwise specified, stratified analyses and analyses “by stratum” will be based on IRT stratification data, while data for other subgroup analyses and model covariates will be based on CRF data.

3.2.2 Diagnosis and extent of cancer

Summary statistics will be tabulated for diagnosis and extent of cancer. This analysis will include the following: primary site of cancer, histological grade, stage at initial diagnosis, time since initial diagnosis, stage at time of study entry, presence/absence of target and non-target lesions, number and type of metastatic sites involved, HER-2 / estrogen / progesterone receptor status, disease-free interval (DFI) for non-de novo patients, and a categorization of prior (neo-)adjuvant endocrine therapy (ET) as outlined below:

- De novo patients will be identified as those with no date of first recurrence/progression or the first recurrence/progression is within 90 days of initial diagnosis without prior antineoplastic medication (other than the short period of hormonal therapy for advanced breast cancer permitted by the protocol).
- DFI for non-de novo patients will be calculated as the time from initial diagnosis to first recurrence/progression and categorized as ≤12 months and >12 months.
- Patients will be grouped as follows based on prior (neo-)adjuvant ET:
  - No prior (neo-)adjuvant endocrine therapy
  - Progression while on or within 12 months of end of (neo-)adjuvant ET
  - Progression > 12 months after end of (neo-)adjuvant ET.

Time since initial diagnosis will be summarized in months. A month is defined as 365.25/12=30.4375 days.

3.2.3 Medical history

Medical history and ongoing conditions, including cancer-related conditions, will be summarized and listed. Separate summaries will be presented for ongoing and historical medical conditions. The summaries will be presented by primary system organ class and preferred term. Medical history/current medical conditions are coded using the latest version of Medical Dictionary for Regulatory Activities (MedDRA) terminology available at the time of the analyses. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable tables/listings.

3.2.4 Prior anti-neoplastic therapy

The number and percentage of patients recording any prior anti-neoplastic medications, radiotherapy or surgery (biopsy and non-biopsy separately) will be summarized by treatment arm both separately and in a combined fashion.

- Prior anti-neoplastic medications will be summarized by therapy type (e.g. chemotherapy, hormonal therapy etc), and also by lowest ATC class, preferred term and treatment. The total number of regimens along with the type (e.g. hormonal therapy), setting (e.g. adjuvant), best response and time from last treatment to progression for the last therapy will be summarized by treatment arm.
  - Hormonal therapy (tamoxifen/letrozole/anastrozole) for advanced breast cancer received for ≤ 14 days with or without goserelin, or treatment with goserelin for ≤ 28 days prior to randomization, as permitted by the study protocol, will be summarized separately.
  - Hormonal therapies in (neo-)adjuvant setting will also be summarized, including time from last dose of (neo-)adjuvant hormonal therapy to randomization.
- For radiotherapy, the setting (e.g. adjuvant, palliative) for the last therapy will be summarized.
- For surgery (excluding biopsies), the time since last surgery will be summarized by treatment arm.

Separate listings will be produced for prior anti-neoplastic medications, radiotherapy, and surgery.

3.2.5 Other

All data collected at baseline, including source of patient referral, child bearing potential and [redacted], will be listed.

3.3 Protocol deviation summaries

The number and percentage of patients in the FAS with any protocol deviation will be tabulated by deviation category (as specified in the Study Specification Document) and by treatment arm.

Protocol deviations leading to the exclusion from the per protocol set will be tabulated separately by treatment arm.

All protocol deviations, as well as protocol deviations leading to exclusion from the per protocol set, will be listed.

3.4 Groupings for analysis

The number and percentage of patients in each analysis set (definitions are provided in Section 2.3) will be summarized by treatment arm and randomization stratum.

3.5 Patient disposition

Patient disposition for all randomized patients will be summarized based on the FAS. There will be one combined by-treatment summary showing:

1. Number (%) of patients treated/untreated
2. Number (%) of patients who are still on-treatment (based on the absence of the 'End of treatment' page)
3. Number (%) of patients who discontinued study treatment (based on the 'End of Treatment' page)
4. Reasons for study treatment discontinuation (based on 'End of Treatment' page)
5. Number (%) of patients who entered the post-treatment evaluations (based on 'End of Treatment' page)
6. Number (%) of patients who discontinued from the post-treatment evaluations (based on 'End of post treatment follow up disposition' page)
7. Reasons for discontinuation from the post-treatment evaluations phase (based on 'End of post treatment follow up disposition' page)
8. Number (%) of patients who entered survival follow-up.

In a separate summary, the reasons for patients not completing the screening phase will be presented based on the "Screening Phase Disposition" CRF.

3.6 Study treatment

Duration of study treatment exposure, cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized by treatment. The number of patients with dose reductions/interruptions, and the reasons, will be summarized and listed. Details of the derivations and summaries are provided in the following sections.

The safety set will be used for all summaries and listings of study treatment.

3.6.1 Duration of study treatment exposure

The duration of exposure to study treatment will be calculated as

Duration of exposure to study treatment (days) = (***last date of exposure*** to any study treatment component) – (date of first administration of study treatment) + 1.

Duration of exposure to any single component of study treatment will be calculated as

Duration of exposure (days) = (last date of exposure to study treatment component) – (date of first administration of study treatment component) + 1.

The ***last date of exposure*** is defined as follows for the study treatment components:

- For ribociclib/placebo, tamoxifen, anastrozole, and letrozole: the last date of exposure is defined as the date of last administration of the corresponding medication;

- For goserelin, the last date of exposure is defined as the date of last administration of goserelin + 27 days.
  - If a patient died or was lost to follow-up within date of last administration + 27 days, then the last date of exposure is the date of death or last contact date, respectively.

This duration of exposure includes the periods of temporary interruption (of any component of the study treatment for any reason). The duration of study treatment exposure and exposure to each treatment component will be summarized by treatment arm. In addition, the duration of exposure will be categorized into time intervals (e.g. <3 months; 3-<6 months; 6-<9 months, etc); frequency counts and percentages will be presented for the number of patients in each interval.

3.6.2 Cumulative dose and average daily dose

Cumulative dose for any component of study treatment is defined as the total dose of the medication given during the study treatment exposure.

Average daily dose is defined as [Cumulative dose (mg) / Number of dosing days]; drug free day(s) are not counted as dosing days.

Cumulative dose and average daily dose will be summarized using descriptive statistics by treatment arm for each component of study treatment. Patients with no exposure to the study treatment component will be excluded from the corresponding summary.

3.6.3 Dose intensity and relative dose intensity

Dose intensity (DI) for patients with non-zero duration of exposure to each study treatment component is defined as follows:

For tamoxifen, letrozole, and anastrozole, DI is defined as

DI (mg/day) = Cumulative dose (mg) / duration of exposure to study treatment component (days).

For goserelin, DI is defined as

DI (mg/28 days) = Cumulative dose (mg) / {[(date of last administration of goserelin + 27) – (date of first administration of goserelin) + 1] / 28},

to account for the goserelin dosing schedule in the DI computation.

For ribociclib/placebo, DI is defined as

DI (mg/day) = Cumulative dose (mg) / *adjusted* duration of exposure to ribociclib/placebo (days),

where *adjusted* duration of exposure (days) to ribociclib/placebo represents the number of ribociclib/placebo dosing days a patient would be expected to have received per protocol, given their duration of exposure to ribociclib/placebo as defined in Section 3.6.1. Since ribociclib/Placebo follows a 3 weeks on, 1 week off schedule, the adjusted duration of exposure to ribociclib/placebo is the duration of exposure to ribociclib/placebo minus the planned off days. The adjusted duration of exposure to ribociclib/placebo is therefore 21 x (# completed 28 day cycles) + min(21, duration of last incomplete cycle).

For example, if the duration of exposure to ribociclib/placebo is 66 (corresponding to two cycles and 10 days), then the adjusted duration of exposure to ribociclib/placebo is 21*2+10=52. If the duration of exposure to ribociclib/placebo is 108 days (corresponding to three cycles and 24 days), then the adjusted duration of exposure is 21*3+21=84 days.

Specifically, let D1 represent the duration of exposure to ribociclib/placebo as defined in Section 3.6.1. Then the adjusted duration of exposure is defined as

D=21*[D1/28]+min(21,D1-28*[D1/28]) days,

where [x] stands for the integer part of x. In this equation [D1/28] is the number of completed cycles, and D1-28*[D1/28] is the additional number of days in the last, incomplete cycle (if any). For example, if D1=30 then [D1/28]=1, D1-28*[D1/28]=2, and D=23. If D1=7 then D=7; if D1=22 then D=21; if D1=28 then D=21, etc.

Planned dose intensity (PDI) is defined as the assigned dose by unit of time planned to be given to patients as per protocol. The PDI for each study treatment component is displayed in Table 3-1. Note that DI will also be calculated and reported in the units displayed in Table 3-1, whereas duration of exposure itself will be summarized in months.

Table 3-1 Planned dose intensity

Medication	PDI (dose unit/unit of time)
Ribociclib/placebo	600 mg/day
Tamoxifen	20 mg/day
Letrozole	2.5 mg/day
Anastrozole	1 mg/day
Goserelin	3.6 mg/28 days

Relative dose intensity (RDI) is defined as:

RDI = DI (dosing unit / unit of time) / PDI (dosing unit / unit of time).

DI and RDI will be summarized separately for each of the study treatment components.

3.6.4 Dose reductions, interruptions and delays

The number and percentage of patients with dose reductions, interruptions or delays, and the reasons, will be summarized by treatment arm as outlined below.

Interruption: An interruption is defined as a 0mg dose given on one or more days during the period where (for ribociclib/placebo only) a patient is not on the "off" part of a treatment cycle, after which > 0mg dose resumes. For patients who had dose interruption checked but never resumed non-zero dose, the dose interruption will not be counted. For example, in the sequence of 600 mg – 0mg (dose break) -0mg (dose interruption) – 0 mg (dose permanently discontinued) the 0mg (dose interruption) will not be counted as a dose interruption. Interruptions will be summarized for each component of study treatment.

Delay: A special case of an interruption is one that, for ribociclib/placebo, occurs at the start of a new cycle, after a planned rest period. Such instances will be identified as a subset of the interruptions (specifically those occurring directly after a planned dose break) and will be summarized separately as dose delays.

Reduction: A dose reduction for ribociclib/placebo is defined as a decrease from the previous non-zero dose to another non-zero dose less than the protocol-planned dose, even if this decrease has been directly preceded by an interruption. For example, in the sequence 600mg – 0mg – 400mg, the 400mg dose will be counted as a reduction.

If due to dosing error, a patient took a dose during the dosing break with a dose that is lower than the previous dose, this will not be considered as a dose reduction. For example, a patient took 600 mg from day 1-21, then mistakenly took 200 mg per day on day 22-28, which is supposed to be a dosing break, then resumed 600 mg dosing on day 29. This will not be considered as a dose reduction.

Missing data: If dose is recorded but frequency is missing or entered as 'none', it is assumed that the study drug was taken as per-protocol.

In this study, dose reductions for letrozole, anastrozole, tamoxifen or goserelin are not permitted and will not be summarized, however all dosing data will be listed.

3.6.5 Discontinuation of study treatment components

The reasons for discontinuation of each study treatment component will be summarized by treatment arm, based on the information on the component-specific DAR CRFs (i.e. using the "Reason Dose Permanently Discontinued" field on the applicable DAR CRF for patients who have the "dose permanently discontinued" box checked).

Partial discontinuation: Partial discontinuations and their reasons will also be summarized for ribociclib/placebo and goserelin (separately). Partial discontinuation refers to discontinuation of ribociclib/placebo or goserelin, respectively, while continuing on study treatment.

- For ribociclib/placebo, partial discontinuation is defined as the event when the last non-zero dose of ribociclib/placebo is more than 21 days before the last non-zero dose of study treatment, with the permanent discontinuation box checked in the ribociclib/placebo DAR page;
- For goserelin, partial discontinuation is defined as the last non-zero dose of goserelin plus 27 days being more than 21 days before the last non-zero dose of study treatment, with the permanent discontinuation box checked on the goserelin DAR page.

3.7 Concomitant and post treatment therapy

Concomitant therapy

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) besides the study treatment that were administered to a patient, coinciding with the study assessment period (even if started before the study assessment period).

Concomitant medications will be coded using the World Health Organization (WHO) Drug Reference Listing (DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (ATC) classification system.

Concomitant medications will be summarized by lowest ATC class, preferred term and treatment arm. These summaries will include:

1. medications starting on or after the start of study treatment but no later than 30 days after last dose of study treatment, and
2. medications starting prior to the start of study treatment and continuing after the start of study treatment.

All concomitant therapies will be listed. Any concomitant therapies starting and ending prior to the start of study treatment or starting more than 30 days after the last date of study treatment will be flagged in the listing.

The safety set will be used for all concomitant medication tables and listings.

Concomitant medications that have the potential to impact some specific analyses (e.g. PK, efficacy or safety analyses) will be identified prior to database lock. Separate summaries of these concomitant medications will be produced using the appropriate analysis set (e.g. FAS for those potentially affecting efficacy).

- Strong inhibitors or inducers of CYP3A4/5, substrates of CYP3A4/5 with a narrow therapeutic window, medications with a known risk of QT prolongation and other prohibited medications described in Section 2.4 will be identified. They will be tabulated by ATC class and preferred term.
- Any anti-neoplastic therapy administered concomitantly with study treatment will be listed based on their identification through the protocol deviation process.

Post treatment anti-cancer therapy

Anti-neoplastic therapies after discontinuation of study drug will be listed and tabulated by ATC class, preferred term and treatment arm by means of frequency counts and percentages using the FAS. In addition, 1st and 2nd new anti-neoplastic medication regimens after discontinuation of study treatment will be summarized by medication type, e.g. hormonal therapy, chemotherapy, etc. The medication type will be determined by clinical review.

[redacted]

3.8 Efficacy evaluation

The efficacy endpoints based on the tumor assessments will be derived according to the RECIST guideline version 1.1 (see Section 2.5 and Appendix 3 of the Clinical Study Protocol for details). The tumor endpoint derivation is based on the sequence of overall lesion responses at each assessment/time point. However, the overall lesion response at a given assessment/time point may be provided from different sources as illustrated in Table 3-2.

Table 3-2 Sources for overall lesion response

Source 1	Investigator (local radiology) reported overall lesion response
Source 2	Novartis-calculated overall lesion response based on raw (i.e. individual lesion) measurements from investigator (local radiology)
Source 3	Final BIRC-reported overall lesion response

The primary efficacy analysis will be based on the investigator/local radiology review. The investigator reported overall lesion response at each assessment/time point (Source 1 in Table 3-2) will be used to derive the efficacy endpoints.

The overall response at each assessment will also be calculated using the raw lesion measurements (Source 2 in Table 3-2). The calculated responses will be listed along with the responses given by the investigator. As a sensitivity analysis, PFS based on calculated overall lesion response (Source 2 in Table 3-2) will also be summarized.

Tumor assessment data based on BIRC assessment (Source 3) will be used for selected supportive efficacy analyses. The BIRC comprises of two independent radiologists and an adjudicator. The BIRC-reported overall lesion response data will be used to derive the BIRC-based endpoints. Data from independent readers will be listed together with the adjudication. . Differences in overall responses between local radiology (Source 1) and central radiology (Source 3) will be listed.

3.8.1 Primary efficacy

PFS based on local radiology assessment is the primary efficacy variable in this study. PFS is defined as the time from the date of randomization to the date of the first documented disease progression or death due to any cause. If a patient has not progressed or died at the analysis cutoff date, PFS will be censored at the time of the last adequate tumor assessment before the cut-off date. Definitions and further details on PFS can be found in Appendix 3 of the study protocol.

Discontinuation due to disease progression (collected on the 'End of treatment' and 'End of post treatment follow up' disposition page) without supporting objective evidence (as defined in Section 2.5.2) satisfying progression criteria per RECIST will not be considered disease progression for PFS derivation.

3.8.1.1 Primary analysis

The primary analysis of PFS will be based on the local radiological assessments (Source 1 in Table 3-2). The analysis will be performed on the FAS and will use the default censoring and event date options from Table 3-2 of study protocol Appendix 3 i.e. event/censoring rules will be based on options A(1), B(1), C1(1), C2(1), D(1), E(1), and F(1) (summarized in Table 3-3). In particular, PFS will be censored at the last adequate tumor assessment if a patient didn't have an event or the event occurred after two or more missing tumor assessments (see Section 2.5.5). In the primary analysis in this study PFS will not be censored if a new antineoplastic therapy is started; instead, an ITT approach will be used and this new antineoplastic therapy will be ignored for the purposes of PFS derivation (and tumor assessments will continue), i.e. option F(1) in Table 3-2 of protocol Appendix 3 will be used. A sensitivity analysis will be performed censoring PFS at the last adequate tumor assessment prior to start of new antineoplastic therapy, i.e. using option F(2). Discontinuation of study treatment (for any reason) will not be considered as a reason for censoring.

Table 3-3 Outcome and event/censor dates for PFS analysis

Situation	Date	Outcome

No baseline assessment	Date of randomization*	Censored
Progression or death at or before next scheduled assessment	Date of progression (or death)	Progressed
Progression or death after exactly one missing assessment	Date of progression (or death)	Progressed
Progression or death after two or more missing assessments	Date of last adequate assessment prior to missed assessments	Censored
No progression (or death)	Date of last adequate assessment	Censored
New anticancer therapy given prior to protocol defined progression	Ignore the new anticancer therapy and follow situations above	As per above situations
Treatment discontinuation due to 'Disease progression' without documented progression, i.e. clinical progression based on investigator claim	Ignore clinical progression and follow situations above	As per above situations
* The rare exception to this is if the patient dies no more than D2 days (see Section 2.5.5 for definition) after randomization, in which case this is a PFS event at the date of death		

3.8.1.2 Hypothesis and test statistic

The primary efficacy analysis will be the comparison of PFS between the two treatment arms using a stratified (strata based on IRT data) log-rank test at one-sided 2.5% level of significance.

Assuming proportional hazards model for PFS, the following statistical hypotheses will be tested to address the primary efficacy objective:

$$H_{01}: \theta_1 \geq 0 \text{ vs. } H_{a1}: \theta_1 < 0$$

where θ_1 is the log-hazard ratio (Tamoxifen or a NSAI+goserelin+ribociclib arm vs. tamoxifen or a NSAI+goserelin + placebo arm) of PFS.

3.8.1.3 Kaplan-Meier estimates

The survival distribution of PFS will be estimated using the Kaplan-Meier method. The results will be plotted graphically (Kaplan-Meier curves) by treatment arm. The plots will display the number of patients at risk at equidistant time points. The median, 25th and 75th percentiles for PFS for each treatment arm will be provided with associated 95% confidence intervals. The survival probabilities at 2-month intervals, and the associated 95% confidence intervals, will be summarized by treatment arm. Kaplan-Meier estimates will be obtained using PROC LEFETEST with method=KM option in SAS. The loglog option available within PROC LIFETEST will be used to compute the confidence intervals.

3.8.1.4 Hazard Ratio

The PFS hazard ratio with two-sided 95% confidence interval will be estimated using the stratified Cox proportional hazards model. In this analysis the baseline hazard function will be allowed to vary across strata. SAS PHREG procedure with ties=EXACT option will be used to carry out this analysis in which the model statement will include treatment arm variable as the

only covariate and the STRATA statement will include the stratum information as obtained via IRT.

3.8.1.5 Sensitivity and supportive analyses of the primary endpoint PFS

NSAI/tamoxifen subgroup analyses

In addition to the primary analysis in all patients, the treatment effect in terms of PFS for

1. tamoxifen + goserelin + ribociclib vs. tamoxifen + goserelin + placebo, and
2. NSAI + goserelin + ribociclib vs. NSAI + goserelin + placebo,

where tamoxifen/NSAI are identified based on CRF data, will be estimated using separate unstratified Cox regression models. Estimated hazard ratios with corresponding 95% CIs will be presented for each subgroup. Survival distribution of PFS will be estimated using Kaplan-Meier methodology within each subgroup.

Audit-based BIRC assessment of PFS

PFS assessed by Blinded Independent Review Committee (BIRC) will be used as a supportive analysis of the primary endpoint.

For studies with PFS based on local radiology assessment as the primary endpoint, PFS assessed centrally has generally been used as a secondary or supportive analysis of the treatment effect observed in the primary efficacy analysis. Although 100% central review of scans has been performed in many trials, there is a growing body of evidence that an audit based approach for central evaluation is sufficient (Zhang et al, 2012, FDA ODAC 2012).

An audit (sample) based approach will therefore be implemented for the BIRC assessment of PFS, whereby all assessments for a randomly selected subset of randomized patients will be assessed by BIRC. An independent random sampling process, implemented by the third party IRT vendor, will select approximately 40% of randomized patients. This random allocation will be stratified by randomized treatment arm and the strata used for the randomization of patients to treatment arms.

The distribution of PFS based on the audit BIRC sample will be estimated using the Kaplan-Meier method. The median along with two-sided 95% confidence interval (CI) will be presented by treatment group. A Kaplan-Meier figure will also be displayed. A stratified Cox regression will be used to estimate the hazard ratio (HR), along with two-sided 95% CI based on the audit BIRC sample.

In order to determine whether a 100% BIRC review should be conducted, two additional methods will be used to summarize the data from the sample-based BIRC assessment.

1. NCI method: The NCI (National Cancer Institute) method (Dodd et al. 2011) uses an auxiliary variable estimator of the log-hazard ratio that combines information from patient-level investigator assessments from all patients (in the FAS) and the BIRC assessment of patients randomly selected for central review (see Section 4.2.5 for methodology details). This estimate and its one-sided 95% CI will be provided. The NCI method is used for the audit sample size determination (see Section 3.16) and summary of treatment effect for the supportive BIRC assessment.

2. PhRMA method: The data from the BIRC assessment generated following the sampling scheme as above will also be summarized using the method proposed by Amit et al. 2011, referred to as the PhRMA (Pharmaceutical Research and Manufacturers of America) method, based on the early discrepancy rate (EDR) and late discrepancy rate (LDR).

   The EDR quantifies the frequency with which the investigator declares progression early relative to BIRC within each arm as a proportion of the total number of investigator assessed PDs. The LDR quantifies the frequency that the investigator declares progression later than BIRC as a proportion of the total number of discrepancies within the arm (see Section 4.2.5 for further details). If the distribution of discrepancies is similar between the arms this suggests the absence of evaluation bias favoring a particular arm. With this approach, the differential discordance (DD) of the early discrepancy rate (EDR) and late discrepancy rate (LDR) between the two arms will be estimated as the rate on the ribociclib arm minus the rate on the placebo arm. The EDR and LDR results will also be summarized by treatment arm.

The following thresholds based on the NCI and PhRMA methods will be used to define the trigger for a full BIRC review:

- If the upper-bound of the one-sided 95% confidence interval for BIRC-based log-hazard ratio exceeds zero (i.e. HR>1) based on the NCI method

and/or

- If ≥ 15% differential discordance is observed in EDR or LDR based on the PhRMA method.

A full BIRC review will only be considered if the primary endpoint is statistically significant. In the event of a full BIRC read of all patients' imaging data, all analyses based on BIRC assessment will be repeated including all randomized patients, with the exception of the auxiliary variable estimate (NCI method) described above.

Sensitivity analyses and additional supportive analyses

As a sensitivity analysis to assess the impact of stratification, the two treatment groups will be compared using an unstratified log-rank test. The hazard ratio together with associated 95% confidence interval obtained using the unstratified Cox regression model will also be presented.

The primary efficacy analyses; i.e. the stratified log-rank test, Kaplan-Meier estimates, estimate of the median PFS along with 95% confidence interval, and hazard ratio obtained using the Cox proportional hazards model, will be repeated (as appropriate) based on the data obtained:

1. Using the investigator/local assessments (Source 1 in Table 3-2) on the PPS and using the same conventions as for the primary analysis if the primary analysis is found to be statistically significant and if the PPS/FAS are different.
2. Using investigator/local assessment (Source 1 in Table 3-2) on the FAS and taking the event whenever it occurs - even after two or more missing tumor assessments. The following options from Table 3-2 in Section 3.2.10 of Study Protocol Appendix 3 will be used: A(1), B(1), C1(1), C2(3), D(1), E(1) and F(1). In the summary table, this approach is referred as 'Actual event PFS analysis'

3. Using investigator/local assessment (Source 1 in Table 3-2) on the FAS and backdating events occurring after missing tumor assessments. The following options from Table 3-2 in Section 3.2.10 of Study Protocol Appendix 3 will be used: A(1), B(1), C1(2), C2(2), D(1), E(1), and F(1). In the summary tables, this approach is referred as 'Backdating PFS analysis'.
4. Using investigator/local assessment (Source 1 in Table 3-2) on the FAS and censoring PFS at the last adequate tumor assessment before the start of any new antineoplastic therapy (excluding palliative radiotherapy and continuation of combination partner therapy alone after end of study treatment).

A stratified multivariate Cox regression model will be fitted to evaluate the effect of other baseline demographic or disease characteristic on the estimated hazard ratio. This model will include treatment arm and the following key prognostic factors: ECOG performance status (0 vs. ≥1), bone only lesion at baseline (yes or no), age (<40 vs ≥40 years old), and prior (neo-)adjuvant hormonal therapy usage (no; yes – progression while on or within 12 months of end of (neo-)adjuvant ET; yes – progression > 12 months after end of (neo-)adjuvant ET). In additional, stratified multivariate Cox models will be fitted including treatment arm and each of these factors individually, including treatment by factor interaction, to explore the relationship between each factor and treatment.

Further supportive analyses will include:

- Stratified Cox regression using derived local radiology assessment (source 2 in Table 3-2). No p-value will be presented.
- Number of patients and number of events by treatment arm within each stratum and for each level of each stratification factor will be presented along with the hazard ratio for treatment effect obtained using unstratified Cox proportional hazards regression models with corresponding confidence intervals. No p-values will be presented for thee analyses. K-M plots of survival distributions will also be presented.
- If there is a high rate of discrepancy between the strata classifications constructed using the eCRF data and those obtained from the IRT, a sensitivity analysis will be performed in which a stratified Cox regression model will be used to estimate the treatment hazard ratio and the associated 95% confidence interval based on the CRF-derived strata. No other inferential statistics will be provided.
- Number of patients with a PFS event and number of patients censored for the PFS analysis will be summarized. In addition, a summary of reasons for PFS censoring will be provided by treatment arm based on the reasons defined in Section 2.5.5. These summaries on censoring reasons will be produced for PFS by investigator radiology and BIRC review.
- Comparison of PFS event type/censor between local radiology review and BIRC review
- [redacted]
- If more than 5% of the patients are found to have been force randomized, a sensitivity analysis using local radiology assessments (Source 1 in Table 3-2) will be performed on

all randomized patients with the force randomized patients assigned to the alternative treatment group.

3.8.2 Secondary efficacy analyses

The secondary efficacy objectives in this study are to

- compare the two treatment arms with respect to overall survival
- evaluate the two treatment arms with respect to overall response rate and clinical benefit rate;
- describe time to response and duration of response in each treatment arm;
- evaluate the two treatment arms with respect to time to deterioration of ECOG Performance Status;
- evaluate patient reported outcomes for health-related quality of life in the two treatment arms.

OS is the key secondary endpoint. A hierarchical testing strategy, where OS is to be statistically evaluated and interpreted only if the primary efficacy endpoint PFS is statistically significant favoring the treatment arm, will be used to control the overall type-I error rate.

The analysis of all secondary efficacy endpoints will be performed based on the FAS.

3.8.2.1 Key secondary objective: overall survival

Comparing the overall survival (OS) between ribociclib+tamoxifen/NSAI+goserelin and placebo+tamoxifen/NSAI+goserelin is the key secondary objective in this study. OS is defined as the time from date of randomization to date of death due to any cause. A cut-off date will be established for each analysis of OS. If a patient is not known to have died by the date of analysis cut-off, then OS will be censored at the last contact date (see Section 2.1.9).

The analysis of OS will be based on the FAS. If the primary analysis of PFS is significant, the distribution of OS will be compared between the two treatment arms using a stratified log-rank test (strata based on IRT data) at one-sided 2.5% level of significance.

Up to 3 analyses may be performed for OS: (1) at the time of the primary analysis for PFS, at which point a total of 123 (49%) deaths are expected; (2) after approximately 189 (75%) deaths have been documented; and (3) a final analysis for OS after approximately 252 deaths. The type I error probability for OS tests will be controlled by using a Lan-DeMets (O'Brien- Fleming) alpha spending function at the one-sided α=0.025 level (see Section 3.14.2 for further details).

The distribution function of OS will be estimated using the Kaplan-Meier method and displayed for each treatment arm and stratum. The median OS along with 95% confidence intervals will be presented by treatment arm, along with the proportion of patients alive at selected time points and the associated 95% confidence intervals. The stratified Cox regression model, where the baseline hazard function is allowed to vary across strata, will be used to estimate the hazard ratio (HR) of OS, along with 95% confidence interval. Strata information will be obtained from the IRT.

[redacted]

[redacted]

Survival status, reason for censoring and death cause will be listed. Patients not known to have died will be censored for 'Loss to follow-up' if the time between their last contact date and the analysis cut-off date is longer than 12 + 2 weeks = 98 days (i.e. the planned interval between two OS follow-up visits plus the 1 week window on both sides).

Handling missing month/day in date of death

For rare cases when either day is missing or both month and day are missing for the date of death, the follow imputation rules will be implemented:

- If only day is missing, then impute max [(1 mmm-yyyy), min (last contact date +1, cutoff date)].
- If both day and month are missing, then impute max [(1 Jan-yyyy, min (last contact date +1, cutoff date)].

3.8.2.2 Overall response rate

ORR is defined as the proportion of patients with best overall response of confirmed complete response (CR) or confirmed partial response (PR) according to RECIST 1.1 (see Appendix 3 of the study protocol). ORR will be calculated based on the FAS using investigators' review of tumor assessment data. Patients with only non-measurable disease at baseline will be part of the analysis and will be included in the numerator only if a complete response was observed. ORR will be presented by treatment arm along with standard Wald asymptotic (i.e. normal approximation) 95% confidence intervals. The Cochran-Mantel Haenszel chi-square test (strata based on IRT data) will be used to compare the two treatment arms with respect to the ORR at a one-sided 2.5% level of significance.

As sensitivity analysis, ORR will be calculated and summarized including only patients with measurable disease at baseline.

As a supportive analysis, ORR will also be summarized based on the BIRC review of tumor data. However, no p-value will be presented.

Clinical benefit rate

CBR is defined as the proportion of patients with a best overall response of confirmed CR or PR, or SD lasting 24 weeks or longer, according to RECIST 1.1 criteria. A patient will be considered to have SD for 24 weeks or longer if a SD response is recorded at 24-1=23 weeks or later from randomization, allowing for the ±1 week visit window for tumor assessments. Patients with only non-measurable disease at baseline will be part of the analysis and will be

included in the numerator only if they achieve a complete response or have a 'Non-CR/Non-PD' response 23 weeks or more after randomization. CBR will be calculated using the FAS based on the investigators' tumor assessments. CBR will be summarized for the two treatment arms using descriptive statistics. The Cochran-Mantel-Haenszel test (strata based on IRT data) at one sided 2.5% level of significance will be used to compare the two treatment arms with respect to CBR.

As a sensitivity analysis, the above analyses will be repeated including only patients with measurable disease at baseline.

As a supportive analysis, CBR based on BIRC data will be summarized for the two treatment arms using descriptive statistics but no p-value.

3.8.2.3 Time to response

Time to response (CR or PR) is the time from date of randomization to first documented response (CR or PR, which must be confirmed subsequently) according to RECIST 1.1. All patients in the FAS will be included in the time to response calculation. Patients who did not achieve a confirmed PR or CR will be censored at:

- the maximum follow-up time (i.e. FPFV to LPLV used for the analysis) for patients who had a PFS event (i.e. either progressed or died due to any cause);
- the last adequate tumor assessment date for all other patients.

Time to response data will be listed and summarized by treatment arm. The distribution of time to response will be estimated using the Kaplan-Meier method and the median time to response will be presented along with 95% confidence interval only if a sufficient number of responses is observed. A descriptive summary of time to response for the responders will also be presented. No inferential analysis that compares time to response between the two treatment arms will be performed.

3.8.2.4 Duration of response

DoR applies only to patients whose best overall response was CR or PR. The start date is the date of first documented response (CR or PR, which must be confirmed subsequently) and the end date is the date of event defined as the first documented progression or death due to underlying cancer. The start date will be determined using the time the response was first determined and not using the time the response was confirmed. If a patient has not had an event, duration is censored at the date of last adequate tumor assessment, using the same censoring rules described for primary PFS analysis. DoR will be summarized by treatment arm. The distribution of duration of response will be estimated using the Kaplan-Meier method and the median response duration will be presented along with 95% confidence interval only if a sufficient number of responses is observed. No inferential analysis that compares the DoR between the two treatment arms will be performed.

3.8.2.5 ECOG performance status

The ECOG PS scale (Table 3-4) will be used to assess physical health of patients, ranging from 0 (most active) to 5 (least active):

Table 3-4 ECOG Performance Scale

Score	Description
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair
5	Dead

The following intervals will be used to group the ECOG PS data over time. Day in columns 2 and 3 is defined as date of ECOG PS assessment date – randomization date + 1. The correspondence with Day in column 1 assumes that a patient is treated on the day of randomization; however the definition of Day in columns 2 and 3 still applies if this is not the case, i.e. randomization date is taken as the reference for the windows.

Table 3-5 Time windows for ECOG PS assessments

Assessment	Target day of assessment	Time Interval
Baseline		Day 1 (if not available use screening)
Cycle 1 Day 15	15	Day 2 to day 21
Cycle 2 Day 1	29	Day 22 to day 42
Cycle 3 Day 1	57	Day 43 to day 63
Cycle 3 Day 15	71	Day 64 to day 77
Cycle 4 Day 1	85	Day 78 to day 98
Cycle k Day 1 (k≥5)	d=(k-1)*28+1	Day d-14 to day d+13
End of Treatment		Assessment taken at the end of treatment visit

If more than one assessment is done within the same time window, the assessment performed closest to the target date will be used. If 2 assessments within a time window are equidistant from the target date, then the worst of the 2 assessments will be used.

Frequency counts and percentages of patients in each score category will be provided by treatment arm and time point.

An analysis of time to definitive deterioration of performance status will be performed. The time to definitive deterioration is calculated from the date of randomization. Deterioration is defined as an increase in ECOG PS by at least one category from the baseline score or death due to any cause. Deterioration is considered definitive if no return to baseline or better in ECOG PS is observed subsequent to the deterioration during the treatment period.

If a definitive deterioration is observed after two or more missing assessments (using an analogous rule to that defined for PFS in Section 2.5.5, based on the windows specified above),

time to deterioration will be censored at the date of the last ECOG PS assessment prior to the deterioration.

Patients receiving any further anti-neoplastic therapy before definitive deterioration will be censored at the date of their last assessment before the start date of the therapy. Patients that have not deteriorated as of the cutoff date will be censored at the date of their last assessment before the cutoff.

Kaplan-Meier estimates will be constructed for each treatment arm. The median, 25th and 75th percentiles for time to definitive deterioration for each treatment group will be obtained along with 95% confidence intervals.

A stratified log-rank test at one-sided 2.5% level of significance will be used to test the difference in time to definitive deterioration of performance status between treatment arms. A (stratified) Cox proportional hazards model will be used to estimate the hazard ratio (with 95% confidence interval).

3.8.2.6 Patient reported outcomes

The European Organization for Research and Treatment of Cancer's core quality of life questionnaire (EORTC-QLQ-C30, version 3.0) and breast cancer specific questionnaire (EORTC QLQ-BR23, version 1.0) and the EuroQoL 5-level instrument (EQ-5D-5L, Version 4.0) will be used to evaluate patient-reported outcome measures of health-related quality-of-life, functioning, disease symptoms and treatment-related side effects.

[redacted]

The PRO instruments are planned to be administered during screening and every 8 weeks after randomization in the first 18 months, and every 12 weeks thereafter until the end of treatment. PRO assessments [redacted] will continue to be collected during the efficacy follow-up after the end of treatment.

The following time based intervals will be used to group the PRO data over time. Day is defined as date of PRO assessment date – randomization date + 1.

Table 3-6 Time windows for patient reported outcomes

	Time Interval
Baseline	Last assessment before or on date of randomization
Day 1 of Cycle 3, 5, 7, 9 until cycle 19	+/- 4 weeks centered around the planned assessment date (except for the first and last windows): i.e. days 2 to 85 for Cycle 3 Day 1 days 86 to 141 for Cycle 5 Day 1 days t-27 to t+28 for Cycle k Day 1 (k= 7, 9, ..., 17), where t=(k-1)*28+1 is the target day of the assessment day 478 to 547 for Cycle 19 Day 1
Day 1 of Cycle 22, 25, 28, …	+/- 6 weeks centered around the planned assessment date: i.e. days 548 to 631 for Cycle 22 Day 1

	days 632 to 715 for Cycle 25 Day 1 days t-41 to t+42 for Cycle k Day 1 (k= 28, 31,...), where t=(k-1)*28+1 is the target day of the assessment
End of treatment	At the end of treatment visit
Efficacy follow-up [redacted]	At each of the efficacy follow-up visits

If more than one assessment is done within the same time window, the assessment performed closest to the target date will be used. If 2 assessments within a time window are equidistant from the target date, then the assessment obtained prior to target date will be used.

The FAS will be used for all PRO summaries and listings.

EORTC-QLQ and EQ-5D-5L

The global health status/global QoL scale score of the EORTC QLQ-C30 is identified as a primary PRO variable of interest. The physical functioning, emotional functioning and social functioning sub-scale scores of the EORTC QLQ-C30, the breast cancer symptoms scale of the EORTC QLQ-BR23, and the visual analog scale (VAS) of the EQ-5D-5L are identified as secondary PRO variables of interest. High scores in the EORTC QLQ-C30 and QLQ-BR23 represent a higher response level. Thus a high score for a functional scale represents a high / healthy level of functioning; a high score for the global health status / QoL represents a high QoL, but a high score for a symptom scale / item represents a high level of symptomatology /problems. Higher scores in the EQ-5D-5L also correspond to better health states.

The number of patients completing QoL questionnaires and the number of patients missing/expected to have QoL assessments will be summarized by treatment arm for scheduled assessment time points (the number of ongoing patients will be used as denominator). Furthermore, the amount and the pattern of missing data may be explored by treatment arm and over time using summary statistics. The following categories will be used to describe whether the questionnaire was completed at a specific time point:

- yes, fully completed
- yes, partly completed
- no.

Scoring of raw data and methods for handling of missing items or missing assessments will be handled according to scoring manuals for each respective patient questionnaire (Fayers et al 2001, Rabin 2011).

A time-to-event approach will be considered in the analysis of the time to definitive 10% deterioration in the primary and the secondary PRO variables of interest.

Definitive 10% deterioration is defined as a worsening in score by at least 10% compared to baseline, with no later improvement above this threshold observed during the treatment period, or death due to any cause. Time to deterioration is the number of days between the date of randomization and the date of the assessment at which deterioration is seen. If a patient has not had an event prior to analysis cut-off, end of treatment, start of new anti-neoplastic therapy, lost to follow-up, or withdrawal of consent, the time to deterioration will be censored at the date of

the last evaluation before the earliest of these times. Only assessments collected while the patient is on treatment and on or before the end of treatment visit will be included in the PRO time to deterioration analysis.

If deterioration is observed after two or more missing assessments, time to deterioration will be censored at the date of the last assessment prior to the deterioration. The rules for calculating number of missing assessments are the same as those for tumor assessments (see Section 2.5.5)

Time to 10% deterioration in the global health status/QOL scale, and secondary PRO scales of interest as listed above, will be compared between the two treatment arms using a stratified log-rank test (strata based on IRT data) at one-sided 2.5% level of significance. The survival distributions will be presented descriptively using Kaplan-Meier curves. Summary statistics from the Kaplan-Meier distributions will be determined, including the median time to 10% deterioration and the proportions of patients without 10% deterioration at 2-monthly intervals. Both point estimates and 95% CIs will be presented. A stratified Cox regression model will be used to estimate the hazard ratio (HR) of time to deterioration, along with 95% confidence interval.

Change from baseline in all sub-scales obtained from EORTC QLQ-C30 and the breast symptoms score of QLQ-BR23, and the VAS of EQ-5D-5L will be analyzed using a linear mixed effects model for longitudinal data to assess the treatment effect over time including terms for treatment, stratification factors, time of visit (in weeks counting from the time of randomization to the time of a particular post baseline measurement), treatment by time of visit interaction, and baseline score. Time will be considered as a continuous variable in this analysis. The data for selected time points will also be fitted using a linear model including treatment stratification factors, and baseline value as covariates. The differences in least square means between treatment and control group, and the corresponding 2-sided 95% CI at selected time points will be presented.

For the mixed effects model, patients with baseline and at least one non-missing post-baseline assessments will be included. This analysis will only include assessments up to the time point where there are at least 50 patients on each of the treatments. Data collected under treatment (i.e. up to the EOT assessment) will be included. As a first approach, an unstructured correlation matrix will be used to model the correlation within patients. The structure of the correlation matrix will be investigated and simplified using likelihood ratio test if appropriate. If PRO is found not to follow a linear trend, the time variable might be considered as a categorical variable instead of continuous in the model.

Descriptive statistics (n, mean, median, SD, min, max) will be used to summarize the individual item and subscale scores from the EORTC QLQ-C30, QLQ-BR23, and EQ-5D-5L at each scheduled assessment time point. Additionally, change from baseline in the subscale scores at the time of each assessment will be summarized. Patients with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

[redacted]

[redacted]

3.9 Safety evaluation

The assessment of safety will be based mainly on the frequency of adverse events and on the number of laboratory/ECG values that fall outside of pre-determined ranges. Other safety data (e.g. vital signs and special tests) will be considered as appropriate.

All safety outputs will use the safety set. The safety summary tables will include only 'on-treatment' events/assessments, i.e. those collected on or after the first date of study treatment and collected no later than 30 days after the date of last study treatment administration. The AEs started before the first dose but worsening during the treatment period are also considered as 'on-treatment' events. All safety events/assessments will be listed and those collected outside of the on-treatment window will be flagged.

3.9.1 Adverse events (AEs)

3.9.1.1 Coding of AEs

Adverse events are coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology.

3.9.1.2 Grading of AEs

AEs will be assessed according to the Common Terminology Criteria for Adverse Events (CTCAE) version 4.0.3.

The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE v4.0.3 grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening, and death.

If CTCAE grading does not exist for an adverse event, grades 1 – 4 corresponding to the severity of mild, moderate, severe, and life-threatening will be used. CTCAE grade 5 (death) will not be used in this project; if an AE results in death it will be documented in the outcome ("fatal"). Information on deaths will also be collected on the 'Death' CRF.

3.9.1.3 General rules for AE Reporting

AE summaries will include all AEs starting on or after study Day 1 (i.e. on or after the day of the first intake of study treatment) and starting no later than 30 days after the last administration of study treatment (see Section 2.1.5). All AEs will be listed. AEs starting prior to study Day 1 and AEs starting later than 30 days after the last treatment date will be flagged in the listings.

AEs will be summarized by presenting the number and percentage of patients having at least one AE, having at least one AE in each primary system organ class, and for each preferred term using MedDRA coding. A subject with multiple occurrences of an AE will be counted only once in the AE category.

Separate AE summaries will be presented by primary system organ class, preferred term, and maximum grade . A patient with multiple grades for an AE will be summarized under the maximum grade recorded for the event. AE with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

In AE summaries the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for the preferred term will be based on their frequency in the ribociclib arm.

The frequency of grade 3 and 4 AEs will be summarized separately.

Any information collected (e.g. grades, relationship to study treatment, action taken etc.) will be summarized and listed as appropriate.

3.9.1.4 AE summaries

The following adverse event summaries will be produced:

- Summary of deaths and adverse events
- AEs, irrespective of causality, by primary system organ class, preferred term and maximum grade
- AEs with suspected relationship to study treatment by primary system organ class, preferred term and maximum grade
- Most common grade 3-4 adverse events, irrespective of causality, by preferred term and maximum grade (greater than x% in either arm)
- Adverse events, irrespective of causality, by primary system organ class and maximum grade
- Adverse events, irrespective of causality, by preferred term and maximum grade

- Adverse events with suspected relationship to study treatment by preferred term and maximum grade
- Grade 3 or 4 AEs, irrespective of causality, by primary system organ class, preferred term and maximum grade
- Grade 3 or 4 AEs with suspected study treatment relationship by primary system organ class, preferred term and maximum grade
- On treatment deaths, by preferred term
- All Deaths(on-treatment + post-treatment), by primary system organ class and preferred term
- Serious AEs, irrespective of causality, by primary system organ class, preferred term and maximum grade
- Serious AEs with suspected relationship to study treatment, by primary system organ class, preferred term and maximum grade
- AEs leading to discontinuation, irrespective of causality, by primary system organ class, preferred term and maximum grade
- AEs requiring dose adjustment , irrespective of causality, by primary system organ class, preferred term and maximum grade
- AEs requiring treatment interruption, irrespective of causality, by primary system organ class, preferred term and maximum grade
- AEs requiring additional therapy, irrespective of causality, by primary system organ class, preferred term and maximum grade
- Non-serious AEs (at least x% incidence rate in either treatment arm) by primary system organ class and preferred term.
- Serious and non-serious adverse events with number of occurrences (an occurrence is defined as >1 day between start and prior end date of record of same preferred term)
- On-treatment deaths and SAEs with fatal outcome, by SOC and PT

AEs of special interest will also be summarized. See Section 3.9.1.5 for the grouping details.

3.9.1.5 Grouping of AEs of special interest

An adverse event of special interest is a grouping of adverse events that are of scientific and medical concern specific to ribociclib. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HLGTs (high level group terms), HLTs (high level terms) and PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad. For each specified AESI, the number and percentage of patients with at least one event of the AESI occurring during the on-treatment period will be summarized.

Summaries of these AESIs will be provided by treatment arm, (specifying grade, SAE, relationship, leading to treatment discontinuation, leading to dose adjustment/interruption, etc.).

A Case Retrieval Sheet (CRS) with the exact composition of the AE groupings is to be used to map reported AEs to the AESI groupings. This file may be updated (i.e. it is a living document) based on review of accumulating trial data, and therefore the groupings are also subject to potential change. Table 3-7 provides the latest groupings at the time of the finalization of the SAP, from the CRS dated 13Dec2016. The most up-to-date version of the CRS will be used at the time of the analysis.

Table 3-7 AESI groupings

AESI grouping	MedDRA category
Anemia	SMQ
Diarrhea	SMQ
Hepatobiliary toxicity	SMQ
Infections	SMQ and SOC
Leukopenia	HLT
Nausea, emesis	HLT
Neutropenia	HLT and PT
Pneumonitis, interstitial lung disease	SMQ
Pulmonary embolism	SMQ and HLT
QTc prolongation	SMQ
Renal impairment	SMQ
Reproductive toxicity	SMQ
Thrombocytopenia	SMQ

3.9.2 Laboratory data

On analyzing laboratory data, data from all sources (central and local laboratories) will be combined. The summaries will include all laboratory assessments collected no later than 30 days after the last administration of study treatment. All laboratory assessments will be listed and those collected later than 30 days after the last treatment date will be flagged in the listings.

Laboratory data will be classified (by biostatistics/statistical reporting) into CTC grades according to the NCI CTCAE v4.0.3. The calculation of laboratory CTC grades will be based on the observed laboratory values only; clinical assessments will not be taken into account. The criteria to assign CTC grades are given in Appendix 1.

For laboratory tests where grades are not defined by CTCAE, results will be graded by the low/normal/high classifications based on laboratory normal ranges.

The following summaries will be produced for the laboratory data (by laboratory parameter):

- Number and percentage of patients with each CTC grade as their worst post-baseline value (regardless of the baseline status). Each patient will be counted only for the worst grade observed post baseline.
- Shift tables using CTC grades to compare baseline to the worst post-baseline value will be produced for hematology and biochemistry laboratory parameters with CTC grades.

- For laboratory parameters where CTC grades are not defined, shift tables to the worst post-baseline value will be produced using the low/normal/high classifications based on laboratory reference ranges.
- Number and percentage of patients meeting categorical liver function test criteria, including ALT, AST and ALT/AST (>3x, 5x, 8x, 10x, 20x ULN), Total Bilirubin (>1x, 2x ULN), ALP (>1.5x, 2x, 3x, 5x, 8x, 10x ULN), combined categories of ALT/AST and total bilirubin (e.g., ALT/AST>3x ULN & total bilirubin > ULN) as well as potential Hy's Law criteria (ALT or AST > 3 x ULN and TBIL > 2 x ULN and ALP < 2 x ULN). For the combined categories, the assessments need not be concurrent, i.e. patients are counted based on their most extreme value for each parameter (highest in the case of ALT, AST and TBIL; lowest in the case of ALP).
- Box plots of laboratory values by scheduled time point and treatment arm.

The following listings will be produced for the laboratory data:

- Listing of patients with CTC grade 3 or 4 laboratory abnormalities;
- Listing of all laboratory data with values flagged to show the corresponding CTC grades and the classifications relative to the laboratory reference ranges.

Time to first occurrence of grade 2 or worse laboratory toxicity and time to first occurrence of grade 3 or worse laboratory toxicity will be performed for (i) neutropenia and (ii) ALT/AST. Median time to first occurrence and 95% CI will be provided based on the Kaplan-Meier method. In addition, Kaplan-Meier plots will be generated.

Time to first occurrence of grade X or worse laboratory toxicity is defined as the time from the start of treatment to the start date of the first incidence of grade X or worse laboratory toxicity, i.e. time in days is calculated as (start date of first occurrence) – (date of first dose of study treatment) +1. A patient will be censored if:

- The patient did not report any post-baseline grade X or worse event on or before the analysis cut-off date.
- The patient discontinued treatment without reporting any grade X or worse event up to 30 days after study treatment discontinuation.
- The patient died without reporting any grade X or worse event.
- The patient received a new anticancer therapy before reporting any grade X or worse event.

The censoring date will be the earliest of the following dates: end of treatment + 30 days, analysis cut-off, new anti-cancer therapy start date, death date and last non-missing assessment for the lab parameter. Note that patients who have grade X or worse toxicity at baseline or missing baseline evaluation will be excluded from this analysis.

In addition, the median time to first occurrence for the subset of patients who experienced the event of interest will be calculated. Simple descriptive statistics, median, min and max as well as 25th percentile and 75th percentile, will be presented.

Duration of grade X or worse laboratory toxicity may also be summarized. Duration of grade X or worse event is calculated as:

(Date when the grade of the event decreases to below X) – (date of onset of grade X or worse event) + 1

For patients experiencing any grade X or worse event, the duration of the first such event will be summarized using the Kaplan-Meier method. Median duration and 95% CI will be presented. In addition, Kaplan-Meier plots will be generated.

A patient will be censored for the duration of grade X or worse event, if:

- The patient dies without reporting a decrease to below grade X
- The patient receives a new anticancer therapy before reporting a decrease to below grade X
- The patient discontinues from the study treatment without reporting a decrease to below grade X up to 30 days after study treatment discontinuation
- The patient is still ongoing at the analysis cut-off date without reporting a decrease to below grade X

The censoring date is the earliest of the following dates: end of treatment + 30 days, analysis cut-off, new anti-cancer therapy start date, death date and last non-missing assessment for the lab parameter.

Urinalysis

The following parameter will be summarized using shift table to compare baseline to the worst-post baseline values: urine bilirubin dipstick, urine blood dipstick, urine glucose dipstick, urine ketones dipstick, urine leukocyte dipstick, and urine nitrate dipstick. For all these parameters, both negative and trace are considered as normal and the more pluses the worse. The urine pH dipstick will be summarized using shift table with low/normal/high classifications based on laboratory reference ranges.

3.9.3 Vital signs

Vital signs assessments are performed in order to characterize basic body function. The parameters expected to be collected include: height, weight, body temperature, heart rate, and systolic and diastolic blood pressure.

The criteria for clinically notable abnormalities are defined as follows:

Clinically notable elevated values

- Systolic BP: ≥ 180 mmHg and an increase ≥ 20 mmHg from baseline
- Diastolic BP: ≥ 105 mmHg and an increase ≥ 15 mmHg from baseline.
- Body temperature: ≥ 39.1°C
- Heart rate: ≥ 120 bpm with increase from baseline of ≥15 bpm

Clinically notable below normal values

- Systolic BP: ≤ 90 mmHg and a decrease ≥ 20 mmHg from baseline
- Diastolic BP: ≤ 50 mmHg and a decrease ≥ 15 mmHg from baseline
- Body temperature: ≤ 35°C
- Heart rate: ≤ 50 bpm with decrease from baseline of ≥ 15 bpm

The following summaries will be produced for each vital sign parameter:

- Summary statistic for change from baseline to the worst post-baseline value (in both directions, i.e. from baseline to highest post baseline and from baseline to lowest post baseline value).
- Number and percentage of patients with at least one post-baseline vital sign abnormality (in both directions, i.e. both elevated and below normal values).

In addition, the following two listings will be produced by treatment arm:

- Patients with clinically notable vital sign abnormalities.
- All vital sign assessments will be listed by patient and vital sign parameter.

In both listings, the clinically notable values will be flagged and also the assessments collected later than 30 days after the last treatment date will be flagged.

3.9.4 ECG

All analyses of ECG data will be based on the average of all available replicate ECGs assessed at each time point for each patient. For unscheduled assessments, 15-minute windows will be applied to group assessments for averaging.

ECG data will be summarized by presenting summary statistics of the raw data and change from baseline by treatment arm and time point. The following parameters will be assessed: QT, QTcF, QTcB, PR, and QRS intervals in msec, heart rate (bpm), and the overall interpretation if clinically significant abnormalities are present.

The number and percentage of patients with notable abnormalities will be summarized. Summary statistics and shift tables will be presented. Individual listings will be provided by subject.

Table 3-8 Clinically notable ECG values

ECG parameter (unit)	Clinically notable criteria
QT, QTcF, QTcB (ms)	New > 450
	New > 480
	New > 500
	Increase from Baseline > 30
	Increase from Baseline > 60
PR duration (ms)	Increase > 25% from baseline and to PR duration > 200 New > 200
QRS duration (ms)	Increase > 25% from baseline and to QRS duration > 110 New > 110
Heart Rate (bpm)	< 50 and decrease from Baseline of > 25% > 100 and increase from Baseline of > 25%

A newly occurring ECG abnormality is defined as an abnormal post-baseline ECG finding that is not present at baseline. Baseline is defined as the average of the last replicate ECG measurements taken

on or before treatment start date. The percentage of patients having notable ECG interval values is based on the number of patients at risk for the change with a value at baseline and post-baseline.
Time to grade 2 or worse QT prolongation will be analyzed using the Kaplan-Meier method and the median time to grade 2 or worse QT prolongation will be presented along with a 95% confidence interval if there are sufficient numbers of events for each treatment group.

In addition, the median time to event for the subset of patients who experienced grade 2 or worse QT prolongation will be calculated. Simple descriptive statistics, median, min and max as well as 25th percentile and 75th percentile, will be presented.
Note that patients who have grade 2 or worse toxicity at baseline or missing baseline evaluation will be excluded from these analyses.

ECG by dose group will also be explored as described in Section 3.10.4.3.

3.9.5 Cardiac imaging (MUGA / ECHO)

Shift tables comparing baseline to worst post-baseline cardiac imaging (MUGA or ECHO) overall interpretation will be provided. Percentages will be based on all patients in the Safety set.

Note: If there is any change in the methodology used throughout the study compared to baseline, the post-baseline values for which the methodology differs from baseline will be discarded in the tables presenting comparisons to baseline.

Descriptive statistics of the left ventricular ejection fraction (LVEF) at baseline, worst post-baseline value and change from baseline to worst post-baseline value will be provided.

A listing of patients with newly occurring clinically significant abnormality will be produced by treatment arm.

3.9.6 Other safety data

Data from other tests will be listed, notable values will be flagged, and any other information collected will be listed as appropriate.

All assessments collected later than 30 days after the last treatment date will be flagged in the listings.

Any statistical tests performed to explore the data will be used only to identify any interesting comparisons that may warrant further consideration.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

3.14 Interim analyses

3.14.1 Primary Endpoint: PFS

No interim analyses are planned for PFS. The primary endpoint PFS will be assessed after approximately 329 PFS events have been documented.

3.14.2 Key Secondary Endpoint: OS

Overall survival will be compared between the two treatment groups, provided the primary endpoint PFS is statistically significant favoring the test treatment arm (tamoxifen or a NSAI + goserelin + ribociclib). A hierarchical testing procedure will be adopted in this study and the OS analyses will be performed only if the primary efficacy endpoint PFS is statistically significant. A maximum of three analyses are planned for OS: at the time of the primary analysis for PFS (provided PFS is significant), at which point a total of 123 deaths are expected, after 189 events have been documented, and a final analysis for OS when 252 deaths are expected (expected 40 months from date of first patient to be randomized).

An α-spending function according to Lan-DeMets (O'Brien-Fleming), (as implemented in East 5.4) along with the testing strategy outlined below will be used to maintain the overall type I error probability (Lan and DeMets 1983). This guarantees the protection of the 2.5% overall level of significance across the two hypotheses and the repeated testing of the OS hypotheses in the interim and the final analysis (Glimm 2010). The trial allows for the stopping of the study for a superior OS result, provided the primary endpoint PFS has already been shown to be statistically significant favoring the test treatment arm (tamoxifen or a NSAI + goserelin + ribociclib). Further, the exact nominal p-values that will need to be observed to declare statistical significance at the time of these analyses for OS will depend on the number of OS events that have been observed at the time of these analyses and the α for OS already spent at the time of earlier analyses.

The operating characteristics for OS are shown in Table 3-11 considering the hierarchical testing strategy of PFS and OS. The probabilities shown in Table 3-11 are conditional on PFS being statistically significant.

Table 3-11 Timelines for interim and final analyses

Months after randomization of the first patient	PFS events	Cumulative Power (%) against a hazard ratio of 0.67	OS events	Cumulative Conditional Power (%) against hazard ratio of 0.7
22	329 (100%)	95	123 (49%)	15
30	--	--	189 (75%)	54
40	--	--	252(100%)	80

Statistical significance of OS will only be declared if significance for primary PFS analysis has been declared.

3.14.3 Confidentiality of Interim OS results

At the time of primary PFS analysis, both PFS and interim OS analysis will be performed by the Sponsor's clinical team. Investigators and patients will remain blinded to study treatment and all patients will continue to be followed for OS until the final OS analysis (or earlier if OS reaches statistical significance at any of the interim analyses).

[redacted]

[redacted]

[redacted]

[redacted]

3.16 Sample size calculation

The assumption of median PFS of 9 months in the control arm (tamoxifen/NSAI + goserelin) for sample size calculation is based on the meta analysis of 4 trials in premenopausal advanced breast cancer treated with combination of tamoxifen and luteinizing hormone-releasing hormone agonist (Klijn 2001) and the results of the letrozole study in breast cancer patients (Mouridsen 2003). It is expected that the addition of ribociclib will result in 33% reduction in the hazard rate (corresponding to an increase in median PFS to 13.4 months assuming PFS has an exponential model assumption).

If the true hazard ratio is 0.67 (under alternative hypothesis), a total of 329 PFS events provide 95% power at a one-sided overall 2.5% level of significance to reject the null hypothesis (hazard ratio=1) using a log-rank test. Considering a recruitment period of approximately 18 months at a uniform rate of 33 patients/month, 600 patients will need to be randomized to the two treatment arms in a 1:1 ratio. Assuming about 10% patients will be lost to follow-up for PFS, a total of 660 patients will need to be randomized. Given the above assumptions, it is estimated that the 329th PFS event will be observed at approximately 22 months from the date of first patient randomized in the study. The sample size calculation was conducted with software package East 6.3.

It is expected that both the endocrine treatment types (tamoxifen and NSAI) will have an adequate number of patients randomized for the respective subgroup analyses. However, this will be monitored closely in a blinded fashion and the protocol may be amended for further randomization in a given group if enrollment is found to be slow.

Audit size calculation for BIRC assessment of PFS

The audit size of the sample-based BIRC assessment will be 40% of all randomized patients. Based on the audit size calculation approach proposed by Dodd et al (2011), assuming investigator and BIRC assessments are similar and the estimated log of investigator-based HR is -0.40 (i.e. HR=0.67), the audit size of 40% will ensure that the upper bound of a one-sided 95% CI for BIRC-based log-hazard ratio has 86% probability of being below zero (i.e. HR < 1) if the correlation between investigator assessment and BIRC assessment is 0.7 (the estimated

correlation based on data from the BELLE-2 [CBKM120F2302] study in metastatic breast cancer).

3.17 Power for analysis of key secondary variables

OS will be compared between the two treatment arms, provided that the primary endpoint, PFS, is statistically significant. Based on available data (Klijn 2001, Jonat 1995) the median OS in control arm is expected to be around 33 months. It is hypothesized that test treatment arm (LEE011 + tamoxifen or a NSAI + goserelin) will result in a 30% reduction in the hazard rate for OS (corresponding to an increase in median survival to 47 months under the exponential model assumption). If the true hazard ratio is 0.7 (under alternative hypothesis), a total of 252 deaths need to be observed to have 80% power at a one-sided overall 2.5% level of significance to reject the null hypotheses (hazard ratio = 1) using a log-rank test and a 3-look group sequential design. Based on the same number of patients that are planned to be enrolled in this study to detect the primary endpoint, it is estimated that these 252 deaths will be observed at approximately 40 months from the date of first patient to be randomized. Therefore the estimated time for the final OS evaluation will be 18 months after the primary analysis of the PFS endpoint has been conducted. The power calculation was conducted with software package East 6.3.

[redacted]

[redacted]

4 Details of the statistical analysis

4.1 Baseline comparability

Appropriate descriptive summary statistics of baseline variables (see Section 3.2) will be provided as in-text tables in the core CSR and also in Section 14 in the post-text tables. The summaries will be grouped by treatment arms, but no p-values will be provided.

4.2 Time-to-event analyses

The following sections present a general methodology to be used to analyze time-to-event variables, e.g.:

- Progression-free survival
- Overall survival
- Time to definitive deterioration of the ECOG score by one category of the score from baseline
- Time to response: defined as the time between date of randomization until first documented response (CR or PR) according to RECIST
- Time to definitive deterioration of PRO scores (eg, global health status/QoL, physical functioning, emotional functioning, social functioning scales of the EORTC QLQ-C30, total score of breast cancer symptoms scale of the EORTC QLQ-BR23, and the VAS of the EQ-5D-5L)

4.2.1 Analysis of time-to-event data with ties

The STRATA statement in LIFETEST procedure will be used to analyze time to event data with ties. The PHREG procedure in SAS with option TIES=EXACT will be used to fit the Cox proportional hazards model.

4.2.2 Hazard ratio

The hazard ratio as a measure of treatment effect will be derived from the Cox proportional hazards model using SAS procedure PHREG with TIES=EXACT option in the MODEL statement. The stratified unadjusted Cox model will be used (where the baseline hazard function is allowed to vary across strata) for the primary analysis, i.e. the MODEL statement will include only the treatment arm variable as a covariate and the STRATA statement will include stratification variable(s).

Hazard ratio with two-sided 95% confidence interval will be based on Wald test.

4.2.3 Hypothesis and test statistic

The primary efficacy analysis will be the comparison of the distribution of PFS between the two treatment arms using a stratified log-rank test at one-sided 2.5% level of significance.

Assuming proportional hazards model for PFS, the following statistical hypothesis will be tested to address the primary efficacy objective

H_{01}: $\theta_1 \geq 0$ vs. H_{a1}: $\theta_1 < 0$

where θ_1 is the log-hazard ratio (Tamoxifen or a NSAI+goserelin+ribociclib arm vs. tamoxifen or a NSAI+goserelin + placebo arm) of PFS..

The ***stratified log-rank*** test (strata based on IRT data) will be implemented as follows: For each of the K=8 strata, the LIFETEST procedure will be run with the STRATA statement including only the treatment variable. The TIME statement will include the survival time and a (right) censoring variable. The rank statistic S_k and the corresponding variance $var(S_k)$ (k=1, 2) will be estimated from this analysis.

The final test statistics will then be reconstructed using the formula:

$Z=[S_1+ \ldots + S_K]/\sqrt{[var(S_1) + \ldots + var(S_K)]}$. One-sided p-value will be computed using this Z statistic. Note: Under the null hypothesis, the asymptotic distribution of the test statistic Z is approximately normal (and correspondingly, Z^2 is approximately distributed as chi-square with one degree of freedom).

4.2.4 Kaplan-Meier estimates

The survival function in each treatment arm will be estimated using the Kaplan-Meier (product-limit) method as implemented in PROC LIFETEST with METHOD=KM option. The PROC LIFETEST statement will use the option CONFTYPE=LOGLOG.

Median survival for each treatment group will be obtained along with 95% confidence intervals calculated from PROC LIFETEST output using the method of [Brookmeyer and Crowley 1982]. Kaplan-Meier estimates of the survival function with 95% confidence intervals at specific time points will be summarized. Median survival for each treatment arm will be obtained along with 95% confidence intervals calculated from PROC LIFETEST output using the method of [Brookmeyer and Crowley 1982]. Kaplan-Meier estimates with 95% confidence intervals at specific time points will be summarized. The standard error of the Kaplan-Meier estimate will be calculated using Greenwood's formula [Collett 1994].

4.2.5 Audit-based BIRC assessment of PFS

NCI method

The auxiliary variable estimator of the NCI audit-based method (Dodd et al. 2011) has the form

$$\tilde{\theta}_C = \hat{\theta}_{CA} + \hat{\lambda}\left(\hat{\theta}_{L\bar{A}} - \hat{\theta}_{LA}\right),$$

where $\hat{\theta}_{CA}$, $\hat{\theta}_{L\bar{A}}$ and $\hat{\theta}_{LA}$ are estimators of the log-hazard ratio based on the central assessment in the audited subset of patients, the local assessment in the nonaudited subset of patients, and the local assessment in the audited subset, respectively. $\hat{\lambda}$ is defined as $\hat{\rho}\sqrt{\delta}(1-\delta)\sqrt{(\hat{V}_{CA}/\hat{V}_L)}$,

where $\hat{V}_{CA}$ and $\hat{V}_L$ are variance estimators of $\hat{\theta}_{CA}$ and $\hat{\theta}_L$ (the estimator of log-HR based on the local assessment in all patients) respectively, δ is the proportion of patients in the audited subset, and $\hat{\rho}$ is an estimator of the correlation between $\hat{\theta}_{LA}$ and $\hat{\theta}_{CA}$. For the latter, a bootstrap approach will be used:

- Within the audited subset of size *m*, *m* patients will be sampled with replacement. Using this sample of *m* patients, the log-hazard ratio will be estimated based on the local and central assessments separately;
- This procedure will be repeated 10 000 times, giving rise to 10 000 pairs (local and central) of estimates of the log-HR;
- The sample correlation coefficient between these pairs of estimates will be used for $\hat{\rho}$.

The log-hazard ratio estimates contributing to the auxiliary variable estimate and corresponding variance estimates will be based on stratified Cox proportional hazards models, with stratification based on the randomization stratification factors. The upper bound of a 95% CI for θ_C will be calculated assuming asymptotic normality of $\tilde{\theta}_C$ and using the variance estimator for $\tilde{\theta}_C$ provided in Dodd et al., 2011, i.e. $\hat{V}_{CA}\{1 - \hat{\rho}^2(1-\delta)\}$.

PhRMA method

The early discrepancy rate (EDR) and late discrepancy rate (LDR) will be calculated using the equations below together with information in Table 4-1.

EDR = $(b + a3)/(a + b)$;

LDR = $(c + a2)/(b + c + a2 + a3)$.

Table 4-1 Local versus central disease progression assessments

		Central
Local	PD	No PD
PD	*a* = *a*1 + *a*2 + *a*3	*b*
No PD	*c*	*d*

*a*1: number of agreements on timing and occurrence of PD
*a*2: number of times local PD declared later than central PD
*a*3: number of times local PD declared earlier than central PD

The timing of local and central response assessments will be considered to agree if they occur within ±7 days of each other, aligned with the protocol-specified window for tumor assessments.

4.3 Group sequential design used in Phase III studies

The statistical methodology for the interim OS analyses will be based on group sequential methodology with efficacy stopping boundaries defined by type I error spending functions.

Since the exact number of events available for interim and final analyses cannot be predicted exactly in the clinical trial setting, the group sequential design will be implemented using the α spending function approach. This approach is flexible in dealing with any deviations from the targeted event totals, or unexpected changes to the plan.

If the exact number of events observed at the interim and final analyses deviates from the target numbers described in the protocol, the actual critical boundaries will be derived using the pre-specified error spending functions and the actual numbers of events observed.

- At interim analyses, information fractions will be computed as the ratio of the number of events observed at the considered interim analysis relative to the number targeted for the final analysis, as described in the sample size section of the protocol.
- At the final analysis, the critical value will be calculated using the exact number of observed events at the final cut-off date, considering the α-levels spent at interim analyses and considering the actual correlation among the test statistics, in order to achieve a cumulative type I error smaller than the desired significance level (i.e. smaller than 2.5% for a one-sided test).

4.3.1 Alpha-spending function

The stopping boundaries to be used for the efficacy test will be calculated using the α-spending function approach described in Lan and DeMets (Lan and DeMets, 1983). The spending function for one-sided test has the following functional form:

$$\alpha(t) = 2 - 2\Phi(Z_{\alpha/2} / \sqrt{t})$$

This function generates stopping boundaries that closely resemble the O'Brien-Fleming boundaries (O'Brien and Fleming, 1979).

4.3.2 Methodology

A maximum of three analyses are planned for OS: at the time of the primary analysis for PFS (provided PFS is significant), at which point a total of 123 deaths are expected, after 189 events have been documented, and a final analysis for OS when 252 deaths are expected (expected 40 months from date of first patient to be randomized).

If the exact number of events observed at the interim and final analyses deviates from numbers defined in the protocol, the actual critical boundaries to be used at each of the interim analyses will be recalculated and will be derived from the pre-specified error spending functions using the actual number of events observed and assuming the final events number is the number derived in the sample size section of the protocol. The critical value for the final analysis will be calculated using the exact number of observed events at the cut-off date, and considering the α-levels spent at interim analyses, in order to achieve a cumulative type I error smaller than 2.5%.

It is recognized that circumstances (that are either internal or external to the trial) may require changes in the scheduling of the interim analyses. In case an additional unscheduled interim analysis is requested (e.g. the DMC might request this analysis if the study duration is much longer than expected) the procedure to calculate stopping boundaries needs to be adapted accordingly. An adaptation is also required if the interim analysis is skipped, e.g. is not considered necessary anymore. Both scenarios can be implemented without inflating the type-I error, thanks to the error spending approach used for the group sequential design. The calculation of stopping boundaries needs to be adapted accordingly.

4.3.3 Calculation of stopping boundaries

At the interim and the final analysis the critical boundaries for the group sequential test (stratified log-rank test) will be derived from the predefined spending functions. The calculations will be performed with East 6.3 software.

4.4 Duration of follow-up

Study follow-up will be summarized using the following methods:

- Summary of duration between randomization and cut-off date, and follow-up times for PFS/OS, which are defined as follows:
  - Randomization (recruitment) period = (Date of last patient randomized - Date of first patient randomized + 1) / 30.4375 (months
  - Duration between randomization and data cut-off date = (Cut-off date – Date of randomization + 1) / 30.4375 (months). This item will be summarized overall.
  - Follow-up time = (Date of event or censoring – Date of randomization + 1) / 30.4375 (months) regardless of censoring. Date of censoring is defined as the last adequate tumor assessment date for PFS or last contact date for OS. This item will be summarized by treatment arm.

All summaries will be reported in months. The calculations for PFS will be based on local assessment. Date of censoring is the same as defined for the PFS and OS analysis.

The time from PFS and OS censoring date to data cut-off date will be summarized by time intervals in months: <3, 3 - < 6, 6 - < 12, 12 - < 18, 18 - <24 and by 12 month intervals thereafter if necessary. The gap time is calculated as ([analysis cut-off date] - [censoring date] + 1)/30.4375.

4.5 Confidence intervals for response rate and clinical benefit rate

Response rate and clinical benefit rate will be summarized as percentages with 95% confidence intervals. A standard Wald asymptotic confidence interval, i.e., normal approximation, (implemented using SAS procedure FREQ for one-way tables) will be calculated.

5 References

Amit O, Mannino F, Stone AM, Bushnell W, Denne J, Helterbrand J and Burger HU. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. European Journal of Cancer 2011; 47: 1772-1778.

Bretz F, Posch M, Glimm E, et al (2011). Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes or parametric tests. Biometrical Journal; 53, 894-913.

Bretz, F., Maurer, W., and Hommel, G. (2009) Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. Statistics in Medicine 30, 1489-1501.

Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38, 29 - 41

Clopper, C.; Pearson, E.S. (1934). "The use of confidence or fiducial limits illustrated in the case of the binomial". Biometrika 26: 404–413

Collet D (1994). Modelling survival data in medical research. London, Chapman & Hall

Dodd EL, Korn LE, Freidlin B, Gray R, Bhattacharya S. An audit strategy for progression-free survival. Biometrics, 2011, 67 (3): 1092-99.

Dmitrienko A., Molenberghs G., Chuang-Stein Ch., Offen W. (2005). Analysis of clinical trials using SAS. SAS Institute Inc., SAS press

Dowsett M, Pfister C, Johnston RD, et al. Impact of tamoxifen on the pharmacokinetics and endocrine effects of the aromatase inhibitor letrozole in postmenopausal women with breast cancer. Clin Can Res 1999;5:2338-2343.

Dowsett M, Tobias JS, Howell A et al. (1999) 79(2):311-315. The effect of anastrozole on the pharmacokinetics of tamoxifen in post-menopausal women with early breast cancer. Br J Cancer. 1999;79(2):311-315

Eisenhauer E, et al (2009) New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). European Journal of Cancer, Vol.45: 228-47

Fayers PM, Aaronson NK, Bjordal K, et al, on behalf of the EORTC Quality of Life Group (2001) The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001

FDA briefing Document Oncologic Drugs Advisory Committee Meeting, July 24, 2012, Evaluation of radiologic review of progression-free survival in Non-hematologic malignancies. http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/oncologicdrugsadvisorycommittee/ucm311141.pdf

Fleming TR, Harrington DP, O'Brien PC (1984). Designs for group sequential tests. Controlled Clinical Trials 5: 348-361.

Fleming TR, Rothmann MD, Lu HL (2009). Issues in Using Progression-Free Survival When Evaluating Oncology Products. Journal of Clinical Oncology 27, 2874-2880

Glimm, E., Maurer, W., and Bretz, F. (2010) Hierarchical testing of multiple endpoints in group-sequential trials. Statistics in Medicine 29, 219-228

Hernan M., Brumback B., Robins J. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology,* 11(5), 561-570 (2000).

Hung HM, Wang SJ and O'Neil R (2007) Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. Journal of Biopharmaceutical Statistics 17: 1201-1210.

Hwang, I.K. Shih, W.J. and DeCani, J.S. (1990). Group sequential designs using a family of type I error probability spending functions. Statistics in Medicine 9, 1439-1445

Jonat W, Kaufmann M, Blamey RW, Howell A, Collins JP, Coates A, Eiermann W, Janicke F, Njordenskold B, Forbes JF, Kolvenbag GJCM (1995), A randomized study to compare the effect of the Luteinising Hormone Releasing Hormone (LHRH) analogue Goserelin with or without Tamoxifen in pre- and perimenopausal patients with advanced breast cancer. European Journal of Cancer, 31A(2), 137-142

Klijn JG, Blamey RW, Boccardo F, Tominaga T, Duchateau L and Sylvester R (2001), Combined Tamoxifen and Luteinizaing Hormone-Releasing Hormone (LHRH) agonist versus LHRH agonist alone in premenopausal advanced breast cancer: a meta-analysis of four randomized trials. Journal of Clinical Oncology, 19: 343-353

Lan, K. K. G. & Demets, D. L. (1983). Discrete sequential boundaries for clinical trials. Biometrika 70, 659-63

Lien EA1, Anker G, Lønning PE, Solheim E, Ueland PM. Decreased serum concentrations of tamoxifen and its metabolites induced by aminoglutethimide. Cancer Res. 1990 Sep 15;50(18):5851-7

Mouridsen H, Gershanovich M, Sun Y, et al (2003) Phase III study of letrozole versus tamoxifen as first-line therapy of advanced breast cancer in postmenopausal women: analysis of survival and update of efficacy from the International Letrozole Breast Cancer Group. J Clin Oncol; 21: 2101-09

O'Brien PC, Fleming TR (1979). A multiple testing procedure for clinical trials, Biometrics. 35(3):549-56.

Rabin R, Oemar M, Oppe M (2011) EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. Version 1.0, April 2011. © EuroQol Group 2011

Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993 Nov;4(5):353-65

Schuirmann DJ (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 15(6):657-80.

Zhang JJ, Chen H, He K, Tang S, Justice R, Keegan P, Pazdur R and Sridhara R. Evaluation of blinded independent central review of tumor progression in oncology clinical trials: a meta-analysis. Therapeutic innovation & regulatory science: 2012, 47(2): 167-174.

Appendix 1 CTC grades for laboratory values in Novartis Oncology

CTC grades for laboratory values in Novartis Oncology (based on CTCAE v4.03 – June 2010) **Page 1**

				CTC Grades[1]				
Lab test (toxicity)	**SI unit**	**Lab test (NCDS)**	**Normal ranges (Merck manual, July 2015) and *conversion factors***	**0**	**1**	**2**	**3**	**4**
Hematology								
WBC ↓	10^9/L	WBC	**3.9 – 10.7 x 10^9/L**	≥ LLN	< LLN - 3.0 x 10^9/L	< 3.0 – 2.0 x 10^9/L	< 2.0 – 1.0 x 10^9/L	< 1.0 x 10^9/L
WBC [2] (Leukocytosis)	10^9/L	WBC			-	-	> 100 x 10^9/L	-
Hemoglobin [2] (Anemia)	g/L	HGB	**120 - 160 g/L** or 7.4 - 9.9 mmol/L (F) **140 - 170 g/L** or 8.7 – 10.6 mmol/L (M) *(16.113 x mmol/L = g/L)*	≥ LLN	< LLN - 100 g/L < LLN - 6.2 mmol/L	< 100 - 80 g/L < 6.2 - 4.9 mmol/L	< 80 g/L < 4.9 mmol/L	-
Hemoglobin ↑	g/L	HGB			Increase >0-20 g/L above ULN	Increase >20-40 g/L above ULN	Increase >40 g/L above ULN	-
Platelets ↓	10^9/L	PLAT	**150 - 350 x 10^9/L**	≥ LLN	< LLN - 75.0 x 10^9/L	< 75.0 - 50.0 x 10^9 L	< 50.0 - 25.0 x 10^9/L	< 25.0 x 10^9/L
Neutrophils [3]↓	10^9/L	NEUT		≥2x10^9/L	< 2.0 - 1.5 x 10^9/L	< 1.5 - 1.0 x 10^9/L	< 1.0 - 0.5 x 10^9/L	< 0.5 x 10^9/L
Lymphocytes [3]↓	10^9/L	LYM		≥1.5x10^9/L	< 1.5 - 0.8 x 10^9/L	< 0.8 - 0.5 x 10^9/L	< 0.5 - 0.2 x 10^9/L	< 0.2 x 10^9/L
Lymphocytes ↑	10^9/L	LYM			-	> 4 - 20 x 10^9/L	> 20 x 10^9/L	-
Biochemistry								
AST ↑	U/L	AST	**0 - 35 U/L** or 0 – 0.58 ukat/L *(60 x ukat/L = U/L)*	≤ ULN	> ULN – 3.0 x ULN	> 3.0 - 5.0 x ULN	> 5.0 - 20.0 x ULN	> 20.0 x ULN
ALT ↑	U/L	ALT	**0 - 35 U/L** or 0 – 0.58 ukat/L *(60 x ukat/L = U/L)*	≤ ULN	> ULN – 3.0 x ULN	> 3.0 - 5.0 x ULN	> 5.0 - 20.0 x ULN	> 20.0 x ULN
Total bilirubin ↑	umol/L	BILI	**5.1 – 20.5 umol/L** or 0.3 – 1.2 mg/dL *(17.1 x mg/dL = umol/L)*	≤ ULN	> ULN - 1.5 x ULN	> 1.5 - 3.0 x ULN	> 3.0 - 10.0 x ULN	> 10.0 x ULN
Alk. Phosphatase ↑	U/L	ALP	**36 - 92 U/L** or 0.5 - 1.5 ukat/L *(60 x ukat/L = U/L)*	≤ ULN	> ULN - 2.5 x ULN	> 2.5 - 5.0 x ULN	> 5.0 - 20.0 x ULN	> 20.0 x ULN
Creatinine [4]↑	umol/L	CREAT	**61.9 - 115 umol/L** or 0.7 – 1.3 mg/dL *(88.4 x mg/dL = umol/L)*	≤ ULN	> ULN - 1.5 x ULN	> 1.5 - 3.0 x ULN	> 3.0 - 6.0 x ULN	> 6.0 x ULN
Creatinine kinase [4]↑	U/L	CK	**30 - 170 U/L** or 0.5 – 2.83 ukat/L *(60 x ukat/L = U/L)*	≤ ULN	> ULN - 2.5 x ULN	> 2.5 - 5.0 x ULN	> 5.0 - 10.0 x ULN	> 10.0 x ULN
Albumin [2] (Hypoalbuminemia)	g/L	ALB	**35 - 55 g/L** or 3.5 to 5.5 g/dL	≥ LLN	< LLN - 30 g/L	< 30 - 20 g/L	< 20 g/L	-
Total Cholesterol ↑	mmol/L	CHOL	**3.88 – 5.15 mmol/L** or 150 - 199 mg/dL *(38.67 x mg/dL = mmol/L)*	≤ ULN	> ULN - 7.75 mmol/L > ULN - 300 mg/dL	> 7.75 -10.34 mmol/L > 300 – 400 mg/dL	>10.34-12.92 mmol/L > 400 – 500 mg/dL	>12.92 mmol/L > 500 mg/dL
Lipase ↑	U/L	LIPASE	**<95 U/L** or <1.58 ukat/L *(60 x ukat/L = U/L)*	≤ ULN	> ULN - 1.5 x ULN	> 1.5 - 2.0 x ULN	> 2.0 - 5.0 x ULN	> 5.0 x ULN
Amylase ↑	U/L	AMYLASE	**0 - 130 U/L** or 0 – 2.17 ukat/L *(60 x ukat/L = U/L)*	≤ ULN	> ULN - 1.5 x ULN	> 1.5 - 2.0 x ULN	> 2.0 - 5.0 x ULN	> 5.0 x ULN
Uric acid [2] (Hyperuricemia)	umol/L	URATE	**150 - 470 umol/L** or 2.5 – 8 mg/dL *(59.48 x mg/dL = umol/L)*	≤ ULN	> ULN – 10 mg/dL > ULN – 595 umol/L	-	-	> 10 mg/dL > 595 umol/L

ULN = Upper Limit of Normal range; LLN = Lower Limit of Normal range

LAB - CTC grades in Novartis Oncology (26Oct15)

CTC grades for laboratory values in Novartis Oncology (based on CTCAE v4.03 – June 2010)

Page 2

				CTC Grades[1]				
Lab test (toxicity)	**SI unit**	**Lab test (NCDS)**	**Normal ranges (Merck manual, July 2015) and *conversion factors***	**0**	**1**	**2**	**3**	**4**
Phosphorus[2] (Hypophosphatemia)	mmol/L	PHOS	**0.97 – 1.45 mmol/L** or 3.0 - 4.5 mg/dL *(0.32 x mg/dL = mmol/L)*	≥ LLN	< LLN - 2.5 mg/dL < LLN - 0.8 mmol/L	< 2.5 - 2.0 mg/dL < 0.8 - 0.6 mmol/L	< 2.0 - 1.0 mg/dL < 0.6 - 0.3 mmol/L	< 1.0 mg/dL < 0.3 mmol/L
Calcium (corrected) [2] (Hypercalcemia)	mmol/L	CACALC	**2.2 - 2.6 mmol/L** or 9 - 10.5 mg/dL *(0.2495 x mg/dL = mmol/L)*	≤ ULN	> ULN - 11.5 mg/dL > ULN - 2.9 mmol/L	> 11.5 - 12.5 mg/dL > 2.9 - 3.1 mmol/L	> 12.5 - 13.5 mg/dL > 3.1 - 3.4 mmol/L	> 13.5 mg/dL > 3.4 mmol/L
Calcium (corrected) [2] (Hypocalcemia)	mmol/L	CACALC		≥ LLN	< LLN - 8.0 mg/dL < LLN - 2.0 mmol/L	< 8.0 - 7.0 mg/dL < 2.0 - 1.75 mmol/L	< 7.0 - 6.0 mg/dL < 1.75 - 1.5 mmol/L	< 6.0 mg/dL < 1.5 mmol/L
Magnesium [2] (Hypermagnesemia)	mmol/L	MG	**0.62 – 0.99 mmol/L** or 1.5 – 2.4 mg/dL *(0.4114 x mg/dL = mmol/L)*	≤ ULN	> ULN - 3.0 mg/dL > ULN - 1.23 mmol/L	-	> 3.0 – 8.0 mg/dL > 1.23 – 3.3 mmol/L	> 8.0 mg/dL > 3.3 mmol/L
Magnesium [2] (Hypomagnesemia)	mmol/L	MG		≥ LLN	< LLN - 1.2 mg/dL < LLN - 0.5 mmol/L	< 1.2 - 0.9 mg/dL < 0.5 - 0.4 mmol/L	< 0.9 - 0.7 mg/dL < 0.4 - 0.3 mmol/L	< 0.7 mg/dL < 0.3 mmol/L
Glucose (non-fasting) [2] (Hyperglycemia)	mmol/L	GLUCSN	**<7.8 mmol/L** or <140 mg/dL *(0.05551 x mg/dL = mmol/L)*	≤ ULN	-	> ULN - 250 mg/dL > ULN - 13.9 mmol/L	> 250 - 500 mg/dL > 13.9 - 27.8 mmol/L	> 500 mg/dL > 27.8 mmol/L
Glucose (fasting) [2] (Hyperglycemia)	mmol/L	GLUCSF	**3.9 – 5.8 mmol/L** or 70 - 105 mg/dL *(0.05551 x mg/dL = mmol/L)*	≤ ULN	> ULN - 160 mg/dL > ULN - 8.9 mmol/L	> 160 - 250 mg/dL > 8.9 - 13.9 mmol/L	> 250 - 500 mg/dL > 13.9 - 27.8 mmol/L	> 500 mg/dL > 27.8 mmol/L
Glucose [2] (Hypoglycemia)	mmol/L	GLUCSN/ GLUCSF		≥ LLN	< LLN - 55 mg/dL < LLN - 3.0 mmol/L	< 55 - 40 mg/dL < 3.0 - 2.2 mmol/L	< 40 - 30 mg/dL < 2.2 - 1.7 mmol/L	< 30 mg/dL < 1.7 mmol/L
Potassium [2] (Hyperkalemia)	mmol/L	K	**3.5 - 5.0 mmol/L** *(0.2558 x mg/dL = mEq/L = mmol/L)*	≤ ULN	> ULN - 5.5 mmol/L	> 5.5 - 6.0 mmol/L	> 6.0 - 7.0 mmol/L	> 7.0 mmol/L
Potassium [2] (Hypokalemia)	mmol/L	K		≥ LLN	< LLN - 3.0 mmol/L	-	< 3.0 - 2.5 mmol/L	< 2.5 mmol/L
Sodium [2] (Hypernatremia)	mmol/L	SODIUM	**136 - 145 mmol/L** *(0.435 x mg/dL = mEq/L = mmol/L)*	≤ ULN	> ULN - 150 mmol/L	> 150 - 155 mmol/L	> 155 - 160 mmol/L	> 160 mmol/L
Sodium [2] (Hyponatremia)	mmol/L	SODIUM		≥ LLN	< LLN - 130 mmol/L	-	< 130 - 120 mmol/L	< 120 mmol/L
Triglyceride [2] ↑	mmol/L	TRIG	**< 2.82 mmol/L** or < 250 mg/dL *(0.01129 x mg/dL = umol/L)*	< 150 < 1.71	≥ 150 - 300 mg/dL ≥ 1.71 – 3.42 mmol/L	> 300 - 500 mg/dL > 3.42 – 5.7 mmol/L	> 500 - 1000 mg/dL > 5.7 – 11.4 mmol/L	> 1000 mg/dL > 11.4 mmol/L
Coagulation								
INR [2]↑	1	INR	**0.8 – 1.2**	≤ ULN	> ULN - 1.5 x ULN	> 1.5 - 2.5 x ULN	> 2.5 x ULN	-
Activated partial thromboplastin time[2]↑	sec	APTT	**25 - 35 sec**	≤ ULN	> ULN - 1.5 x ULN	> 1.5 - 2.5 x ULN	> 2.5 x ULN	-
Fibrinogen [4]↓	g/L	FIBRINO	**1.5 – 3.5 g/L** or 150 – 350 mg/dL *(0.01 x mg/dL = g/L)*	≥ LLN	< LLN - 0.75 x LLN	< 0.75 - 0.5 x LLN	< 0.5 - 0.25 x LLN	< 0.25 x LLN

ULN = Upper Limit of Normal range; LLN = Lower Limit of Normal range

(1) = **LAB CTC grades 1, 2, 3, 4 overrule the study specific (central or local) normal range criteria**, e.g. if ULN of Sodium is 151 mmol/L and the value is 151 mmol/L, CTC grade 2 is assigned although the value is ≤ ULN.

(2) = **Life-threatening consequences and/or hospitalization are <u>not</u> considered** for determination of LAB CTC grades 3 and 4. Concomitant usage of anticoagulation therapy (for INR and Fibrinogen) is not considered either.

(3) = Values and LNRs for blood differentials can be given as %, absolute values should then be calculated using WBC. Generally, **≥ 1.5 x 10^9 /L (lymphocytes) and ≥ 2 x 10^9 /L (neutrophils) are considered as LAB CTC grade 0**

(4) = For Creatinine and Fibrinogen, the **comparison with baseline is <u>not</u> considered** for derivation of LAB CTC grades