NOVARTIS

Clinical Development

BYL719 (alpelisib)

Study Number: CBYL719C2301 / NCT02437318

SOLAR-1: A phase III randomized double-blind, placebo controlled study of alpelisib in combination with fulvestrant for men and postmenopausal women with hormone receptor positive, HER2-negative advanced breast cancer which progressed on or after aromatase inhibitor treatment

Statistical Analysis Plan (SAP)

Author:	[redacted]
Document type:	SAP Documentation
Document status:	Final version number 2.0
Release date:	03 July 2018
Number of pages:	102

Document History – Changes compared to previous version of SAP.

Version	Date	Changes
Amendment 1	17th October 2016	Changes due to Protocol Amendments #1 and #2; new Novartis protocol template text; new Novartis SAP template language; new Novartis guidelines for safety analyses; updated Novartis standards for data reporting; updated Novartis RECIST guidelines; minor clarifications and corrections
Amendment 2	23rd March 2017	Design/endpoint changes due to Protocol Amendment 3; Novartis guidelines for safety analyses; updated Novartis standards for data reporting; minor clarifications and corrections
Amendment 3	27 November 2017	Section 2.1.8 Baseline • Added clarification that tumor assessments (and investigator's overall response) occurring within 7 days following treatment start date may considered as the baseline RECIST assessment. This allows for the fact that subjects have a +/- 7 day window around their scheduled tumor assessments Section 2.2 Data included in the analysis • Added clarification on the respective cut-off dates used for the efficacy analysis in each cohort: the interim analysis in the mutant cohort and final analysis in the non-mutant cohort occurred at different time points. The cut-off for safety in both cohorts was set at the same time point as the interim analysis in the mutant cohort. • Clarified that any data collected in the clinical database after a subject withdraws informed consent from all further participation in the trial, will not be included in the analysis or analysis data sets Section 2.4.5 Determination of missing adequate assessments • Amended the window for the baseline RECIST assessment, given the change regarding baseline definition outlined in Section 2.1.8 Section 3.2.1 Basic demographic and background data • Analysis using summary statistics updated to use Novartis standards Section 3.2.4 Prior anti-neoplastic therapy • Added definition to provide summary of patient population in terms of primary & secondary endocrine resistance and endocrine sensitivity status Section 3.6.1 Duration of study treatment exposure • Amended study treatment exposure categories Section 3.8 Efficacy Evaluation • Clarified that for central radiology assessments, only the final adjudicated data will be listed. Section 3.8.1.5 Sensitivity analyses of the primary endpoint (Different Censoring Mechanisms) • Updated the PFS censoring options to match the correct option in Table 3-3.

Version	Date	Changes
		Section 3.8.1.6 Supportive analyses of the primary endpoint (PFS assessed by BIRC) • Clarify that a negative observed differential discordance for the EDR or a positive differential discordance for the LDR meeting the 15% threshold is required to trigger a full BIRC review (per Amit et al, 2011) • Clarified that the timing of local and central response assessment (for subjects with complete agreement of local and central sources) will be considered to agree if they occur within ±7 days of each other, aligned with the protocol-specified window for tumor assessments Section 3.8.2.5 Overall response rate • Added new analysis to present ORR in the subset of patients with measurable disease at baseline Section 3.8.2.6 Clinical benefit rate • Added new analysis to present CBR in the subset of patients with measurable disease at baseline Section 3.8.2.9 Patient reported outcomes • Clarified that data collection for PRO are scheduled up until disease progression and not just end of study treatment (as per protocol) • Amended text for inclusion data into statistical model. Model fit based on unblinded data will be assessed to decide which time points to include  Section 3.9.1.4 AE summaries • Removed table for CTC Grade 3 or 4 AEs as this data is presented elsewhere Section 3.9.3 Vital Signs • Removed summary statistics for change from baseline to the worst post-baseline vital signs value Section 3.9.4 ECGs • RR parameter is not collected in the study • Removed analysis using summary statistics and shift tables • Table 3-8: corrected parameters as per Novartis Oncology Guideline for Safety Analyses Section 3.10 Pharmacokinetic (PK) analyses

Version	Date	Changes
		• Clarified the use of all available PK samples to derive empirical Bayes estimates however reporting C1D15 and steady-state assessments only in the summary tables of PopPK-derived individual exposures. Section 3.10.3 Population PK parameters for alpelisib • Clarified 3-step procedure to obtain individual exposures from PopPK-derived empirical Bayes estimates: o Use of the most recent version of the Phase I PopPK model instead of the model published by de Buck et. al. o Restriction to patients with sparse PK profile sampling only o Specification of the parameters to be derived • Predicted concentrations will not be reported, only the individual parameters (see TFLs) [redacted] Section 3.13 Subgroup analyses • Clarified age categories for geriatric subgroups • Removed gender as a subgroup analysis as number of male subjects randomized was too low (one male patient randomized in mutant cohort) • Clarified that CDISC controlled terminology will be used for subgroup analyses based on race (and not ethnicity)

Version	Date	Changes
		• Clarified definition of region based on countries with randomized patients • Clarified definition of first and second line therapy groups • Clarified that prior chemotherapy use is based on the last setting prior to study entry • Added efficacy subgroup analysis in patients with primary- & secondary endocrine resistance vs endocrine sensitivity Section 3.15 Sample Size Calculation (Overall Survival for Patients with PIK3CA non-mutant status (secondary endpoint) • Corrected reference to primary endpoint for overall survival in this cohort Section 4.8 Audit-based BIRC assessment of PFS (PhRMA method) • Clarified that the timing of local and central response assessment (for subjects with complete agreement of local and central sources) will be considered to agree if they occur within ±7 days of each other, aligned with the protocol-specified window for tumor assessments
Amendment 4	03 July 2018	Section 3.6.1 Duration of study treatment exposure • New summaries for treatment exposure are added for specific subgroups based on AESI Section 3.9.1.5 Adverse Events of Special Interest • An analysis for time to first onset of CTC Grade 3 or worse is added for Hyperglycemia and Rash (AESI) Section 3.9.2.1 Blood glucose parameters • Section added to present new analyses for time to onset and time to resolution of hyperglycemia based on plasma glucose Section 3.10.3 Population PK parameters for alpelisib • The population PK analysis will be provided in a separate report. [redacted] Section 3.13.1 Safety [subgroups] • For the AESI of Hyperglycemia only, a subgroup analysis by hyperglycemia diagnosis status at baseline per American Diabetes Association (ADA) 2017 will be presented Section 3.13.2 Efficacy [subgroups]

Version	Date	Changes
		• Subgroup analyses added for the primary endpoint by (i) PIK3CA somatic mutation and (ii) PIK3CA somatic mutation by location

Table of contents

List of tables

List of figures

List of abbreviations

AE	Adverse event
AESI	Adverse event of Special Interest
ATC	Anatomical Therapeutic Chemical
BIRC	Blinded Independent Review Committee
BOR	Best overall response
[redacted]	[redacted]
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
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer's core quality of life questionnaire
EOT	End of treatment
[redacted]	[redacted]
FAS	Full analysis set
HER2	Human epidermal growth factor receptor 2
HR.	Hazard Ratio
HRQoL	Health-Related Quality of Life

IRT	Interactive Response Technology
LLOQ	Lower Limit of Quantification
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
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
PFS	Progression-free survival
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
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
SMQ	Standardized MedDRA queries
SOC	System organ class
TA	Tumor assessment
TEAE	Treatment-emergent adverse event
UNK	Unknown

1 Introduction

This document describes the detailed statistical methodology to be used for the primary Clinical Study Report (CSR) of study CBYL719C2301, based on the interim or final PFS analysis in the PIK3CA mutant cohort (primary objective).

The content of this SAP is based on protocol CBYL719C2301 Amendment 3. All decisions regarding analysis, as defined in the SAP document, have been made prior to database lock and unblinding of the study data.

This section provides an introduction to the document and describes the study design and objectives as outlined in the Clinical Study Protocol. Section 2 provides definitions and the general methodology that will be used to analyze data. Section 3 describes the analyses and summaries that will be produced. Finally, Section 4 provides more detailed specifications on the statistical methodology used.

CSR deliverables (shells for tables, figures, listings) and further programming specifications are described in SAP documents ”Tables, Figures and Listings (TFL) Shells” and “Programming Dataset Specifications”, respectively.

1.1 Study Design

This is a phase III, randomized, double-blind, placebo-controlled study of BYL719 (alpelisib) or placebo in combination with fulvestrant for the treatment of men and premenopausal women with hormone receptor positive, HER2-negative, advanced breast cancer which progressed on or after aromatase inhibitor treatment.

A total of approximately 560 patients will be enrolled; in which approximately 340 and 220 patients will be enrolled respectively to each of two cohorts: *PIK3CA* mutant and *PIK3CA* non-mutant. Within each of these two cohorts, randomization will be stratified by:

1. Lung and/or liver metastases (yes versus no)
2. Previous treatment with any CDK4/6 inhibitor (yes versus no)

PFS in the PIK3CA mutant cohort, as assessed by the local radiologists/investigators and using RECIST 1.1 criteria will be the primary endpoint. PFS in the PIK3CA mutant cohort as assessed through Blinded Independent Review Committee (BIRC) using an audit-based approach will be used for supportive evidence of the primary efficacy endpoint.

An independent Data Monitoring Committee (DMC) will monitor semi-blinded safety and efficacy data during the trial. A separate DMC SAP specifies the analyses to be performed for the DMC reviews.

One futility interim analysis is planned for the primary efficacy endpoint (PFS) in the PIK3CA mutant cohort at the expected time given in Section 3.12.1. There is no intention to stop for superiority at this interim analysis. Another interim analysis that allows the study to stop for superior efficacy is planned in the *PIK3CA* mutant cohort, after all patients have been randomized and approximately 75% of the total PFS events have been documented, as per local assessments. If PFS is statistically significant, interim analyses for OS in the PIK3CA mutant cohort will also be conducted as detailed in Section 3.12.2.

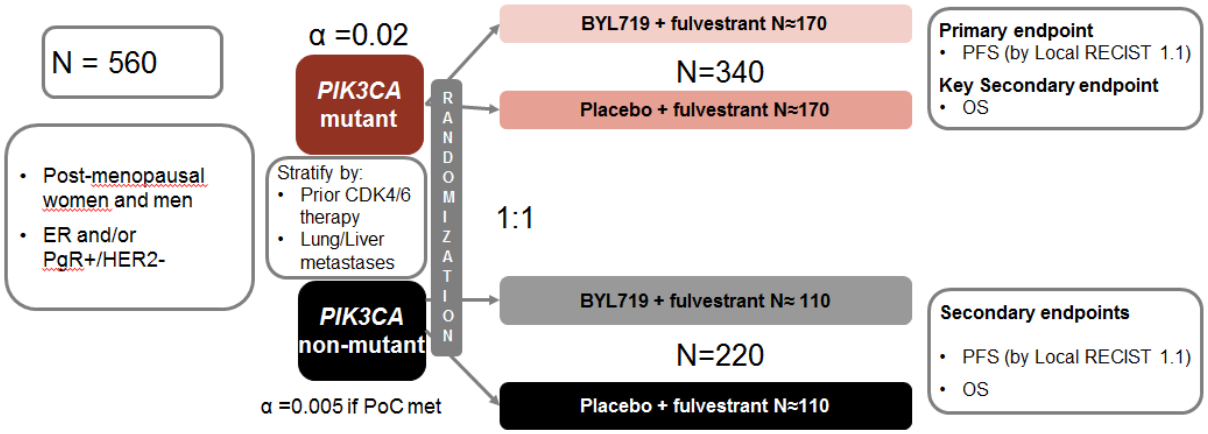
Overall survival in the PIK3CA mutant cohort is a key secondary endpoint in this study and will be analyzed provided the primary endpoint PFS is statistically significant. The type I error rate for OS will be controlled by using a separate Lan-DeMets (O'Brien-Fleming) alpha spending function for OS, independent of the Haybittle-Peto boundary used for the primary efficacy analysis of PFS.

PFS and Overall survival in the PIK3CA non-mutant cohort are secondary endpoints in this study. OS will be tested provided the PFS is statistically significant in this cohort.

The futility and efficacy interim analyses for PFS in the PIK3CA mutant cohort and the final PFS analysis in the PIK3CA non-mutant cohort will be performed by an independent statistician external to Novartis and the results will be provided to the DMC by the independent statistician. Novartis will be unblinded if the study stops early for futility or efficacy in the PIK3CA mutant cohort or at the time of the final PFS analysis in the PIK3CA mutant cohort. Further details regarding the group sequential design are provided in Section 4.5.

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		Refer to Section
To determine whether treatment with alpelisib in combination with fulvestrant prolongs PFS compared to treatment with placebo in combination with fulvestrant for patients with *PIK3CA* mutant status	PFS based on local radiology assessments and using RECIST 1.1 criteria in the *PIK3CA* mutant cohort	3.8.1

Objective	**Endpoint**	**Analysis**
Key secondary		Refer to Section
To determine whether treatment with alpelisib in combination with fulvestrant prolongs overall survival (OS) compared to treatment with placebo in combination with fulvestrant for patients with *PIK3CA* mutant status	OS in the *PIK3CA* mutant cohort	3.8.2.1
Other secondary		Refer to Section
To establish proof of concept of treatment benefit with alpelisib in combination with fulvestrant with respect to PFS for patients with PIK3CA non-mutant status	PFS based on local radiology assessments and using RECIST 1.1 criteria in the *PIK3CA* non-mutant cohort	3.8.2.2
To evaluate the two treatment arms with respect to OS for patients with *PIK3CA* non-mutant status	OS in the *PIK3CA* non-mutant cohort	3.8.2.3
● To evaluate the two treatment arms and cohorts of interest with respect to overall response rate (ORR), clinical benefit rate.	ORR and CBR in each of the *PIK3CA* mutant and non-mutant cohorts	3.8.2.5 3.8.2.6
● To evaluate the two treatment arms and cohorts of interest with respect to time to deterioration of ECOG performance status.	Time to definitive deterioration of the ECOG performance status of the score from baseline in each of the *PIK3CA* mutant and non-mutant cohorts	3.8.2.8
● To evaluate the safety and tolerability of alpelisib in combination with fulvestrant	- Type, frequency and severity of adverse events per CTCAEv4.03 - Type, frequency and severity of laboratory toxicities per CTCAEv4.03	3.9
● To evaluate change in global health status/QOL in the two treatment arms and cohorts of interest	- 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 in each of the *PIK3CA* mutant and non-mutant cohorts	3.8.2.9

Objective	Endpoint	Analysis
• To characterize the pharmacokinetics (PK) of fulvestrant and alpelisib when given in combination with fulvestrant.	Summary statistics for PK: plasma concentration-time profiles of alpelisib given in combination with fulvestrant and appropriate individual PK parameters based on population PK model Summary statistics of fulvestrant trough plasma concentrations in each treatment arm (alpelisib/placebo)	3.10
• To evaluate the association between PIK3CA mutation status as measured in ctDNA at baseline with PFS upon treatment with alpelisib.	PFS based on local radiology assessments and using RECIST 1.1 criteria for each of (i) patients with *PIK3CA* mutant status and (ii) patients with *PIK3CA* non-mutant status as measured in ctDNA at baseline.	3.8.2.7
[redacted]		

2 Definitions and general methodology

2.1 Definitions

2.1.1 Study drug and study treatment

Study drug refers to alpelisib, alpelisib matching placebo or fulvestrant.

Study treatment refers to alpelisib in combination with fulvestrant or alpelisib matching placebo in combination with fulvestrant.

Alpelisib matching placebo will be referred to as "placebo" in the remainder of this document.

2.1.2 Date of first administration of study drug

The date of first administration of study drug is derived as the first date when a non-zero dose of study drug (alpelisib/placebo or fulvestrant) is administered.

For the sake of simplicity, the date of first administration of study drug is referred to start date of study drug. Start date of study drug is defined for each drug which is part of study treatment.

The date of first administration for alpelisib/placebo or fulvestrant is recorded on the corresponding "Dosage Administration Record" (DAR) eCRF.

Note: Dates from "DAR –PK sampling" eCRF will not be used for this derivation.

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 non-zero dose of study drug is administered and recorded on the DAR eCRF.

This date will also be referred as ***last date of study drug.*** Last date of study drug is defined for each drug which is part of study treatment.

The date of last administration for alpelisib/placebo or fulvestrant is recorded on the corresponding "DAR" eCRF.

Note 1: Dates from "DAR –PK sampling" eCRF will not be used for this derivation.

Note 2: Last date of study drug exposure may not be the same as the last date of study drug (see Section 2.1.6).

2.1.4 Date of first administration of study treatment

The date of first administration of study treatment is derived as the first date when a non-zero dose of any component of the study treatment (alpelisib/placebo or fulvestrant) is administered

For the sake of simplicity, the date of first administration of study treatment will also be referred as start date of study treatment.

Note: Dates from 'DAR – PK sampling' eCRF will not be used for this derivation.

For example: if the first dose of alpelisib/placebo is taken on 05JAN2011, and first dose of fulvestrant is taken on 03JAN2011, then the date of first administration of study treatment is 03JAN2011.

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 non-zero dose of ***any component*** of the study treatment (alpelisib/placebo or fulvestrant) is administered.

Note: Dates from 'DAR – PK sampling' eCRF will not be used for this derivation.

For example: if the last dose of alpelisib/placebo is taken on 15APR2011, and the last dose of fulvestrant is taken on 17APR2011, then the date of last administration of study treatment is on 17APR2011.

2.1.6 Last date of exposure to study drug/treatment

The study treatment schedule is organized in cycles of 28 days.

The ***last date of exposure to study treatment*** is derived to be the latest date of the last date of exposure to alpelisib/placebo and fulvestrant. The last date of exposure to alpelisib/placebo and last date of exposure to fulvestrant will be derived as follows.

Alpelisib/placebo is administered daily on a continuous once daily dosing schedule. Hence, the ***last date of exposure to alpelisib/placebo*** is the date of last administration of alpelisib/placebo.

Fulvestrant is administered on

- Cycle 1 Day 1
- Cycle 1 Day 15, and
- The first day of every cycle thereafter (e.g. Cycle 2 Day 1, Cycle 3 Day 1 etc.)

Due to the irregularly spaced fulvestrant dose administration, the ***last date of exposure to fulvestrant*** is calculated using a different method depending on the cycle at which fulvestrant was discontinued:

1. If the patient discontinues fulvestrant between Cycle 1 Day 1 and Cycle 1 Day 14 inclusive:
   - The last date of exposure to fulvestrant is calculated as (last date of administration of fulvestrant) + (length of time interval to next scheduled dose - 1) i.e. [last date of fulvestrant administration+ (14-1)].
   - If the patient died or was lost to follow-up within last date of administration of fulvestrant + 13 days, the last date of exposure to fulvestrant is the date of death or the date of last contact, respectively.
2. If the patient discontinues fulvestrant between Cycle 1 Day 15 and Cycle 2 Day 1:
   - The last date of exposure to fulvestrant is calculated as (last date of administration of fulvestrant) + (length of time interval - 1) i.e. [last date of fulvestrant administration+ (14-1)].
   - If the patient died or was lost to follow-up within last date of administration of fulvestrant + 13 days, the last date of exposure to fulvestrant is the date of death or the date of last contact, respectively.
3. If the patient discontinues fulvestrant on or after Cycle 2 Day 1, then:

- the last date of exposure to fulvestrant is calculated as (last date of administration of fulvestrant) + (length of time interval - 1) i.e. [last date of fulvestrant administration+ (28-1)].
- If the patient died or was lost to follow-up within last date of administration of fulvestrant + 27 days, the last date of exposure to fulvestrant is the date of death or the date of last contact, respectively.

'Date of last administration of study drug' and 'Date of last contact' are defined in sections 2.1.3 and 2.1.10 respectively.

2.1.7 Study day

The study day, describes the day of the event or assessment date, relative to the reference start date (randomization date or start date of study treatment).

The reference start date is designated as Study Day 1. Study Day –1 is the day that precedes Day 1. Study Day 0 is not defined. Study day is not to be used in numerical computations, for example in calculating exposure.

The study day will be calculated as follows:

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

The reference start date ***for all safety assessments*** (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption etc.) and PK data will be the start date of study treatment.

The reference start date ***for all efficacy assessments*** (e.g. tumor assessment, death, ECOG performance status, PRO) will be the randomization date.

For any ***non-safety screening assessments or events*** such as baseline disease characteristics or medical history (e.g., time since diagnosis of disease) [redacted] that occurred prior to randomization the reference start date will be the randomization date.

The study day will be displayed in the data listings.

2.1.8 Baseline

Baseline is the result of an investigation describing the "true" uninfluenced state of the patient, defined as the period from the date of signing any informed consent document to the start date of study treatment or the date of randomization. Assessments, specified to be collected post-dose on the first date of treatment are not considered as baseline values.

For efficacy evaluations, the last non-missing assessment, including unscheduled assessments on or before the date of randomization is taken as "baseline" value or "baseline" assessment. For RECIST-based endpoints including PFS, ORR, CBR, time to response and duration of response, a window of 7 days from the start of study treatment will be allowed, i.e. the

investigator/BIRC-reported responses will be maintained and baseline considered valid if the baseline assessment is within 7 days of treatment start date. In the context of baseline definition, the efficacy evaluations also include ECOG performance status and PRO. If a patient has 2 ECOG PS (or PRO) values at the same date, the worst ECOG PS value will be taken as 'baseline'. [redacted]

For safety evaluations (i.e. laboratory assessments, ECGs and vital signs), the last available assessment, including unscheduled assessments on or before the start date of study treatment (Cycle 1 Day 1) as described in Section 2.1.4 is taken as "baseline" value or ''baseline" assessment.

In rare cases, where multiple safety measurements meet the baseline definition, with no further flag or label that can identify the chronological order, then the following rule should be applied: If values are from central and local laboratories, the value from central assessment should be considered as baseline. If multiple values are from the same laboratory (local or central) or collected for ECGs or vital signs, then the last value should be considered as baseline.

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

2.1.9 On-treatment assessment/event

Safety summary tables and selected summaries of deaths will summarize only on-treatment assessments/events.

An on-treatment adverse event is defined as any adverse event reported in the following time interval (including the lower and upper limits):

- date of first administration of study treatment; date of last administration of study treatment + 30 days

In addition, an AE that started in the screening phase and was ongoing in the treatment phase will not be summarized unless it has worsened in severity.

An on-treatment assessment is defined as any assessment performed after the date of first administration of study treatment i.e. assessments performed in the following time interval (including the lower and upper limits):

- date of first administration of study treatment + 1; date of last administration of study treatment + 30 days

If the last date of study treatment is missing, on-treatment assessments/events include any assessment/event recorded in the database and which occur after the start date of study treatment.

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

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.

2.1.10 Last contact date

The last contact date is derived for patients not known to have died at the analysis cut-off date based on the latest date among the following:

- Assessment dates (e.g. laboratory, vital signs, ECOG performance status/PRO, ECG, cardiac imaging, tumor imaging, PK assessment, EOT completion etc.).
- Medication and procedures dates including study medication, concomitant medications, surgical and medical procedures, antineoplastic therapies administered after study drug discontinuation (with non-missing medication/procedure term).
- Adverse event start and end dates (with non-missing verbatim AE term present)
- "Last known date patient alive" collected on the "Survival information" eCRF
- Study treatment start/end date
- Randomization date

The last contact date is defined as the latest complete date from the above list or the cut-off date, whichever comes first. 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. Partial date imputation is allowed for event (death)/censoring is coming from 'Survival information' eCRF.

The last contact date is used for censoring of patients in the analysis of overall survival and analyses for time to onset of adverse events.

2.1.11 Screening failure

Screening failures are patients who have signed informed consent and failed screening criteria in the study. These patients will not be enrolled into the treatment phase.

2.1.12 Time Units

A month length is 30.4375 days (365.25 / 12). If duration is to be reported in months, duration in days is divided by 30.4375. If duration is to be reported in years, duration in days will be divided by 365.25.

2.2 Data included in the analysis

A unique cut-off date will be established after the targeted number of events for each of the planned interim and final efficacy analyses has been documented.

For the primary analysis CSR, all efficacy analyses will be based on separate cut-off dates determined for each of the PIK3CA mutant and non-mutant cohorts, based on the target number of PFS events specified in Section 3.12 (i.e. final PFS analyses data cut-off date for PIK3CA non-mutant cohort; interim efficacy or final PFS analyses data cut-off date for PIK3CA mutant cohort). For the primary analysis CSR, all safety analyses for both PIK3CA cohorts will be based on the data cut-off date for the interim efficacy or final PFS analysis in the PIK3CA mutant cohort.

For each of the analyses, all statistical analyses will be performed using all data collected in the database up to the data cut-off date. All data with an assessment date or event start date (e.g. vital sign assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.

All events with start date before or on the cut-off date and end date after the cut-off date will be reported as 'continuing at the cut-off date'. The same rule will be applied to events starting before or on the cut-off date and not having documented end date. This approach applies, in particular, to adverse event and concomitant medication reports. For these events, the end date will not be imputed and therefore will not appear in the listings.

Interim review of safety data will be provided in each cohort as well as pooled across both cohorts, at each DMC meeting where safety data is reviewed. It is planned that the first DMC safety review meeting will be based on at least 2 months' exposure data from approximately the first fifty patients, or available data after at least 6 months after the first patient is randomized, whichever date occurs first. Subsequent formal safety reviews will be based on data available every six months (+/-30 days window) after the first meeting, unless otherwise requested by the DMC.

Withdrawal of Informed Consent

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

2.2.1 Pharmacokinetic (PK) data review

No PK data review is planned by DMC.

2.3 Analysis sets

2.3.1 Full analysis set (FAS)

The Full analysis set (FAS) comprises all patients who were randomized to study treatment (alpelisib + fulvestrant or matching placebo + fulvestrant). According to the intent to treat principle, patients will be analyzed according to the cohort, treatment and strata they have been assigned to during the randomization procedure. FAS will be the main population for analyses of patient disposition, demographics and other baseline characteristics. The FAS will be the primary population for the efficacy analyses. Patients who do not provide main study informed consent will be excluded from the FAS.

2.3.2 Safety set

The Safety Set includes all patients who received any study treatment. Patients will be analyzed according to the study treatment received, where treatment received is defined as the randomized treatment if the patient took at least one dose of that treatment or the first treatment received if the randomized treatment was never received.

2.3.3 Per protocol set

The Per-protocol set (PPS) comprises all patients in the FAS for the PIK3CA mutant cohort who do not have any protocol deviations that could confound the interpretation of the primary analyses conducted on the FAS. The PPS will be used to perform sensitivity analysis for the primary efficacy endpoint (i.e. PFS in the PIK3CA mutant cohort) if the primary endpoint is statistically significant. Patients with any of the following protocol deviations will be excluded from the PPS.

Exclusion Criteria for Per Protocol Set

- Patient did not provide main study informed consent
- Patient does not have HER2 negative breast cancer (Protocol Inclusion Criterion 7)
- Patient does not have histologically and/or cytologically confirmed diagnosis of ER and/or PgR positive breast cancer by local laboratory (Protocol Inclusion Criterion 6)
- Patient is female and not post-menopausal (Protocol Inclusion Criterion 4)
- Patient does not have at least one measureable lesion or at least one predominantly lytic bone lesion (Protocol Inclusion Criterion 8)
- Patient is newly diagnosed with endocrine naive advanced breast cancer (Protocol Inclusion Criterion 9)
- Patient relapsed on/or within 12 months from completion of (neo) adjuvant endocrine therapy and then subsequently progressed after one line of endocrine therapy (either an antiestrogen or an aromatase inhibitor) for metastatic disease (Protocol Inclusion Criterion 9)
- Patient has no radiological or objective evidence of recurrence or progression (i.e.: clinical progression not allowed) (Protocol Inclusion Criterion 5)
- Patient has no recurrence or progression of disease during or after aromatase inhibitor therapy (Letrozole, anastrozole, exemestane) (Protocol Inclusion Criterion 10)
- Patient does not have an ECOG status of 0 or 1 (Protocol Exclusion Criterion 11)
- Patient has symptomatic visceral disease or any disease burden that makes the patient ineligible for endocrine therapy per the investigator's best judgment (Protocol Exclusion Criterion 1)
- Patient has received prior treatment with chemotherapy (except for neoadjuvant/ adjuvant chemotherapy), or any PI3K, mTOR or AKT inhibitors (Protocol Exclusion Criterion 2)
- Patient is concurrently using other anti-cancer therapy (Protocol Exclusion Criterion 5)
- Patient is currently receiving any of the following substances and cannot be discontinued 7 days prior to Cycle 1 Day 1: Herbal preparations / medications
- Patient has already been treated by Fulvestrant (Protocol Exclusion Criterion 2)

Other Criteria

- Patient enrolled into the wrong cohort, either:
- Patient with PIK3CA non-mutant status (as per confirmation by Novartis-designated laboratory) enrolled into PIK3CA mutant cohort
- Patient with unknown PIK3CA mutation status (as per confirmation by Novartis-designated laboratory) enrolled into the PIK3CA mutant cohort
- Patient did not take any study treatment

2.3.4 Pharmacokinetic (PK) set

The PK Analysis Set (PAS) will consist of all patients who receive at least one dose of study treatment (alpelisib/placebo or fulvestrant) and have at least one post-treatment concentration measurement.

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 2-1.

Table 2-1 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 study treatment taken
Per Protocol set	Any major protocol deviation as listed in definition of per protocol set	Patient not evaluable for FAS

2.4 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.

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

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 3.2.8 of RECIST Novartis guidelines as described in detail in Appendix 3 of the Clinical Study Protocol).

Patients with at least one predominantly lytic bone lesion but not having measurable disease per RECIST 1.1 are allowed to enter the study and the same rules as specified above apply for these patients.

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.

2.4.2 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.

2.4.3 Best overall response (BOR)

The evaluation of BOR will be assessed using RECIST 1.1. The definitions and the details on the derivation of BOR are given in Appendix 3 of the study protocol.

The best overall response will usually be determined from response assessments undertaken while on treatment. In addition, only tumor assessments performed before the start of any further anti-neoplastic therapies (i.e. any additional secondary anti-neoplastic therapy with the exception of palliative radiotherapy) will be considered in the assessment of BOR.

Further anti-neoplastic therapies will be identified from the data collected on 'Antineoplastic therapy since discontinuation of study treatment- Medication' eCRF and "Antineoplastic Therapy - Radiotherapy" eCRF.

Palliative radiotherapy is the only setting of radiotherapy allowed during a study. It will not be considered as an antineoplastic therapy for the assessment of BOR.

Continuation of fulvestrant monotherapy as 1st new anti-neoplastic therapy after end of treatment without prior PD and collected in the 'antineoplastic therapy since discontinuation of study treatment- Medication' eCRF, will not be considered as an anti-neoplastic therapy for the assessment of BOR.

Since the tumor assessments are performed every 8 weeks (+/- 7 days) after randomization during the first 18 months and every 12 weeks (+/- 7 days) thereafter, the standard definition of a best overall response evaluation of "stable disease", "progressive disease" or "unknown" given in the Appendix 3 of the study protocol requires an adjustment. Best overall response for each patient is determined from the sequence of overall (lesion) responses 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) > 7 weeks after randomization date (and not qualifying for CR or PR).
- Non-CR/non-PD = at least one non-CR/non-PD assessment (or better) > 7 weeks after randomization date (and not qualifying for CR or PR). This applies only for patients with non-measurable disease alone at Baseline.
- PD = progression ≤ 17 weeks after randomization date (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 7 weeks or early progression within the first 17 weeks).

Patients with 'unknown' BOR 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 lesion response UNK
- New anti-neoplastic therapy started before first post-baseline assessment
- SD or non-CR/non-PD too early
- PD too late

Note 1: A SD or Non-CR/Non-PD is considered as "SD too early"' if the SD or Non-CR/Non-PD is documented within first 7 weeks after randomization date.

Note 2: A PD is considered as "PD too late" if the first documentation of PD is recorded more than 17 weeks after randomization date with no qualifying CR, PR or SD or Non-CR/Non-PD in between.

Note 3: Special (and rare) cases where BOR is "unknown" due to both too early SD and too late PD will be classified as "SD too early".

2.4.4 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 change in technique (e.g. from CT to MRI, or vice-versa), or a change in any other imaging modality. A change in methodology will result by default in an UNK (unknown) calculated overall lesion response assessment. However, another response assessment than the Novartis calculated UNK response may be accepted from the investigator or the central blinded reviewer if a definitive response assessment can be justified based on the available information.

2.4.5 Determination of missing adequate assessments

The term 'missing adequate tumor assessment' is defined as a tumor assessment not done or tumor assessment with overall lesion response 'Unknown'. For the sake of simplicity, a 'missing adequate tumor assessment' will also be referred to as a 'missing assessment'.

As described in Table 14-9 in Appendix 3 of the study protocol, the PFS censoring and event date options depend on the presence and the number of missing tumor assessments (TAs). In the analysis of PFS, an event occurring after two or more missing assessments or non-adequate tumor assessments is censored at the last adequate tumor assessment.

An exact rule to determine whether there is no, one or two missing TAs is therefore needed. This rule is based on the time 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 windows for tumor assessments, and the thresholds for LATA to belong to a visit can be found in the following table.

Schedule for tumor assessment and time windows

Assessment schedule		Scheduled date – 1 week	Scheduled date (weeks from randomization)	Scheduled date +1 week	Threshold (weeks)*
Every 8 weeks for the first 18 months	Baseline	0	0^	1	n/a
	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.

^ Day of randomization is taken as 0.

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 the table above (essentially whichever scheduled assessment it is closest to).

Two additional thresholds, D1 and D2 are calculated for that scheduled assessment based on the protocol-specified schedule and windows.

- The threshold D1 is defined as the protocol-specified time interval between the TAs plus 2x the protocol-allowed time window around the assessments.
- The threshold D2 is defined as twice the protocol-specified time interval between the TAs plus 2x the protocol-allowed time window around the assessments (except when the matched scheduled tumor assessment is C17D1, in which case D2 is defined in Rule 2 below).

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

Rule 1: if LATA happens within 60 weeks from randomization (the matched scheduled tumor assessment is C15D1 or before). For example, D1=8+2=10 weeks and 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). For example, D1=8+2=10 weeks and 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). For example, D1=12+2=14 weeks and D2=2*12+2=26 weeks.

The number of missing events is defined as:

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

.

The same definition of D2 will be used to determine the PFS censoring reason. If there is no post-baseline adequate tumor assessment available (before an event or a censoring reason occurred), the randomization date will be used to compute the interval.

If the time interval between the last adequate TA date and the earliest of the following dates is smaller or equal to D2 days:

Analysis cut-off date

Date of consent withdrawal

Visit date of study treatment discontinuation due to lost to follow-up or end of post-treatment follow-up discontinuation due to lost to follow-up.

Then the PFS censoring reason will be respectively:

- ‘Ongoing’
- ‘Withdrew consent’
- ‘Lost to follow-up’

However if the time interval is larger than D2 days with no event then the PFS censoring reason will always default to ‘Adequate assessment no longer available’. If the time interval between the last adequate tumor assessment 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'.

2.4.6 No baseline tumor assessments

As specified in Table 14-9 in Appendix 3 of the 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 'progressive disease' component of the PFS assessment.

Patients without any baseline tumor assessment who die within D2 time interval (Section 2.4.5) from date of randomization will be counted as having an event in the analysis of PFS at the date of death. All deaths will be counted in the overall survival analysis regardless of presence or absence of the baseline tumor assessment.

2.4.7 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 2-2.

Table 2-2 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

Based on the above considerations, the following algorithm will be used to construct the graph:

1. Select "valid" post-baseline assessments to be included, i.e. for each patient and each assessment repeat the following four steps.

1.1. Check the target lesion response and overall lesion response. If at least one of them is UNK then exclude the whole assessment. Otherwise, go to step 1.2.
1.2. Check the overall lesion response. If it is PD then go to step 1.3. Otherwise go to step 1.4.
1.3. Check target response. If it's PD then go to step 1.4. Otherwise flag the assessment with ✶ .
1.4. Calculate the % change from baseline in target lesions.

2. For each patient, go through all valid assessments identified in step 1 and find the assessment with best % change from baseline in target lesions. The "best" means best for the patient, i.e. the largest shrinkage or if a patient only has assessments with tumor growth take the assessment where the growth is minimal.
3. Construct the waterfall graph displaying the best % change from baseline for each patient. Patients having only ✶ flagged assessment(s) will be displayed separately.

The graph will be constructed using the data from the investigator/local radiologist assessments.

The best overall response (BOR) will be shown above each of the displayed bars in the graph, if the number of patients displayed in the graph is small enough for the labels to be legible.

The order of the display from left to right will be as follows:

1. Bars under the horizontal axis representing tumor shrinkage
2. Bars above the horizontal axis representing tumor growth
3. "Zero" bars with ✶ symbol.

For each of the 3 categories above, n (%) (where % uses the total number of patients displayed in the graph) will be displayed. If there are any patients with zero change they will be as a separate category following patients with tumor shrinkage.

3 Statistical methods used in reporting

3.1 Enrollment status

The following summaries will be provided in each cohort separately for the FAS overall, and for both treatment groups:

1. Number (%) of patients who were randomized
2. Number (%) of patients who received at least one dose of study treatment after randomization

Number (%) of patients screened will be summarized by country and center. In addition, the number (%) of patients randomized will be summarized by country, center and treatment group.

For patients who are screen failures, the reasons for not completing screening will be summarized based on "Screening Phase Disposition" eCRF.

3.2 Background and demographic characteristics

The FAS will be used for all baseline disease characteristics and demographic summaries and data listings. Summaries will be presented in each cohort separately for the FAS overall and for both treatment groups to assess baseline comparability. No inferential statistics will be provided.

3.2.1 Basic demographic and background data

Descriptive statistics (mean, median, standard deviation, minimum, maximum) will be presented for continuous variables. The number and percentage of patients in each category will be presented for categorical variables. This analysis will include the following: Age, gender, race, ethnicity, body mass index (BMI) and ECOG performance status at baseline.

BMI at Baseline will be calculated using the following formula, i.e., BMI (kg/m^2) = weight (kg) / (height (cm)/100)**2) using weight at Baseline.

A breakdown of the age distribution according to EUDRACT required categories will be provided: Adolescents (12-17 years), Adults (18-64 years), Elderly (>= 65 years).

3.2.2 Diagnosis and extent of cancer

This analysis will include the following: primary site of cancer, predominant histology/cytology, histological grade, stage at initial diagnosis, stage at time of study entry, time since initial diagnosis of primary site, time from initial diagnosis to first recurrence/progression, time since most recent relapse/progression (where last treatment was in the (neo-)adjuvant or advanced disease settings), presence/absence of target and non-target lesions, number of metastatic sites, metastatic sites, HER2 receptor status, estrogen receptor (ER) status, progesterone receptor (PgR) status and hormone receptor (ER and/or PgR) status.

Note: The variables 'presence/absence of target and non-target lesions' will be based on the data collected on target/non-target lesion assessment according to RECIST 1.1 (Appendix 3 of the protocol) and documented in the eCRF.

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 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 (%) of patients receiving prior anti-neoplastic medication, prior anti-neoplastic radiotherapy and prior anti-neoplastic surgery respectively will be summarized.

Prior anti-neoplastic therapy will be summarized for each treatment group by therapy type (surgery, radiotherapy, chemotherapy, hormonal therapy etc.), disease setting, last therapy

before study entry and its outcome (type of therapy, setting, and best response). Number of lines of prior medication therapy, prior chemotherapy, and prior hormonal therapy will be summarized by treatment group in any setting and in the metastatic setting. In addition, number of lines of prior chemotherapy in the neo-/adjuvant setting will also be summarized by treatment group.

In addition, number (%) of patients who received any hormonal therapy (irrespective of given in combination with chemotherapy or biologic or targeted therapy), any aromatase inhibitors, any hormonal therapy other than aromatase inhibitors in any setting and in the metastatic setting, any prior CDK4/6 inhibitors will be summarized by treatment group.

Patients exhibiting primary & secondary endocrine resistance will be summarized based on the ESMO (ref) definition, vs patients exhibiting endocrine sensitivity:

- **Primary resistance** = Relapse < 24 months while on ET in adjuvant setting or progression < 6 months while on ET in metastatic setting
- **Secondary resistance** = Relapse ≥ 24 months while on ET in adjuvant setting or relapse <12 months after end of ET in adjuvant setting or progression ≥ 6 months while on ET in metastatic setting
- **Endocrine sensitive** = Relapse ⩾ 12 months after end of ET in adjuvant setting or progression ⩾ 12 months after end of ET in metastatic setting

The patient population will be summarized additionally by presenting the number (%) of patients with:

- relapsed with documented evidence of progression more than 12 months from completion of (neo)adjuvant endocrine therapy with no treatment for metastatic disease
- relapsed with documented evidence of progression more than 12 months from completion of (neo)adjuvant endocrine therapy and then subsequently progressed with documented evidence of progression while on or after only one line of endocrine therapy for metastatic disease
- relapsed with documented evidence of progression while on (neo) adjuvant endocrine therapy or within 12 months from completion of (neo)adjuvant endocrine therapy with no treatment for metastatic disease
- newly diagnosed advanced breast cancer, then relapsed with documented evidence of progression while on or after only one line of endocrine therapy

The medication therapy type of any combination therapy will be classified based on the following order: chemotherapy, biologic therapy, targeted therapy, hormonal therapy. For example, a combination therapy of chemotherapy and hormonal therapy will be classified as chemotherapy.

Last hormonal therapy prior to study entry will be summarized by type of aromatase inhibitors, anti-estrogen therapy, disease setting, duration of hormonal therapy in adjuvant setting and best overall response at last hormonal therapy in the metastatic setting.

Last hormonal therapy refers to hormonal medication received in the last regimen. If multiple hormonal therapy medications are given as part of last regimen, then last hormonal therapy medications are selected based on the latest start date among the hormonal therapy medications. Standard imputation rules will be applied for dates of prior anti-neoplastic therapy (medication, radiotherapy and surgery).

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

[redacted]

3.3 Protocol deviation summaries

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

Protocol deviations leading to the exclusion from the Per Protocol Set will also be summarized.

All protocol deviations will be listed. Summaries will be presented in each cohort separately for the FAS overall and for both treatment groups.

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 in each cohort separately, by treatment arm and randomization stratum (presence of lung and/or liver metastases, previous treatment with any CDK4/6 inhibitor based on data obtained from the IRT system).

Discrepancies between stratum recorded in IRT at the time of randomization and actual stratum recorded in the clinical database will be summarized.

3.5 Patient disposition

The FAS will be used for the patient disposition summaries, which will be summarized in each cohort separately overall, and for both treatment groups.

Based on the ‘End of Treatment Disposition’ and ‘End Post Trt Phase Disposition’ CRF there will be one combined by-treatment summary showing:

1. Number (%) of patients who are still on-treatment (based on the absence of the ‘End of Treatment Disposition’ eCRF)
2. Number (%) of patients who discontinued study treatment (based on the ‘End of Treatment Disposition’ eCRF)
3. Reasons for study treatment discontinuation (based on ‘End of Treatment Disposition’ eCRF)

4. Number (%) of patients who entered the post-treatment evaluations (based on 'End of Treatment Disposition' eCRF)
5. Number (%) of patients who discontinued from the post-treatment evaluations (based on the 'End Post Trt Phase Disposition' eCRF)
6. Reasons for discontinuation from the post-treatment evaluations phase (based on 'End of post treatment follow up disposition' eCRF).

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.

Summaries will be presented for the Safety set by treatment arm, in each cohort separately as well as both cohorts combined.

Data handling

The following rule should be used for the imputation of date of last administration (please refer to Section 2.1.3) for a given study treatment component:

Scenario 1: If the date of last administration is completely missing and there is no EOT eCRF, the subject is considered as on-going:

The subject should be treated as on-going and the cut-off date should be used as the last dosing date.

Scenario 2: If the date of last administration is completely or partially missing and the EOT eCRF is available (prior to any death date or withdrawal of consent date, if available):

Case 1: The date of last administration is completely missing, and the EOT visit date is complete, then this latter date should be used.

Case 2: Only Year(yyyy) of the dose end date is available and yyyy < the year of EOT date:

Use Dec31yyyy

Case 3: Only Year(yyyy) of the dose end date is available and yyyy = the year of EOT date:

Use EOT date

Case 4: Both Year(yyyy) and Month (mm) are available for the date of last administration, and yyyy = the year of EOT date and mm < the month of EOT visit:

Use last day of the Month (mm).

After imputation, compare the imputed date with the start date of that specific record, if the imputed date is < start date of that record

Use the start date of that record.

Subjects with missing start dates are to be considered missing for all study treatment component related calculations described in Section 3.6 and no imputation will be made. If the date of first administration is missing, then the date of last administration should not be imputed.

3.6.1 Duration of study treatment exposure

Duration of exposure to study drug (for alpelisib/placebo and fulvestrant) is defined according to dosing regimen for each study drug as outlined in Section 2.1.6.

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

Duration of exposure to study treatment is considered by taking into account the duration of exposure to each study drug:

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

The duration includes the periods of temporary interruption. 'Date of first administration of study drug/treatment' and 'last date of exposure to study drug/treatment' are defined in Sections 2.1.2/2.1.4 and 2.1.6 respectively.

Duration of exposure to study drug/treatment will be categorized into time intervals (<1 month, at least1 month, at least 2 months etc.). In addition, summary statistics will be displayed.

Note: If the last record in DAR CRF is a zero dose, this record will not be used in the analyses.

To assess the impact of specific adverse events of special interest on exposure to study treatment the following summaries are added:

- Duration of exposure to study drug for patients who developed hyperglycemia during the study
- Duration of exposure to study drug for patients who developed hyperglycemia during the study and started anti-diabetic medication
- Duration of exposure to study drug for patients who developed hyperglycemia during the study and started anti-diabetic medication by hyperglycemia diagnosis status
- Duration of exposure to study drug for patients who developed rash during the study
- Duration of exposure to study drug for patients who developed rash during the study and started anti-rash medication

Patients who developed hyperglycemia [rash] during the study are defined as those patients with any on-treatment CTC Grade 1 or higher AESI. Hyperglycemia diagnosis status is defined in Section 3.13.1. Anti-diabetic medications are defined by medications where ATC code=A10 [Drugs used in diabetes]. Anti-rash medications are defined by medications where ATC codes in the following categories:

- R06A [Antihistamines for systemic use]
- D07 [Topical dermatological corticosteroids]
- H02 [Corticosteroids systemic]

3.6.2 Cumulative dose

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

Cumulative dose will be summarized using descriptive statistics by treatment arm for each component of study treatment. For patients who do not receive any drug the cumulative dose will be set to zero.

The cumulative dose is defined according to the type of dosing schedule and is calculated from the DAR eCRF. It is expressed in mg for alpelisib/placebo and fulvestrant.

Alpelisib/placebo

Cumulative dose (mg) = Sum of doses of the study drug administered to the patient from the start date to the last date of study drug.

Fulvestrant

The cumulative dose for fulvestrant with cyclic administration should be defined based on the days when the subject is assumed to have taken a non-zero dose during dosing periods.

3.6.3 Dose intensity and relative dose intensity

Dose intensity (DI) for patients with non-zero duration of exposure is defined as follows.

DI (dosing unit / unit of time) = Cumulative dose (dosing unit) / Duration of exposure (unit of time).

For patients who did not take any drug the DI is equal to zero. Planned dose intensity (PDI) is the assigned dose by unit of time planned to be given to patients as per protocol in the same dose unit and unit of time as that of the Dose Intensity. DI, PDI and Relative dose intensity (RDI) is defined as:

For alpelisib/placebo:

- DI (mg/day) = Cumulative dose (mg) / duration of exposure (days)
- PDI is 300 mg/day
- RDI (%) = DI (mg/day) / PDI (mg/day) *100

For fulvestrant:

- DI (mg/day) = Cumulative dose (mg) / duration of exposure (days)

Categorical summaries of alpelisib RDI and the continuous summaries of RDI (i.e. mean, standard deviation etc.) will be presented.

3.6.4 Dose reduction, interruption and permanent discontinuation

The number (%) of patients with dose reductions or interruptions and permanent discontinuations, and associated reasons, will be summarized separately for each study drug (alpelisib/placebo and fulvestrant). In addition, reasons for permanent discontinuation from the study drug will be summarized for both alpelisib/placebo and fulvestrant.

Dose administered (mg) and dosing frequency from the DAR eCRF will be used to determine the dose reductions and interruptions.

'Dose permanently discontinued' ticked box from the DAR eCRF will be used to determine permanent discontinuation.

Dose interruption

For alpelisib/placebo, an interruption is defined as a zero dose on one or more days between two non-zero doses.

Any two or more consecutive zero doses of alpelisib (e.g. in the sequence 300 mg daily, 0 mg, 0 mg, 300 mg daily) or fulvestrant will be counted as 1 interruption if the reasons for these two consecutive dose interruption are the same. It will be counted as two different interruptions only if the reasons are different.

For fulvestrant, an interruption is defined as a zero dose on one or more days between two non-zero doses.

For example, a 500mg dose on C1D1, a zero dose on C1D15 and a 500mg dose on C2D1 will constitute a single interruption.

The number (%) of dose interruptions along with reasons will be summarized.

Note: The last zero dose of alpelisib/placebo or fulvestrant is not considered as a dose interruption.

Dose reduction

For alpelisib/placebo, a dose reduction is defined as a decrease in dose from the protocol planned starting dose (e.g. from 300 mg daily to 250 mg daily) even if the dose decrease has been directly preceded by an interruption. On the other hand, if the dose decrease is followed by an interruption, with the dose resuming at the same level prior to the interruption (e.g. in the sequence 300 mg daily – 0 mg - 300 mg daily), the second dose decrease or change in dosing frequency will not be counted as dose reduction.

If, due to a dosing error, a patient receives a higher than planned starting dose and moves down to the planned starting dose then this is not considered a dose reduction. However if the dose change is from a higher than planned starting dose down to a lower than protocol planned starting dose, then this is considered a dose reduction (e.g. in the sequence: 350 mg daily, 300 mg daily, 250 mg daily; 250 mg is considered a dose reduction).

If, due to a dosing error, a patient receives a lower than previous non-zero dose and resumes later at the protocol specified dose reduction, then the lower dose received due to dosing error and protocol specified dose reduction are dose reductions (e.g. in the sequence 300 mg daily – 200 mg daily - 250 mg daily, then 200 mg and 250 mg are considered dose reductions).

If, due to a dosing error, a patient receives a lower than previous non-zero dose and resumes later at a lower than previous non-zero dose, then 2 dose reductions will be counted (e.g. in the sequence 300 mg daily - 250 mg daily – 200 mg daily, 250 mg and 200 mg are dose reductions).

Table 3-1 Examples of Dose Reduction for alpelisib

Sequence	Reduction
With dose change	
300 mg daily - 250 mg daily – 0 mg - 250 mg daily	1 reduction (the 1st 250 mg)
300 mg daily – 300 mg daily – 0 mg - 250 mg daily	1 reduction (250 mg)
300 mg daily – 0 mg – 250 mg daily	1 reduction (250 mg)
With interruption	
300 mg daily – 0 mg - 300 mg daily	0 reductions
With dosing error	
300 mg daily – 250 mg daily – 200 mg daily*	2 reductions (250 mg, 200 mg)
300 mg daily – 200 mg daily* - 300 mg daily	1 reduction (200 mg)
300 mg daily – 200 mg daily* - 250 mg daily	2 reductions (200 mg, 250 mg)
300 mg daily – 400 mg daily* - 350 mg daily*	0 reductions since 400 mg and 350 mg are dose escalations not reduction
300 mg daily – 150 mg daily* - 300 mg daily	1 reduction (150 mg)
With dosing error at the 1st administration	
150 mg daily* - 300 mg daily	1 reduction (150 mg)
150 mg daily* - 0 mg – 150 mg*- 300 mg daily	1 reduction (150 mg)
150 mg daily* - 300 mg daily – 0 mg - 250 mg daily	2 reductions (150 mg and 250 mg)

*dosing error

There is no planned dose reduction for fulvestrant; in addition the reason for fulvestrant dose reduction is not collected in the eCRF. No analysis for fulvestrant will be done on the number of, and reasons for, reductions.

3.7 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.

Concomitant therapy include medications (other than study drugs) starting on or after the start date of study treatment or medications starting prior to the start date of study treatment and continuing after the start date of study treatment.

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 and summarized by lowest ATC class and preferred term

using frequency counts and percentages. Surgical and medical procedures will be coded using MedDRA and summarized by SOC and preferred term.

Concomitant medications, concomitant medications with specific impact on the analysis (See Section 3.7.1), procedures and significant non-drug therapies taken concurrently with study treatment will be listed and summarized by ATC class/SOC using frequency counts and percentages. Any prior medications, procedures or significant non-drug therapies starting and ending prior to the start date of study treatment will be listed. Analysis will be based on Safety set.

3.7.1 Concomitant medications with specific impact on the analysis

According to study protocol Table 14-1, the following medications are either prohibited during the treatment period in this study, or to be used with caution:

- CYP3A4, CYP2C8, CYP2C9 or CYP2C19 substrates
- Medications with a known risk of QT prolongation
- QT prolonging drugs to be used with caution
- BCRP inhibitors to be used with caution

A corresponding list for programming purposes will be saved in a separate document.

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.

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 central radiology review committee 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.

Source 2 will be listed against Source 1 and discrepancies between calculated and assigned responses will be identified for data review purposes.

Data from Source 3 will be used for selected supportive efficacy analyses. For central radiology, the final adjudicated data will be listed. Differences in overall responses between local radiology (Source 1) and central radiology (Source 3) will be listed.

3.8.1 Primary efficacy

The primary objective of the study is to determine whether treatment with alpelisib in combination with fulvestrant prolongs PFS compared to treatment with placebo in combination with fulvestrant in men and postmenopausal women with HR+, HER2-negative advanced breast cancer which progressed on or after AI treatment for patients with PIK3CA mutant status as measured in tissue.

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 Disposition' and 'End Post Trt Phase Disposition' eCRF) without supporting objective evidence (as defined in Section 2.4.2) 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) up until the cut-off date defined in Section 2.2. The analysis will be performed on the FAS and will use the default censoring and event date options from Table 3-3, with the exception of the rules for new antineoplastic therapy given, 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). 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 2.4.5). 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 3-3 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 Options for event dates used in PFS, duration of response

Situation		**Options for end-date (progression or censoring)[1] (1) = default unless specified differently in the protocol or RAP**	**Outcome**
A	No baseline assessment	(1) Date of randomization/start of treatment[3]	Censored
B	Progression at or before next scheduled assessment	(1) Date of progression (2) Date of next scheduled assessment[2]	Progressed Progressed
C1	Progression or death after **exactly one** missing assessment	(1) Date of progression (or death) (2) Date of next scheduled assessment[2]	Progressed Progressed
C2	Progression or death after **two or more** missing assessments	(1) Date of last adequate assessment[2] (2) Date of next scheduled assessment[2] (3) Date of progression (or death)	Censored Progressed Progressed
D	No progression	(1) Date of last adequate assessment	Censored
E	Treatment discontinuation due to 'Disease progression' without documented progression, i.e. clinical progression based on investigator claim	(1) Ignore clinical progression and follow situations above (2) Date of discontinuation (visit date at which clinical progression was determined)	As per above situations Progressed
F	New anticancer therapy given	(1) Ignore the new anticancer therapy and follow situations above (ITT approach) (2) Date of last adequate assessment prior to new anticancer therapy (3) Date of secondary anti-cancer therapy (4) Date of secondary anti-cancer therapy	As per above situations Censored Censored Event
G	Deaths due to reason other than deterioration of 'Study indication'	(1) Date of last adequate assessment	Censored (only TTP and duration of response)

[1].=Definitions can be found in Protocol Appendix 3 Section 14.3.25
[2].=After the last adequate tumor assessment. "Date of next scheduled assessment" is defined in Protocol Appendix 3 Section 14.3.25.
[3].=The rare exception to this is if the patient dies no later than the time of the second scheduled assessment as defined in the protocol in which case this is a PFS event at the date of death

3.8.1.2 Hypothesis and test statistic

The overall Type I error for the trial is one-sided 2.5%. The primary efficacy analysis of PFS based on the population of patients with PIK3CA mutant status will be performed at a one-sided 2.0% level of significance. A secondary efficacy analysis of PFS based on the population of patients with PIK3CA non-mutant status will be performed at a one-sided 0.5% level of significance (see Section 3.8.2.2). This approach guarantees the protection of the overall type I error at 2.5% (based on a Bonferroni adjustment).

The primary efficacy analysis will be the comparison of PFS between the two treatment arms using a stratified log-rank test at a one-sided 2.0% level of significance for the PIK3CA mutant

cohort. The strata information will be based on the data obtained from the IRT system that was utilized for randomization.

Assuming proportional hazards model for PFS for the PIK3CA mutant cohort, the following statistical hypotheses will be tested at the one-sided 2.0% level of significance:

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

where θ_1 is the log-hazard ratio (alpelisib + fulvestrant treatment arm vs. placebo+fulvestrant treatment 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 6, 12, 18 months, and the associated 95% confidence intervals will be summarized by treatment arm. Kaplan-Meier estimates will be obtained using PROC LIFETEST with method=KM option in SAS. The loglog option available within PROC LIFETEST will be used to compute the confidence intervals.

The purpose of the two-sided 95% CI is to give an estimate of the respective treatment effect together with a comparable measure of reliability but not for reconstructing the test decisions.

3.8.1.4 Hazard Ratio

The PFS hazard ratio with two-sided 95% confidence interval will be derived from the stratified Cox proportional hazards model for each cohort. 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 analyses of the primary endpoint

Sensitivity analyses are performed only if the analysis of primary endpoint in the PIK3CA mutant cohort shows statistically significant results. Subgroup analyses to explore the intrinsic consistency of any treatment effect found overall are specified in Section 3.13.2.

Per protocol Population

The PPS will be used to perform sensitivity analysis for the primary efficacy endpoint (i.e. PFS) if the primary endpoint is statistically significant.

Different Censoring Mechanisms

Depending on the statistical significance of the primary efficacy endpoint, the following sensitivity analyses will be performed to address the impact of tumor assessment features and censoring rules on primary analyses. The primary efficacy analyses in each cohort; 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. If more than 10% patients have two or more consecutive missing assessments prior to PFS event, primary PFS analyses will be repeated 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 3-3 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'
2. If more than 10% patients have two or more consecutive missing assessments prior to PFS event, primary PFS analyses will be repeated 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 3-3 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' where, the date of next scheduled assessment is defined as the date of the last adequate tumor assessment plus the protocol specified time interval for assessments.
3. Using investigator/local assessment (Source 1 in Table 3-2) on the FAS and censoring all PFS events that occur after the start of new anti-neoplastic therapy to the time of last adequate tumor assessment done prior to the therapy start date.

Unstratified Analysis

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

Baseline Demographic and Disease Characteristic factors

If the primary endpoint is statistically significant, a multivariate stratified Cox regression model for PFS will be fitted to evaluate additionally the effect of other baseline demographic or disease characteristics on the estimated hazard ratio. This model will include the following key prognostic factors: ECOG performance status (0 vs. 1), bone lesions only at baseline (yes or no), number of prior lines of therapy in any setting (1st line vs 2nd line) and region (see Section 3.13.2 for definitions).

All covariates will be included in the model regardless of their observed significance (p-value for given covariate). Forest plots for these factors will be provided, please see Section 3.13.2.

3.8.1.6 Supportive analyses of the primary endpoint

The following analyses will be conducted in the PIK3CA mutant cohort to support the interpretation of the primary endpoint, regardless of the outcome in the primary endpoint.

- If there is a high rate of discrepancy (>10%) between the strata classifications constructed using CRF data and those obtained from the IRT (considering also PIK3CA mutation status), 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.

- 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 2.4.5.

The following summaries on censoring reasons will be produced for PFS by investigator radiology and central radiology. The censoring patterns will be compared between investigator and central review (in patients selected for BIRC assessment).

- Comparison of PFS event type/censor between local radiology review and central radiology review
- Summary for the difference in days to progression as per local radiology review and as per central radiology review

PFS assessed by Blinded Independent Review Committee (BIRC) will serve as supportive evidence of the primary endpoint in the PIK3CA mutant cohort.

For studies with PFS based on local radiology assessment as the primary endpoint, PFS assessment done 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 50% 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 audit BIRC sample will be estimated using the Kaplan-Meier method for the FAS. The median along with two-sided 95% confidence intervals (CI) will be presented by treatment group. 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.

Two additional methods will be used to summarize the data from the BIRC assessment in order to decide whether a 100% BIRC review should be conducted.

- 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 assessment from all patients in the PIK3CA mutant cohort and the BIRC assessment of these patients randomly selected for central review (see Section 4.8 for methodological details). This estimate and its one-sided 95% CI will be provided. The NCI method will be used for audit sample size determination (see Section 3.14) and summary of treatment effect (HR, 95% confidence intervals) based on the supportive BIRC assessment.
- 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 in 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. If the distribution of discrepancies is similar between the arms this suggests the absence of evaluation bias favoring a particular arm. (see Section 4.8 for details on the calculations). 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 alpelisib+fulvestrant arm minus the rate on the placebo+fulvestrant arm. The EDR and LDR results will also be summarized by treatment arm.

If the analysis of primary endpoint in the PIK3CA mutant cohort shows a statistically significant treatment effect, a full BIRC review may be conducted.

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 negative observed differential discordance for the EDR or a positive differential discordance for the LDR)

Cross-tabulation of 'PFS by central radiology' vs. 'PFS by investigator' by PFS event type (i.e. 'death', 'PD', 'censor' for each of the two sources) and by treatment will be constructed to investigate discordance between the two sources (in patients selected for BIRC assessment). The discrepancy rate between central radiology and investigator will be calculated and presented as % as follows: $100 \times (n_{13} + n_{23} + n_{31} + n_{32})/ N$ by treatment arm.

A cross-tabulation will be produced displaying the PFS timings for the local investigators' assessment compared to the BIRC assessment (in patients selected for BIRC assessment). For progression assessments, the frequency and percent of subjects with complete agreement [occurring on the same date plus or minus 7 days of each other], progression later, progression earlier, and cases where progression was called by one method and censored by the other will be displayed. Similarly, if censoring was recorded, the frequency and percent of subjects with complete agreement, censoring called later, censoring called earlier, and cases where censoring was called by one method and progression was called by the other method will be displayed.

3.8.1.6.1 Missing tumor assessments

The number of patients with at least one missing/unknown TA based on local assessment will be presented together with the following breakdown categories: number of patients with 1, 2, 3, 4, 5, >5 missing/unknown TAs. The purpose of this analysis is to gain an insight as to whether the TAs have been carried out in accordance with the protocol and to understand if any meaningful discrepancies exist between the pattern of missing assessments by treatment arms.

Since the planned tumor assessments are every 8 weeks in the first 18 months and every 12 weeks thereafter, the following time windows (in weeks) will be constructed for each patient: (note the open parenthesis such as (12, 20] indicates that week 12 doesn't belong to this interval and week 12+1 day belongs to that interval)

- Until ~18 months post randomization [0, 12], (12, 20], (20, 28], …(60, 68], (68, 76]
- After ~18 months post randomization (76, 90], (90, 102], (102, 114], …

where '0' is the patient's date of randomization. Every time-window (with the exception of the initial, broader one) is centered at the scheduled time of a TA, i.e., around week 16, week 24 for second and third window respectively, etc. A patient will be considered 'at risk' of missing his/her TA for any one of these time-windows if he/she either:

- is 'on study' for at least the first 4 weeks of the time-window for the first 18 months (8 weeks for the first time window), or at least the first 6 weeks of time-window thereafter, i.e., if the patient is ongoing at the time of the scheduled TA, or
- discontinued treatment due to documented disease progression within the specific time window.

For example, if a patient discontinued due to documented disease progression during Week 24, then he/she would have been 'at risk' of a missing/unknown TA for the [20, 28] week time-window.

For the purpose of this analysis, 'unknown' TAs (i.e., evaluations with an overall lesion response of 'unknown') will be considered to be missing. However, a clear distinction between 'truly missing' and 'present but unknown' needs to be made in the derived dataset to allow for both a combined analysis, i.e. missing and unknown treated the same, and separate analyses.

TAs performed after a documented disease progression will not be considered. In other words, the final time-window for which a patient would be at risk of a missing/unknown scan would be that during which the documented progression occurred.

For patients without documented progression, all TAs are considered up to the earliest of the following dates: death, the analysis cut-off, disease progression, withdrawal of consent or loss to follow-up.

[redacted]

3.8.2 Secondary efficacy analyses

The analysis of all secondary efficacy endpoints will be performed, where appropriate, based on the FAS for i) cohort of patients with PIK3CA mutant status, and ii) cohort of patients with PIK3CA non-mutant status.

3.8.2.1 Key secondary objective: overall survival

The key secondary objective of the study are to determine whether treatment with alpelisib in combination with fulvestrant prolongs OS compared to treatment with placebo in the PIK3CA mutant cohort.

In the PIK3CA mutant cohort, a hierarchical testing procedure will be adopted and OS will be tested between the two treatment groups, provided the primary endpoint PFS is statistically significant favouring alpelisib.

OS is defined as the time from date of randomization to date of death due to any cause. 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 2.1.10).

Assuming proportional hazards model for OS, the following statistical hypothesis for OS will be tested using a stratified log-rank test (according to randomization stratification factors) at the one-sided level of significance of 2.0%:

$$H_{02}: \theta_2 \geq 0 \text{ vs. } H_{a2}: \theta_2 < 0$$

where θ_2 is the log-hazard ratio (alpelisib-fulvestrant treatment arm vs. placebo-fulvestrant treatment arm) of OS.

The analysis for OS will be based on the FAS population according to the treatment arm patients were randomized to and the strata they were assigned to at randomization.

The final OS analysis will not be performed at the time point of the final PFS analysis in the PIK3CA mutant cohort, but after additional follow-up. Therefore, a three-look group sequential design is considered for OS, see Section 3.12.2 for further details.

The type I error probability will be controlled by using a separate Lan-DeMets (O'Brien-Fleming) alpha spending function independent of the Haybittle-Peto boundary used for the primary efficacy analysis of PFS at a 2.0% level of significance. This guarantees the protection of the overall type I error ($\alpha = 2.5\%$) across all hypotheses and the repeated testing of the OS hypotheses at the interim and the final analyses (Glimm 2010). This includes hypotheses associated with the secondary endpoints PFS and OS in the *PIK3CA* non-mutant cohort (PFS in the non-mutant cohort will be tested at a 0.5% level of significance if PoC is established).

The OS distribution will be estimated using the Kaplan-Meier method, and Kaplan-Meier curves, medians and 95% confidence intervals of the medians along with the proportion of patients alive at 12, 24, 36, and 48 months will be presented for each treatment group. The hazard ratio for OS will be calculated, along with its 95% confidence interval, using a stratified Cox model using the same stratification factors as the log-rank test. The purpose of the two-sided 95% CI is to give an estimate of the respective treatment effect together with a comparable measure of reliability but not for reconstructing the test decisions.

Supportive Analyses

If the key secondary endpoint OS is statistically significant, a multivariate stratified Cox regression model will be fitted to evaluate the effect of other baseline demographic or disease characteristics on the estimated hazard ratio. This model will include the following key prognostic factors: ECOG performance status (0 vs. 1), bone lesions only at baseline (yes or no), number of prior lines of therapy in any setting (1st line vs 2nd line) and region (see Section 3.13.2 for definitions).

All covariates will be included in the model regardless of their observed significance (p-value for given covariate).

If the analysis for OS is statistically significant, subgroup analyses of OS will be performed to explore homogeneity of the treatment effect across relevant patient subsets. See details in Section 3.14.2.

The pattern of censored data will be examined between the treatment arms: reasons for censoring ('Alive' or 'Lost to follow-up') and death cause will be summarized by treatment arm. 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 either side).

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 (any valid date from data base used for deriving last contact date +1, cutoff date)].
- If both day and month are missing, then impute max [(1 Jan-yyyy, min (any valid date from data base used for deriving last contact date +1, cutoff date)].

3.8.2.2 PFS in patients with *PIK3CA* non-mutant status measured in tissue

PFS in the *PIK3CA* non-mutant cohort will be analyzed at a single look based on the FAS population according to the treatment arm patients were randomized to and the strata they were assigned to at randomization. Refer to Section 2-2 for the cut-off date for efficacy.

PFS treatment effect in this cohort will be considered to be clinically relevant via a Bayesian decision rule if:

- The estimated HR (stratified according to presence of lung and/or liver metastasis and previous treatment with CDK4/6 inhibitor) ≤ 0.60

and

- The posterior probability (HR < 1) $\geq 90\%$

The posterior probability in the second criterion will be derived from the Bayesian posterior distribution of the HR. Assuming a non-informative prior distribution, the distribution of the HR will be updated with all available data from the patients included in the FAS in this cohort.

The cumulative posterior distribution will be used to derive the probability that the true HR is less than 1.

If both these criteria are met then the comparison of PFS between the two treatment arms in this cohort using a stratified log-rank test at a one-sided 0.5% level of significance, will be made. Assuming proportional hazards model for PFS, the following statistical hypotheses will be tested at the one-sided 0.5% level of significance:

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

where θ_1 is the log-hazard ratio (alpelisib + fulvestrant treatment arm vs. placebo+fulvestrant treatment arm) of PFS.

The median PFS along with 95% confidence intervals will be presented by treatment arm.

3.8.2.3 OS in patients with *PIK3CA* non-mutant status measured in tissue

OS analyses will be performed only if the secondary efficacy endpoint, PFS, in this cohort meets the PoC criteria given in Section 3.8.2.2 and is statistically significant. A hierarchical testing procedure will be adopted. Assuming proportional hazards model for OS, the following statistical hypotheses will be tested at the one-sided 0.5% level of significance:

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

where θ_1 is the log-hazard ratio (alpelisib + fulvestrant treatment arm vs. placebo+fulvestrant treatment arm) of OS.

The analysis for OS will be based on the FAS population according to the treatment arm patients were randomized to and the strata they were assigned to at randomization.

The final OS analysis will not be performed at the time point of the final PFS analysis in the *PIK3CA* non-mutant cohort, but after additional follow-up. Therefore, a three-look group sequential design is considered for OS, see Section 3.12.3 for further details.

OS will be estimated using the Kaplan-Meier method. The median OS along with 95% confidence intervals will be presented by treatment arm. Stratified Cox regression will be used to estimate the HR of OS, along with 95% confidence interval. The pattern of censored data will be examined between the treatment arms as described in Section 3.8.2.1.

3.8.2.4 PFS in patients where PIK3CA mutation status is measured in ctDNA

An analysis of PFS based on local radiology assessments and using RECIST 1.1 criteria for each of (i) patients with *PIK3CA* mutant status and (ii) patients with *PIK3CA* non-mutant status as measured in ctDNA at baseline will be conducted using the same analytical conventions as the primary analysis.

3.8.2.5 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 for each cohort. 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 along with the exact binomial two-sided 95% confidence interval (Clopper 1934) by treatment arm for the *PIK3CA* mutant and non-mutant cohorts.

In addition, ORR in the subset of patient with measurable disease at baseline will be presented.

As a supportive analysis, ORR will also be summarized based on the central radiology review of tumor data, if the BICR moves to a 100% audit.

3.8.2.6 Clinical benefit rate

Clinical benefit rate is defined as the proportion of patients with a best overall response of CR or PR or SD or Non-CR/Non-PD lasting 24 weeks or more based on local investigator assessment 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 presented along with the exact binomial two-sided 95% confidence interval (Clopper 1934) by treatment arm for the *PIK3CA* mutant and non-mutant cohorts.

In addition, CBR in the subset of patient with measurable disease at baseline will be presented.

As a supportive analysis, CBR will also be summarized based on the central radiology review of tumor data, if the BICR moves to a 100% audit.

3.8.2.7 Clinical response in patients with *PIK3CA* mutant status measured in ctDNA

An analysis of ORR, CBR based on local radiology assessments and using RECIST 1.1 criteria; and OS for (i) patients with *PIK3CA* mutant status and (ii) patients with *PIK3CA* non-mutant status as measured in ctDNA at baseline will be conducted using the same analytical conventions as the endpoint where mutation status is defined in tissue.

3.8.2.8 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

Score	Description
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 corresponding 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 ECOG PS assessment is not available, use the one performed at screening)
Cycle 2 Day 1	29	Day 2 to day 42
Cycle k Day 1 (k≥3)	d=(k-1)*28+1	Day d-14 to day d+13
End of Treatment		Assessment taken at the end of treatment visit

If 2 assessments within a time window are equidistant from the target date (or if the closest assessment to the target date has two ECOG filled out on the same date), then the worst ECOG PS value will be used.

Time windows are applicable for descriptive summary of ECOG data by visit only. For time to deterioration analysis described hereafter all post-baseline assessments will be considered.

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

Time to definitive deterioration of the ECOG PS is the number of days between the date of randomization and the date of the assessment at which definitive deterioration is seen. The ECOG PS deterioration is considered definitive if there is an increase in the performance status by at least one category relative to the baseline or death due to any cause and if no improvement in ECOG PS is observed subsequent to the deterioration.

Baseline is the last available assessment on or before date of randomization. If a patient has 2 ECOG PS values at the same date, the worst ECOG PS value will be taken as ‘baseline’.

Example: If the ECOG PS is 1 at baseline and then 1, 2, 1, 2, 3 at D28, D57, D83, D115, and D150 respectively, then the time to definitive worsening is D115.

Example: if the ECOG PS is 1 at baseline and then 1, 1, 2 at D28, D57, and D83 respectively, with no assessment of the ECOG PS after D83 then the time to definitive worsening is 83 days.

If a definitive deterioration is observed after any missing assessments, this event will be backdated to the first of the missing assessments before the deterioration. The first missing

assessment date is calculated as the last available assessment before the definitive deterioration plus X, where X corresponds to the planned scheduled time point for ECOG PS (28 days).

For example, if a patient has an assessment at week 6, misses the following two assessments on weeks 10 and 14 and a definitive deterioration is observed on week 15, then the event will be backdated to week 6+28 days.

In addition, death is considered as a worsening of performance status if it occurs close to the last available assessment, where "close" is defined as twice the planned (i.e. protocol scheduled) period between two assessments. This avoids overestimating the time to definitive worsening in patients dying after an irregular assessment scheme. Patients who die after more than twice the planned period between two assessments are censored at the date of their last available assessment of the performance status.

For example, if the last assessment is at week 6 and the patient dies at week 10, the definitive deterioration date will be week 10. On the other hand, if the last assessment is at week 6 and the patient dies at week 16, which is after more than twice the planned period between two assessments since the last assessment (week 6), then the definitive deterioration date will be week 6.

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

Patients without baseline ECOG PS or without any post-baseline ECOG PS will be censored at the date of randomization with censoring reason being 'No baseline score' or 'No post-baseline score', respectively. However, patients without post-baseline ECOG PS who die within 62 days after date of randomization will be counted as having a definitive deterioration of the ECOG PS at the date of death.

This threshold corresponds to twice the protocol defined ECOG assessment interval plus twice the time window around each assessment: (2 x 28 days)+(2 x 3 days), i.e. 62 days.

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

3.8.2.9 Patient reported outcomes

The European Organisation for Research and Treatment of Cancer's core quality of life questionnaire (EORTC-QLQ-C30, version 3.0) [redacted] will be used to evaluate patient-reported outcome measures of health-related quality-of-life, [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 (including at EOT) until disease progression, death, withdrawal of consent, loss to follow-up or subject/guardian decision.

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

Assessment	Time Interval
Baseline	Screening assessment
Cycle 3, 5, 7, 9 until cycle 19	+/- 4 weeks centered around the planned assessment date (except for the first window and the last window): i.e. days (1, 85] for Day 1 of cycle 3 (2nd assessment) days (85, 141] for Day 1 of cycle 5 (3rd assessment) days (k*56-27; k*56+29] for (k+1th assessment) days (477, 547] for 10th assessment on Day 1 cycle 19
Cycle 22, 25, 28, …	+/- 6 weeks centered around the planned assessment date: i.e. days (547, 631] for 11th assessment days (631, 715] for 12th assessment days (715, 799] for 13th assessment

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 global health status/QoL scale score of the EORTC QLQ-C30 is identified as a primary PRO variable of interest. [redacted]

The number of patients completing PRO questionnaires and the number of patients missing/expected to have PRO 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 missing items or missing assessments will be handled according to scoring manuals for each respective patient questionnaire (Fayers 2001; Oemar and Janssen 2013; Cleeland 2009).

Descriptive statistics (n, mean, median, SD, min, max) will be used to summarize the [redacted] scores from the EORTC QLQ-C30, [redacted] at each scheduled assessment time point. Additionally, change from baseline in the [redacted] scores at the time of each

assessment will be summarized. [redacted] 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.

A repeated measures model for longitudinal data will be used to compare the two treatment arms [redacted] This longitudinal model will include terms for treatment, the randomization stratification factors (presence of lung and/or liver metastases [yes versus no] and previous treatment with any CDK4/6 inhibitor [yes versus no]), time (duration in weeks counting from the time of baseline measurement to the time of a particular post baseline measurement) and baseline value as main effects, as well as an interaction term for treatment by time. Time will be explored as both a continuous and categorical variable to assess the best model fit. 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 tests if appropriate. 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. Data collected under treatment (i.e. while the patient is treated) and during post-treatment follow-up up until documented progression will be included. If the model fit based on unblinded data does not achieve convergence then fewer time points for inclusion may be considered.

Time to 10% deterioration in the global health status/[redacted] will be assessed in each cohort. 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, 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 dates. 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. Patients with no baseline data will be censored at Day 1. Death is considered as an event when it occurs within a period of time defined by 2 times the period between two assessments as planned in the study protocol. This avoids overestimating the time to definitive worsening in patients dying after an irregular assessment scheme. Patients who die after more than twice the planned period between two assessments since the last assessment are censored at the date of their last available questionnaire.

Time to 10% deterioration will be compared between the two treatment arms using a stratified log-rank test (strata based on IRT data) and 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 6, 12, 18 months. 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. Sensitivity analysis of time to definitive deterioration with different cut-off definitions (e.g. 5%, 15%) may also be considered if the number of events per arm is judged sufficient to draw relevant conclusions.

[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 and be presented by treatment arm for the *PIK3CA* mutant and non-mutant cohorts individually as well as combined. 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 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 system organ class, preferred term, and maximum CTC grade. A patient with multiple CTC grades for an AE will be summarized under the maximum CTC grade recorded for the event. In the summaries presented by grade, all AEs will be pooled regardless of whether they are CTC gradable or not. AE with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

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

Any information collected (e.g. CTC 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:

- AEs by SOC, PT, maximum CTCAE grade

- Most frequent AEs, regardless of study treatment relationship by PT, maximum CTCAE grade (*at least 5% incidence*)
- Treatment-related AEs by SOC, PT, maximum CTCAE grade
- Most frequent treatment-related AEs by PT, maximum CTCAE grade (*at least 5% incidence*)
- On-treatment deaths by primary SOC and PT
- On-treatment deaths and SAEs with fatal outcome by PT
- All deaths by SOC, PT. *(All deaths (on-treatment + post treatment) are included)*
- Serious adverse events by SOC, PT, maximum CTCAE grade
- Treatment-related serious adverse events by SOC, PT, maximum CTCAE grade
- Most frequent SAEs by PT and maximum CTCAE grade
- Most frequent treatment-related SAEs by PT and maximum CTCAE grade
- AEs leading to study drug discontinuation by SOC, PT and maximum CTCAE grade
- Treatment-related AEs leading to study drug discontinuation by SOC, PT and maximum CTCAE grade
- AEs requiring dose adjustment and/or interruption by SOC, PT, maximum CTCAE grade.
- AEs requiring medication or therapies by SOC, PT, maximum CTCAE grade.

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

3.9.1.5 Adverse Events of Special Interest

Specific groupings of Adverse Events of Special Interest (AESI) will be considered and the number of patients with at least one event in each grouping will be reported. Such groups consist of AEs for which there is a specific clinical interest in connection with alpelisib treatment (i.e. where alpelisib may influence a common mechanism of action responsible for triggering them) or AEs which are similar in nature (although not identical). The groups are defined according to the MedDRA terms defined in the program Case Retrieval Strategy (CRS) document and will be summarized. The latest version of the CRS document available at the time of the analyses will be used.

All AESI groupings are defined through the use of Preferred Terms (PT), High Level Terms (HLT) or System Organ Classes (SOC) or through a combination of these three components. An Excel file with the exact composition of the AEs 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. Note that certain AEs may be reported within multiple groupings. Final deliverables will be aligned with the final excel file. A listing of all grouping levels down to the MedDRA preferred terms used to define each AESI will be generated.

Groupings for Adverse Events of Special Interest

AESI grouping*	Definition*
Hyperglycemia	SMQ Hyperglycaemia /new onset diabetes mellitus (narrow)

AESI grouping*	Definition*
Rash	NMQ Rash (BYL719)
Pneumonitis	SMQ Interstitial lung disease (narrow)
GI Toxicity Nausea, Vomiting, Diarrhea	HLT Diarrhoea (excl infective)
	HLT Nausea and Vomiting symptoms

* At the time of the analyses, MedDRA-defined (SMQ) or Novartis-defined (NMQ) groupings for all identified and potential risks as described in the latest version of the CRS document available will be used

Standard data of analysis for AESI will be conducted as follows:

- The number (%) of patients with AESI will be reported by AESI grouping, maximum CTCAE grade and by treatment group.
- All AEs of special interest will be listed.
- Depending on the observed number of events, the time to first occurrence of any CTC grade ≥ 2 AESI may be summarized using Kaplan-Meier methods for each of the groupings in the table above. Median time to onset and 95% CI will be summarized. Ascending Kaplan-Meier plots will be generated. Please refer to Section 3.9.1.5.1 for definitions. For the AESI of 'Hyperglycemia', 'Rash' and 'GI Toxicity Nausea, vomiting and Diarrhea', additional analyses for time to first occurrence of any CTC grade ≥ 3 AESI will be presented.

3.9.1.5.1 Time to onset of any CTC grade ≥ 2 [≥ 3] event

Time to onset of CTC grade ≥ 2 [≥ *3*] event will be summarized using the Kaplan-Meier method. Median time to onset and 95% CI will be provided. In addition, Ascending Kaplan-Meier plots will be generated.

Time to onset of CTC grade ≥ 2 [≥*3*] event is defined as the time from the start of treatment to the start date of the first incidence of an event of CTC grade ≥ 2 [≥*3*] i.e. time in days is calculated as (start date of first occurrence of the event) – (date of first dose of study treatment) +1.

In the absence of an event during the on-treatment period, the censoring date applied will be the earliest of the following dates:

- end date of on-treatment period (end of study treatment + 30 days).
- death date
- start date of new antineoplastic therapy (with the exception of palliative radiotherapy or fulvestrant monotherapy) before experiencing any CTC grade ≥ 2 [≥*3*] event.
- data cut-off date.
- withdrawal of informed consent date

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.

If laboratory values are provided as '<X' (i.e. below limit of detection) or '>X', prior to conversion of laboratory values to SI unit, these numeric values are set to X.

Grade categorization of lab values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. 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 the document "Novartis internal criteria for CTC grading of laboratory parameters". The latest available version of the document based on the underlying CTCAE version 4.03 at the time of analysis will be used.

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):

- Worst post-baseline CTC grade (regardless of the baseline status). Each subject 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.

3.9.2.1 Blood Glucose parameters

HbA1c, fasting plasma glucose and fasting C-peptide data will be summarized in tables by time point and treatment. Summary statistics include number of patients with available data, mean, standard deviation, median, minimum and maximum. Figures of mean glucose/C-peptide levels with two-sided 95% confidence intervals over time by treatment may also be produced to view the trends over time.

For plasma glucose only, the following summaries will be provided:

- Time to first occurrence of CTC Grade >=2 hyperglycemia
- Time to first occurrence of CTC Grade >=3 hyperglycemia
- Time to resolution of CTC Grade >=2 hyperglycemia
- Time to resolution of CTC Grade >=3 hyperglycemia

Note that CTC Grade 3 and 4 hyperglycemia events use both fasting and non-fasting plasma glucose laboratory values. Median time to onset, duration and 95% CI will be summarized. In addition, ascending Kaplan-Meier plots will be generated. In addition, median and range of time to event will be summarized.

A plot of baseline HbA1c vs. worst post-baseline HbA1c (%) will be presented.

Time to first occurrence of Grade ≥ 2 [≥ *3*] lab event

Time to onset of first grade 2 [≥ *3*] or worse toxicity is defined as the time from the start of treatment to the start date of the first incidence of grade 2 [≥ *3*] or worse toxicity i.e. time in days is calculated as (start date of first occurrence of grade 2 [≥ *3*] or worse toxicity) – (date of first dose of study treatment) +1.

In the absence of grade 2 [≥ *3*] or worse toxicity during the on-treatment period, the censoring date will be the earliest date from the following dates: last date of administration of study treatment + 30 days, analysis cut-off date, new anti-cancer therapy start date, death date, withdrawal of informed consent date and last non-missing assessment for the lab parameter if the event is lab based).

Note: patients who have grade 2 [≥ *3*] or worse toxicity at the baseline will be excluded from this analysis.

Time to resolution (i.e. duration) of Grade ≥ 2 [≥ 3] lab event

Time (days) to resolution of an event is defined as time from first onset to the date of resolution of the event: (date of resolution of event) – (date of first onset of event) + 1. Resolution of an event means that there is a lab value returning to grade ≤ 1. The following lab parameters will be analyzed:

- Grade 2 or worse Hyperglycemia
- Grade 3 or worse Hyperglycemia

Time to resolution of an event will be presented for the subset of the Safety Set who experienced the event.

In the absence of a resolution during the on-treatment period, the censoring date is the earliest of the following dates: end of treatment + 30 days, analysis cut-off, new anticancer therapy start date, death date, withdrawal of informed consent date and last non-missing assessment for the lab parameter.

Ascending Kaplan-Meier curves will be constructed by treatment arm. Medians together with 95% confidence intervals will be presented for each treatment arm. 25th percentile and 75th percentile may be considered as well. In addition, the median and range of time to resolution for patients with an event will also be summarized.

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.
- Urinary parameters for all patients

In order to summarize labs parameters as applicable, collected over time (including unscheduled visits), the assessments will be time-slotted.

Time windows for laboratory parameters

Assessment (1)	Target day of assessment	Time Interval

Baseline		≤ Day 1
Cycle 1 Day 8	8	Day 2 to Day 11
Cycle 1 Day 15	5	Day 12 to Day 18
Cycle 2 Day 1	29	Day 26 to day 35
Cycle 2 Day 15	43	Day 36 to day 49
Cycle 3 Day 1	57	Day 50 to day 70
Cycle 4 Day 1	85	Day 71 to day 98
Cycle k Day 1 (k≥5)	=(k-1)*28+1	Day d-14 to day d+13
End of Treatment		Assessment taken at the EOT visit

(1) HbA1c is measured on-treatment at Cycle 3 Day 1 and EOT and same rule applies.

For laboratory parameters, all scheduled/unscheduled assessments should be assigned to time windows. In case of multiple values per window, the one closest to the planned visit date should be used. If 2 values are equidistant to the planned visit date, the selection should be made by selecting the one assessed by central (if any) and otherwise - for multiple central assessments equidistant to the planned visit - the last value.

Further derivation of laboratory parameters might be required for CTCAE grading. For instance, corrected calcium can be derived using the reported total calcium value and albumin at the same assessment using the following formula:

Corrected Calcium (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4]

3.9.2.2 Liver function parameters

Liver function parameters of interest for alpelisib are total bilirubin (TBIL), ALT, AST and alkaline phosphatase (ALP).

The number (%) of patients with worst post-baseline values (maximum post-baseline values) as per Novartis Liver Toxicity guidelines will be summarized:

- ALT or AST > 3xULN
- ALT or AST > 5xULN
- ALT or AST > 10xULN
- ALT or AST > 20xULN
- ALP > 1.5xULN
- TBL > 1.5xULN
- TBL > 2xULN
- ALT or AST > 3xULN & TBL > 2xULN (without time window)
- ALT or AST > 3xULN & TBL > 2xULN & ALP < 2xULN (without time window)*

*Potential Hy's Law events (candidates) are defined as those patients with AST or ALT > 3xULN and TBL > 2xULN and ALP < 2xULN at any visit during the on-treatment period.

A figure displaying time course of hepatic function tests (ALT, AST, TBL, ALP) in patients with potential Hy's law cases will be displayed.

Additional categories may be added to the above list based on any updates to the internal guidelines on collection, analysis, and presentation of liver safety data.

For lab events the censoring date will be the date of the last scheduled/unscheduled lab assessment with a non-missing value for the lab parameter of interest that was collected prior to the time of occurrence of the censoring reason.

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
- Pulse rate: >=100 and increase from baseline of >25%
- Weight: Increase ≥ 10% from baseline

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
- Pulse rate: <=50 and decrease from baseline of >25%
- Weight: Decrease ≥ 10% from baseline

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

- 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 by the central reader at each scheduled time point for each patient. 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, PR, and QRS intervals in ms, 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. Individual listings will be provided by subject.

Table 3-8 Clinically notable ECG values

ECG parameter (unit)	Clinically notable criteria
QT, QTcF (ms)	New value of > 450 and ≤ 480 ms
	New value of > 480 and ≤ 500 ms
	New value of > 500 ms
	Increase from Baseline > 30 to ≤ 60 ms
	Increase from Baseline > 60 ms
PR duration (ms)	Increase > 25% from baseline and to PR duration > 200
	New value of > 200 ms
QRS duration (ms)	Increase > 25% from baseline and to QRS duration > 120 ms
	New value of > 120 ms
Heart Rate (bpm)	Decrease from Baseline of > 25% and to a value < 50 bpm
	Increase from Baseline of > 25% and to a value > 100 bpm

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 last ECG measurements taken during the screening phase. 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.

3.9.5 Cardiac imaging (MUGA / ECHO)

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.

For left ventricular ejection fraction (LVEF) a shift table using CTC grades for 'Ejection fraction - decrease' as defined per CTCAE version v4.03 to compare baseline to the worst on-treatment 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

Other safety data (e.g. data relating to liver events) will be listed in the safety set.

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.

3.10 Pharmacokinetic (PK) analyses

3.10.1 General principle

All PK analyses for alpelisib and fulvestrant will be based on the PK Analysis set. However, only samples taken within the following time windows around the scheduled time points and at steady state will be used for summary tables and profiles which are summarized by time point:

- Pre-dose: prior to dosing on the assessment day and collected at approximately 24 ± 2 hours after the last dose
- 1 h post-dose: within ± 10 minutes of the scheduled time point
- 2, 4, 6 or 8 h post-dose: within ± 30 minutes of the scheduled time point

In addition the following criteria for alpelisib samples considered at steady state must be met:

- assessments with at least 3 continuous days of daily dosing at the planned dose (dose assigned at study entry) prior to the day of alpelisib PK assessment;
- no vomiting occurs within the first 4 hours of the last dose (pre-dose trough samples);
- no vomiting occurs within the first 4 hours of the current and last dose (post-dose sparse samples).

All samples regardless of time window or steady state status will be used for the individual time profile figures.

3.10.2 PK concentrations

Trough PK concentrations of alpelisib and fulvestrant will be summarized by visit and treatment group. Alpelisib sparse PK concentrations will be reported at Cycle 1 Day 15.

Descriptive statistics of concentrations will be provided using steady state concentrations only and include n, number of non-zero concentrations, arithmetic mean, geometric mean, median,

SD, coefficient of variation CV (%), geometric CV (%), minimum and maximum. Coefficient of variation CV (%) is calculated as follows from non-zero values: 100*(SD/arithmetic mean).

Geometric CV (%) is calculated as follows from non-zero values:

$$CV(\%) = 100 \cdot \sqrt{\exp(\hat{\sigma}^2) - 1}$$

where $\hat{\sigma}^2$ denotes the variance of the log-transformed values.

Geometric mean and arithmetic mean (SD) plots will also be graphically presented for trough concentration-time data. Individual concentration-time profiles will be displayed graphically for trough concentrations (alpelisib and fulvestrant) and sparse concentrations (alpelisib only) separately. All PK concentration data will be listed as appropriate.

3.10.3 Population PK parameters for alpelisib

The population PK analysis will be conducted in a separate report.

[redacted]

[redacted]

[redacted]

3.10.5 Handling missing and invalid values

The lower limits of quantitation (LLOQ) are currently 5.0 ng/mL for alpelisib and 1.0 ng/mL for fulvestrant. Values below the assay LLOQ will be reported as 0 ng/mL. All concentrations below the LLOQ will be displayed in listings as zero with a flag and handled as zero in any calculations of summary statistics, but handled as missing for the calculation of the geometric means and their CV. Any missing PK parameter data will not be imputed.

Invalid concentrations will be flagged in the PK concentration data set by the CP expert after the merge of clinical and bioanalytical data has taken place. Flagged values will not be included in summaries and will be listed only.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

3.12 Interim analyses

3.12.1 Primary endpoint: Progression-free survival

PIK3CA mutant cohort:

Two interim analyses are planned after approximately 97 and 185 of the 243 targeted PFS events (40% and 76% information fractions respectively) have been documented. These analyses are expected to take place around 19 and 25 months from the date of first patient randomized in the study. Approximately 243 patients are expected to be randomized when the 97th PFS event occurs at the time of the first interim analysis if H_0 is true (HR.=1). The primary intent of the first interim analysis is to allow the cohort to stop early for lack of efficacy (futility). There is no intent to carry out an analysis to declare superior efficacy at the time of the first interim analysis. The second interim analysis will allow the study to stop early for outstanding efficacy. The second interim analysis will only be carried out after all patients have been randomized in the PIK3CA mutant cohort and approximately 76% of the 243 targeted PFS events have been observed.

The assessment of futility will be guided based on two criteria.

A user-defined gamma spending function (γ=5) will be used as a beta-spending function to determine the non-binding futility boundary. One important feature of the design is that the

efficacy stopping boundaries will not be affected by the presence of non-binding futility stopping boundaries.

Based on the choice of α-spending and β-spending function described above, the futility boundary in terms of p-value scale at the interim is calculated as p =0.128 (or Z=1.134; hazard ratio=0.794). The observed (i.e. nominal) p-value has to be greater than 0.128 to conclude futility according to this criterion at the time of the first interim analysis.

In addition to the stopping boundary based on the β-spending function described above, DMC members will be instructed to include in their recommendation whether the conditional probability of observing a clinically relevant PFS treatment effect at the final PFS analysis is less than 0.20, i.e.:

$$\text{Conditional probability } (HR_{final} \leq 0.6 | HR_{interim}) < 0.20$$

This criterion uses the observed interim data and an assumption regarding the distribution of future unobserved data at the final analysis, conditioned under the alternative hypothesis (HR.=0.6). The futility boundary for this criterion in terms of p-value scale at the interim is calculated as p=0.068 (or |Z|=1.489; hazard ratio=0.739). Thus the observed (i.e. nominal) p-value has to be greater than the p-value scale futility boundary = 0.068 to conclude futility according to this criterion.

Details of the methodology as well as the operating characteristics for the futility criterion based on conditional probability is described in Section 4.7.

Therefore at the time of the futility analysis, the PIK3CA mutant cohort may be stopped for futility if one or both of the criteria are met.

In addition, the predictive probability of success based on the final planned number of PFS events will be calculated given the interim data, and provided to the DMC at the time of the futility interim analysis as supportive information.

A Haybittle-Peto stopping boundary (as implemented in East 6.3) will be used for interim and final PFS analyses. At the second interim analysis, the observed p-value has to be less than or equal to 0.0001 (or Z=3.719) in order to conclude superior efficacy. If the study continues to final analysis, the p-value that will be used to declare statistical significance at the final analysis will be 0.0199 (Z=2.054).

Since the observed number of events at the interim analyses may not be exactly equal to the planned number of events, the efficacy and futility boundaries will need to be re-calculated (or updated) based on the actual number of observed events using the pre-specified Haybittle-Peto boundary and β-spending functions. Therefore, the observed p-values at the interim analyses will be compared with the updated boundaries.

If the study continues to final analysis, the p-value that will be used to declare statistical significance at the final analysis will be based on the actual number of PFS events documented at the cut-off date for the final analysis and the alpha already spent at the interim analysis. Therefore, if the interim analyses were carried out after exactly 40% and 76% of the planned number of events, and the cohort continued until the final analysis, the observed p-value will have to be less than 0.0199 to declare statistical significance. If the number of events in the final

analysis deviates from the expected number, the final analysis criteria will be determined so that the significance level is maintained at 0.02 in this cohort.

Statistical properties of the group sequential design in this cohort are summarized in Table 3-10.

Table 3-10 Simulated probabilities to stop for futility or efficacy at the interim analysis in the *PIK3CA* mutant cohort

Scenario	Look	# PFS events	Simulated cumulative probabilities (%)		Simulated incremental probabilities (%)	
			Stop for efficacy	Stop for futility	Stop for efficacy	Stop for futility
Under H_0 (HR=1)	Interim 1	97	0	92.98	0	92.98
	Interim 2	185	0.02	-	0.02	-
	Final	243	0.98	-	0.96	-
Under H_a (HR=0.6)	Interim 1	97	0.34	15.31	0.34	15.31
	Interim 2	185	38.50	-	38.50	-
	Final	243	83.95	-	45.45	-
Note: The study will not stop for outstanding efficacy at the first (futility) interim analysis. Simulation is performed in SAS v9.4 with number of simulations = 10,000 and randomization seed =111064						

3.12.2 Key Secondary Endpoint: OS in the PIK3CA mutant cohort

OS will be compared between the two treatment groups, provided the primary endpoint PFS is statistically significant favouring alpelisib. A hierarchical testing procedure will be adopted in this study and OS will be tested only if the primary efficacy endpoint PFS is statistically significant. A maximum of three analyses are planned for OS:

- The first potential time point for OS analysis will be at the time of the PFS efficacy interim analysis after approximately 37% of the expected deaths are observed, at which point approximately 66 deaths are expected. If PFS is not statistically significant at this stage, then OS will not be tested, in which case the next potential time point for OS analysis will be at the time of the final PFS analysis after approximately 57% of the expected deaths are observed, at which point approximately 101 deaths are expected to have been recorded in the clinical database.
- If OS is not statistically significant at the first interim analysis, the 2nd OS analysis will be planned after approximately 85% of the expected deaths are observed, at which point approximately 151 deaths have been recorded in the clinical database. If OS is not statistically significant at this stage, a final analysis is planned at the time approximately 178 deaths have been recorded.
- If PFS is not statistically significant at the final analysis for PFS, then OS will not be tested.

The type I error probability will be controlled by using a separate Lan-DeMets (O'Brien-Fleming) alpha spending function independent of the Haybittle-Peto boundary used for the primary efficacy analysis of PFS at a 2.0% level of significance for the *PIK3CA* mutant cohort.

This guarantees the protection of the overall type I error ($\alpha = 2.5\%$) across all hypotheses and the repeated testing of the OS hypotheses at the interim and the final analyses (Glimm 2010). This includes hypotheses associated with the secondary endpoints PFS and OS in the *PIK3CA* non-mutant cohort (PFS in the non-mutant cohort will be tested at a 0.5% level of significance if PoC is established). The trial allows for the stopping of a cohort for a superior OS result, provided the primary endpoint PFS has already been shown to be statistically significant favouring the alpelisib arm. 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.

Given the hierarchical testing strategy of PFS and OS, the design concerning OS analyses will have the following characteristics based on simulations in East 6.3. The probabilities shown in Table 3-11are conditional probabilities (conditional on PFS being statistically significant) not marginal probabilities.

At the final PFS analysis in the *PIK3CA* mutant cohort:

- The cumulative probability to show efficacy on OS (alternative hypothesis H_a is true) by the final analysis is 71.57%; while the cumulative type I error (rejecting the null hypothesis H_0 if H_0 is true) is 1.91%.
- The cumulative probability to detect efficacy on OS if the alternative hypothesis H_a is true is 19.14% at the first interim analysis, 56.60% at the second interim analysis and 71.57% at the final PFS analysis.

Statistical properties the *PIK3CA* mutant cohort are summarized in Table 3-11.

Table 3-11 Simulated probabilities to stop for efficacy at 1st interim, 2nd interim, or final OS analysis in the *PIK3CA* mutant cohort

Scenario	Look	# deaths	Simulated cumulative probabilities (%)*	Simulated incremental probabilities (%)*
			Stop for efficacy	Stop for efficacy
Scenario 1: The first IA for OS is performed at the time of the PFS IA for efficacy				
Under H_0 (HR.=1)	1st Interim	66	0.01	0.01
	2nd Interim	151	1.11	1.09
	Final OS	178	1.89	0.79
Under H_a (HR.=0.67)	1st Interim	66	1.94	1.94
	2nd Interim	151	57.01	55.07
	Final OS	178	71.62	14.61
Scenario 2: The first IA for OS is performed at the time of the final PFS analysis				
Under H_0 (HR.=1)	1st Interim	101	0.17	0.17
	2nd Interim	151	1.12	0.95
	Final OS	178	1.91	0.79
Under H_a (HR.=0.67)	1st Interim	101	19.14	19.14
	2nd Interim	151	56.60	37.46

Scenario	Look	# deaths	Simulated cumulative probabilities (%)*	Simulated incremental probabilities (%)*
			Stop for efficacy	Stop for efficacy
	Final OS	178	71.57	14.97

Note: Simulations are performed in East 6.3 with number of simulations = 10,000 and randomization seed =37059.
* Probabilities are reported as if OS was tested alone, regardless the testing strategy with PFS. The true probabilities should take into account the probability of PFS at each look. Simulated probabilities shown based on the alpha allocated for testing (p=0.02).

At the time of final PFS analysis in this cohort, 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).

3.12.3 OS in the PIK3CA non-mutant cohort:

OS will be compared between the two treatment groups, provided the PFS is statistically significant favouring alpelisib, in the *PIK3CA* non-mutant cohort. A hierarchical testing procedure will be adopted in this study and the OS analyses will be tested only if PFS is statistically significant. A maximum of three analyses are planned for OS:

- at the time of the final analysis for PFS (provided PFS is statistically significant) when approximately 29% of the expected deaths are observed, at which point approximately 36 deaths in the *PIK3CA* non-mutant cohort have been recorded (after approximately 18 months from the first patient to be randomized in this cohort);
- at the time when approximately 87% of the expected deaths are observed, at which point approximately 109 deaths in the *PIK3CA* non-mutant cohort have been recorded (after approximately 45 months from the first patient to be randomized in this cohort);
- a final analysis for OS when approximately 125 deaths in the *PIK3CA* non-mutant cohort have been recorded (approximately 54 months from date of first patient to be randomized in this cohort).

An α-spending function according to Lan-DeMets (O'Brien-Fleming will be used to maintain the overall type I error probability (Lan and DeMets 1983). 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.

Given the hierarchical testing strategy of PFS and OS, the design concerning OS analyses will have the following characteristics based on simulations in East 6.3. The probabilities shown in Table 3-12 are conditional probabilities (conditional on PFS being statistically significant) not marginal probabilities.

Table 3-12 Simulated probabilities to stop for efficacy at 1st interim, 2nd interim, or final OS analysis in the *PIK3CA* non-mutant cohort

Scenario	Look	# deaths	Simulated cumulative probabilities (%)*	Simulated incremental probabilities (%)*
			Stop for efficacy	Stop for efficacy
Under H_0 (HR.=1)	1st Interim	36	<0.001	<0.001
	2nd Interim	109	0.31	0.31
	Final OS	125	0.59	0.28
Under H_a (HR.=0.67)	1st Interim	36	0.001	0.001
	2nd Interim	109	24.59	24.58
	Final OS	125	36.74	12.15

Note: Simulations are performed in East 6.3 with number of simulations = 10,000 and randomization seed =60030. 1st Interim analysis will be conducted at the time of final analysis for PFS

* Probabilities are reported as if OS was tested alone, regardless the testing strategy with PFS. The true probabilities should take into account the probability of PFS at each look. Simulated probabilities shown based on the alpha allocated for testing (p=0.005).

At the time of final PFS analysis in this cohort, both PFS and interim OS analysis will be performed by the independent statistical group for the DMC. The Novartis Clinical team will remain blinded to study treatment allocations up until such point the PIK3CA mutant cohort can be unblinded. 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).

3.12.4 Confidentiality of Interim OS results

At the time of final PFS analysis in the PIK3CA mutant cohort, 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).

3.13 Subgroup analysis

3.13.1 Safety

The objective for carrying out these subgroup analyses is to identify potential safety issues that may be limited to a subgroup of patients, or safety issues that are more commonly observed in a subgroup of patients

The main safety analyses will be repeated in the following subgroups of patients:

- Age >=65 vs age <65 years
- Age >=75 vs age <75 years
- Race (white vs Asian vs black or African American vs other)

These main safety analyses include:

- AEs, by SOC, PT, maximum CTCAE grade

- Treatment-related AEs by SOC, PT, maximum CTCAE grade
- Serious adverse events (SAE) by SOC, PT, maximum CTCAE grade
-
- For the AESI of Hyperglycemia only, a subgroup analysis by hyperglycemia diagnosis status at baseline per American Diabetes Association (ADA) 2017 will be presented:
    - Diabetic: FPG ≥ 7.0 mmol/l or 126 mg/dl or HbA1c ≥ 6.5 % vs.
    - Pre-diabetic: FPG 5.6- <7.0 mmol/l or 100-125 mg/dl or HbA1c 5.7- <6.5% vs.
    - Normal: [FPG <5.6 mmol/l or <100 mg/dl] and HbA1c <5.7%

3.13.2 Efficacy

[redacted]

If the primary analysis for PFS is statistically significant in a cohort, subgroup analyses will be performed on the FAS with the same statistical model used for the point estimate in the full cohort. Analyses will be performed for each subgroup one by one, i.e. fitting a model using only patients available in the subgroup of consideration.

[redacted]

[redacted]

3.14 Sample size calculation

The median TTP for fulvestrant in first line post-menopausal advanced breast cancer patients with endocrine sensitive disease is estimated to be between 8 months (Howell 2004) and 23 months in FIRST trial (Robertson 2014). For sample size calculation, it is assumed that approximately 8% patients in the PIK3CA mutant cohort and 15% of patients in the PIK3CA non-mutant cohort will comprise these patients with an expected median PFS for fulvestrant of 18 months.

Two main studies have been reported assessing fulvestrant in relapsed advanced breast cancer: SoFEA trial with a median PFS of 4.8 months (Johnston 2013) and CONFIRM trial with a median PFS of 6.5 months (Di Leo 2010). However, in the SoFEA study after a first induction with fulvestrant at 500 mg, the dose of fulvestrant continued at 250 mg; in CONFIRM fulvestrant was given at 500 mg throughout. Approximately 92% patients in the PIK3CA mutant cohort and 85% of patients in the PIK3CA non-mutant cohort enrolled in the current study will have similar clinical features to the population treated in CONFIRM trial, therefore

for sample size calculation, the median PFS for fulvestrant in this group is assumed to be 6.5 months.

For the overall population in the PIK3CA mutant cohort, the median PFS in the control arm (fulvestrant + placebo) is estimated via simulation to be around 7.0 months.

For the overall population in the PIK3CA non-mutant cohort, the median PFS in the control arm (fulvestrant + placebo) is estimated via simulation to be around 7.4 months.

It is expected that treatment with alpelisib + fulvestrant in both cohorts will result in a 40% reduction in the hazard rate (corresponding to an increase in median PFS from 7.0 months to 11.67 months in the PIK3CA mutant cohort and from 7.4 months to 12.33 months in the PIK3CA non-mutant cohort, under the exponential model assumption).

Patients with PIK3CA mutant status:

If the true hazard ratio is 0.6 (under alternative hypothesis), a total of 243 PFS events are required to have 83.80% power at an one-sided overall 2.0% level of significance to reject the null hypothesis (HR.=1) using a log-rank test for a 3-look group sequential design using a Haybittle-Peto boundary to determine the efficacy boundary along with (i) a gamma spending function ($\gamma = 5$) and (ii) a conditional probability function to determine the non-binding futility boundaries. Assuming that 40% of the patients will have a *PIK3CA* mutant status, an enrollment rate of 12 patients per month during the first 6 months (5 per month with *PIK3CA* mutant status), 35 patients per month up to 12 months (14 per month with *PIK3CA* mutant status) and 59 patients per month afterwards (24 per month with *PIK3CA* mutant status) and 10% patients will be lost to follow-up for PFS final analysis, a total of 340 patients will need to be randomized in this cohort to the two treatment arms in a 1:1 ratio. Given the above assumptions, it is estimated that the 243rd PFS event will be observed at approximately 32 months from the date of first patient randomized in the cohort.

The estimated timelines for interim and final PFS analyses are provided in Table 3-13.

Table 3-13 Estimated timelines for interim and final PFS analyses in the *PIK3CA* mutant cohort

Look	Months after randomization of the first patient	Number of PFS events	Number of patients expected to be randomized (HR.=1).
Interim 1 (futility)	19	97	271
Interim 2 (efficacy)	25	185	340
Final	32	243	340

Patients with PIK3CA non-mutant status:

The proof of concept criteria require:

(a) an estimate for PFS HR reaching a critical value i.e. HR ≤ 0.60

(b) strong evidence that the treatment results in a HR that is better than the value of no interest. i.e. Posterior Probability (HR < 1) $\geq 90\%$

Based on the dual criteria a minimum of 102 PFS events are required (please refer to Section 3.8.2.2). Assuming an enrollment rate of 12 patients per month during the first 6 months (7 per month with *PIK3CA* non-mutant status), 35 patients per month up to 12 months (21 per month with *PIK3CA* non-mutant status) and 59 patients per month afterwards (35 per month with *PIK3CA* non-mutant status) and 10% patients will be lost to follow-up, 220 patients will be randomized (110 per arm), in order to observe the required 102 PFS events in approximately 18 months (if the observed HR is 0.60 and the median PFS for the control arm is 7.4 months).

The primary analysis to estimate the HR will be performed after approximately 102 PFS events have been observed. If the true HR is 1, the probability (obtained by simulation) to obtain a positive conclusion is 0.005; if the true HR is 0.50, the probability to meet efficacy criteria is 0.813. If the true HR is 0.60 (reflecting the minimum clinically relevant difference), the probability to meet efficacy criteria is 0.491. Simulation results are provided in Table 3-14.

Table 3-14 Operating characteristics for PoC criteria in *PIK3CA* non-mutant cohort

True HR	True Median PFS alpelisib (months)	Probability to PoC	Expected duration of cohort (months)
0.3	24.67	0.999	21.0
0.4	18.50	0.975	19.7
0.5	14.80	0.813	18.7
0.6	**12.33**	**0.491**	**18.0**
0.7	10.57	0.220	17.4
0.8	9.25	0.076	16.9
0.9	8.22	0.020	16.5
1.0	7.4	0.005	16.0
Assumes: (1) HR.=0.6, (2) true median PFS for fulvestrant = 7.4 months, (3) protocol planned accrual rates, (4) Analysis after 102 PFS events have been observed			

Audit size for BIRC assessed PFS in the PIK3CA mutant cohort

The audit size of the sample-based BIRC assessment will be 50% of all randomized patients in the PIK3CA mutant cohort. 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 -0.51 (i.e. HR.=0.60), the audit size of 50% will ensure that the upper bound of a one-sided 95% CI for BIRC-based log-hazard ratio has 94% probability of being below 0 (i.e. HR. < 1) if the correlation between investigator assessment and BIRC assessment is 0.65 (the estimated correlation based on data from the Bolero-2 [CRAD001Y2301] study in metastatic breast cancer).

3.15 Power for analysis of key secondary variable

For first line patients no phase III data are available with single agent fulvestrant. Data from phase III studies have been reported with letrozole showing a median OS of 34 months (Mouridsen 2003) and with anastrozole showing a median OS of 38 months (Bergh 2012). OS

data with fulvestrant alone in first line setting have been recently presented within the phase II study FIRST. In that study the median OS for fulvestrant was 54 months (Robertson 2014). Therefore, for sample size calculation, median OS for fulvestrant alone in the current study for patients with progression more than 12 months from completion of (neo)adjuvant endocrine therapy is assumed to exceed 50 months.

Median OS for fulvestrant in relapsed post-menopausal advanced breast cancer patients is estimated to be between 19 months (SoFEA trial, Johnston 2013) and 26 months (CONFIRM trial, Di Leo 2013). For sample size calculation, the median OS for fulvestrant in second line is thus assumed to be 26 months.

Based on the expected split of the patient population as mentioned in Section 3.14, the median OS of control arm is estimated via simulation to be approximately 30 months. It is hypothesized that adding alpelisib to fulvestrant will result in a 33% reduction in the hazard rate for OS (corresponding to an increase in median survival to 44.8 months).

Patients with PIK3CA mutant status:

OS will be compared between the two treatment groups, provided that the primary endpoint PFS is statistically significant in this cohort of patients. If the true hazard ratio is 0.67 (under alternative hypothesis), a total of 178 deaths are needed to be observed to have 72% power at an one-sided overall 2.0% level of significance to reject the null hypothesis (HR.=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 and assuming 5% dropout rate by the time of the OS final analysis, it is estimated that these 178 deaths will be observed at approximately 54 months from the date of first patient to be randomized in this cohort.

The estimated timelines for interim and final OS analyses are provided in Table 3-15.

Table 3-15 Estimated timelines for interim and final OS analyses in the *PIK3CA* mutant cohort

Look	Months after randomization of the first patient	Number of OS events
1st OS Interim at time of interim PFS analysis	25	66
1st OS Interim at time of final PFS analysis	32	101
2nd Interim	45	151
Final	54	178

Patients with PIK3CA non-mutant status (secondary endpoint):

OS will be compared between the two treatment groups, provided that the endpoint PFS is statistically significant in this cohort of patients. The final analysis of OS for the *PIK3CA* non-mutant cohort will be performed at approximately 54 months from the date of first patient to be randomized in the *PIK3CA* non-mutant cohort. Based on the same number of patients that are planned to be enrolled in the *PIK3CA* non-mutant cohort and assuming 5% dropout rate by the time of the OS final analysis, it is estimated that approximately 125 deaths will be observed. If the true hazard ratio is 0.67 (under alternative hypothesis), a total of 125 deaths will allow

36.1% power at a one-sided overall 0.5% level of significance to reject the null hypothesis (HR.=1) using a log-rank test and a 3-look group sequential design.

The power calculations were conducted with software package East 6.3.

The estimated timelines for interim and final OS analyses are provided in Table 3-16.

Table 3-16 Estimated timelines for interim and final OS analyses in the *PIK3CA* non-mutant/ cohort

Look	Months after randomization of the first patient	Number of OS events
1st Interim	18	36
2nd Interim	45	109
Final	54	125

3.16 Sample size considerations for PK analysis

Sample size for PK analysis was based on feasibility only.

4 General Statistical Methodology

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 provided by cohort and grouped by treatment arms, but no p-values will be provided.

4.2 Center pooling

All study centers will be combined for the analysis unless otherwise specified. No center effect will be assessed due to expected small size of centers.

4.3 One-sided vs. two-sided test

One-sided tests will be used in this study for the primary endpoint (PFS) and key secondary endpoint (OS) at the alpha levels specified in Section 3.8. Confidence intervals will be estimated as two-sided.

4.4 Time-to-event analyses

The following sections present a general methodology to be used to analyze the following time-to-event variables.

4.4.1 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.

4.4.2 Hypothesis and test statistic

For the primary, key-secondary and secondary efficacy endpoints, the comparison between the two treatment groups will be performed using a stratified log-rank test at levels α described in Section 3.8.

The ***stratified log-rank*** test (strata information will be based on the data obtained from IRT that was utilized for randomization) will be implemented as follows:

General SAS code for the stratified log-rank test

```
PROC LIFETEST data=dataset METHOD=KM;
BY stratum;
TIME survtime*censor(1);
STRATA trt;
RUN;
```

/* *stratum* represents stratum variable (to be included for stratified analysis only);

survtime represents variable containing event/censor times;

censor represents censoring variable (1=censored, 0=event);

trt represents treatment arm variable; */

For each of the K=2 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.

4.4.3 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 (see examples above). Median survival for each treatment arm will be obtained along with 95% confidence intervals calculated from PROC LIFETEST output using the loglog option available within PROC LIFETEST, Kaplan-Meier estimates with 95% confidence intervals at specific time points will be summarized.

The Kaplan-Meier graphs will be constructed using Splus software. The statistics (test statistics, p-value, hazard ratio etc.) displayed on the graph will, however, be obtained from the SAS software.

4.4.4 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). The strata information will be based on the data obtained from IRT that was utilized for randomization.

General SAS code for the stratified Cox model

```
PROC PHREG data=dataset;
MODEL survtime*censor(1)=trt / TIES=EXACT;
STRATA stratum 1 stratum 2;
RUN;
```

/* *survtime* represents variable containing event/censor times;

censor represents censoring variable (1=censored, 0=event);

trt represents treatment arm variable;

stratum 1 and stratum *2* represent IRT stratification variables */

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

Note: Ideally, the hazard ratio and the confidence interval should be derived by a method consistent with the p-value calculation, i.e. in this case with log-rank test. This requirement would lead to the score test based intervals. However, score test-based confidence intervals are not available in SAS procedure PHREG therefore Wald test based intervals will be used instead.

[redacted]

4.5 Group sequential design used in Phase III studies

The statistical methodology for the interim analyses will be based on group sequential methodology.

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- and β-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 in the PIK3CA mutant cohort, 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 2.0% for a one-sided test in the PIK3CA mutant cohort).

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. if recruitment is much faster than planned and an early efficacy comparison cannot be performed or 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.

4.5.1 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).

4.5.2 Beta-spending function

The stopping boundaries to be used for the futility test will be calculated using the *Gamma* family β-spending function approach (Hwang et al, 1990). The spending function for one-sided test has the following functional form:

$\beta(t) = \beta\ (1 - e^{-\gamma t}) / (1 - e^{-\gamma})$, where t represents the information fraction

A non-binding user-defined beta-spending function will be used to determine the futility boundary. The futility boundary will be constructed so that the critical value will be determined based on the gamma function (γ=5). The choice of non-binding nature of the futility stopping boundary ensures that the efficacy stopping boundaries are not affected.

Negative values of γ yield convex spending functions that increase in conservatism as γ decreases, while positive values of γ yield concave spending functions that increase in aggressiveness as γ increases.

4.5.3 Calculation of stopping boundaries

The stopping boundaries for PFS in the PIK3CA mutant cohort were calculated at the design stage by selecting a nominal p-value that met the high threshold of the efficacy criteria as determined in the protocol.

The following description applies to the PIK3CA mutant cohort with 2 interim analyses allowing stopping for futility and outstanding efficacy and assumes a 1-sided test. For this study since there is no futility assessment in the second interim analysis, the formula is altered to fit the particular case.

Let $\boldsymbol{u(t_1)}$, $\boldsymbol{u(t_2)}$, $\boldsymbol{u_F}$ denote the upper critical boundaries for efficacy to be used for the test statistics $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$ and $\boldsymbol{Z_F}$ at the 1st interim, the 2nd interim and the final analysis, respectively, and let $\boldsymbol{l(t_1)}$ denote the lower futility boundaries at the 1st interim analysis (note: $\boldsymbol{l_F = u_F}$ by construct, i.e., the boundaries meet at the final look, thereby ensuring that a decision about the 2nd hypotheses will indeed be made). Let $\boldsymbol{P_0}$ denote probabilities determined under the null hypothesis, and $\boldsymbol{P_a}$ denote probabilities under the alternative hypothesis (used for the sample size or power calculation). Furthermore, $\alpha(t)$ and $\beta(t)$ are the predefined spending functions for efficacy and futility, respectively, and the desired overall significance level is $\alpha = 0.02$ (since a 1-sided test is assumed).

The critical values $u(t_1)$ and $l(t_1)$ for the 1st interim analysis are obtained such that

$P_0(Z_1 \geq u(t_1)) = \alpha(t_1)$ and $Pa(Z_1 \leq l(t_1)) = \beta(t_1)$,

where $\alpha(t_1)$ and $\beta(t_1)$, denote the α-level and β-level, respectively, spent at informationfraction t_1 and determined from the corresponding spending functions (t_1 is calculated using the actual number of events observed at 1st interim and assuming the final number of PFS events is 243).

Critical boundaries for the second interim and the final analysis are calculated recursively such that efficacy boundaries do not depend on the futility boundaries. Given that the boundaries for the 1st interim analysis have been computed, the critical boundary for the 2nd interim analysis $u(t_2)$ is calculated such that

$\alpha(t_1) + P_0(Z_1 < u(t_1), Z_2 \geq u(t_2)) = \alpha(t_2)$

(where t_2 *is* again calculated using the actual number of events observed at 2nd interim and assuming the final number of PFS events is 243).

The boundary for the final efficacy analysis is calculated such that

$\alpha(t_2) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_F \geq u_F) = \alpha = 0.02$,

where α is the cumulative alpha spent up to the final analysis.

As a practical matter it is rather unlikely that the last analysis can be performed at the precise time point that the planned maximum information is attained. If the number of events in the final analysis deviates from the expected number, the final analysis criteria will be determined using the above formula, taking into account the alpha actually spent at each of the interim analyses and the actual correlation among the three test statistics Z_1, Z_2 and Z_F, in such a way that the overall significance level across all analyses is maintained at 0.02.

In practice, the calculation of the final analysis boundaries will be implemented in EAST Version 6.3 by the use of "Interim Monitoring" sheet associated with the study design which allows entering of cumulative events at the interim analysis and at the final analysis.

4.5.4 Group sequential design with PFS as a primary endpoint, OS as a key secondary endpoint and controlling for multiplicity

This statistical testing method described below addresses multiplicity issues arising from using a group sequential design when there are multiple endpoints involved. In this study, the primary endpoint is PFS in the PIK3CA mutant cohort and the key secondary endpoint is OS in the PIK3CA mutant cohort. There are also interim analyses that allows the PIK3CA mutant cohort to stop for futility (in the first interim) and for outstanding PFS efficacy (in the second interim analysis). Therefore, there are 2 sources of multiplicity:

- Multiplicity arising due to the group sequential nature of the study design
- Multiplicity arising due to testing two endpoints PFS and OS

There is yet another complexity that arises due to the fact that the PFS events are expected to accrue faster than the OS events. Since in most situations, there will be some delay between

the PFS events and the OS events, the number of deaths expected at interim analysis as well as at the time of final PFS analysis might not be sufficient to have at least reasonable chance (i.e. statistical power) to achieve statistical significance for the key secondary endpoint. Therefore, it seems reasonable to perform (and pre-specify) the final analysis of overall survival at a later time point, when sufficient survival events have been observed. Since the final analysis for the key secondary endpoint may potentially happen after the primary analysis for PFS has been conducted, there may be up-to 3 analyses for OS.

An intuitive hierarchical testing strategy is to test for the key secondary endpoint only if the primary endpoint is found to be significant, using the same level of significance α that was used to test the primary endpoint. However, as shown [Hung et al, 2007] this hierarchical strategy when applied in a group sequential design does not control the overall type I error rate (or the family-wise type I error probability) in the strong sense. It has been demonstrated that the correlation between the two endpoints and the effect size of the primary endpoint are two nuisance parameters that determine the level of the type I error probability of falsely concluding a positive effect on the secondary endpoint.

The above multiplicity issues will be addressed by the use of a Haybittle–Peto boundary and an alpha spending function $\delta(t)$ for PFS and OS respectively (note: these are two different spending functions on two different information fraction scales), and by implementing the following testing strategy. Notations used in the Table are defined as follows:

PIK3CA Mutant Cohort

- Let $H_{0PFSmut}$ and H_{0OSmut} denote the null hypotheses for testing PFS and OS respectively in the PIK3CA mutant cohort.
- Let $\alpha_{mut}(t)$ and $\delta_{mut}(t)$ denote the alpha-spending functions for PFS and OS, respectively.
- Let $s_{1mut} < s_{2mut} < s_{3mut} < s_{4mut} < s_{5mut}$ denote the time points for:
  - First interim (futility) analysis (driven by PFS events),
  - Second interim analysis (driven by PFS events, first OS interim analysis)
  - Planned final analysis of PFS when targeted number of PFS events is expected to be observed (first OS interim analysis, if not tested at s_2)
  - Planned second interim analysis of OS.
  - Planned final analysis of OS
- Let $t_{PFS(s1mut)}$, $t_{PFS(s2mut)}$, $t_{PFS(s3mut)}$ represent information fractions for PFS at time points s_{1mut}, s_{2mut}, s_{3mut}, respectively.
- Let $t_{OS(s2mut)}$, $t_{OS(s3mut)}$, $t_{OS(s4mut)}$, $t_{OS(s5mut)}$ represent information fractions for OS at time points s_{2mut}, s_{3mut}, s_{4mut}, s_{5mut}, respectively.
- $u_{mut}(t_k)$ and $v_{mut}(t_k)$ are the efficacy stopping boundaries for PFS and OS, respectively, at information fraction t_k.
- $l_{mut}(t_{(PFS(s1mut))})$ is the futility stopping boundaries for PFS at the first interim analysis

Using the following testing strategy, the overall type-I error rate can be controlled:

1. During the first PFS interim analysis, test PFS at $\alpha_{mut}(t_{PFS(s1mut)})$:

- If the futility boundary for PFS, $l_{mut}(t_{PFS(s1mut)})$ is crossed then stop the study for futility, otherwise go to step 2.
- Note there is no intent to test OS at the first PFS interim analysis.

2. During the second PFS interim analysis, test PFS at α=0.0001:
   - If the PFS efficacy boundary $u_{mut}(t_{PFS(s2mut)})$ is crossed then test OS at $\delta_{mut}(t_{OS(s2mut)})$.
   - If the OS efficacy boundary $v_{mut}(t_{OS(s2mut)})$ is also crossed then we will stop the study for superior efficacy, and no more testing will be done for either PFS or OS.
   - If $u_{mut}(t_{PFS(s2mut)})$ is crossed but $v_{mut}(t_{OS(s2mut)})$ is not crossed then we will stop testing for PFS, however OS will be tested as planned at s_{4mut} (Step 4).
   - If $u_{mut}(t_{PFS(s2mut)})$ is not crossed then OS will not be tested during the second interim analysis; and go to step 3.
3. During the final PFS analysis, test PFS at α=0.0199:
   - If the PFS efficacy boundary $u_{mut}(t_{PFS(s3mut)})$ is crossed then test OS at $\delta_{mut}(t_{OS(s3mut)})$.
   - If the OS efficacy boundary $v_{mut}(t_{OS(s3mut)})$ is also crossed then no more testing for either PFS or OS.
   - If $u_{mut}((t_{PFS(s3mut)})$ is crossed but $v_{mut}(t_{OS(s3mut)})$ is not crossed then we will continue test OS as planned (go to step 4).
   - If $u_{mut}(t_{OS(s3mut)})$ is not crossed then OS will not be tested; and no more testing on either PFS or OS.
4. During the planned second OS interim analysis, test OS at $\delta_{mut}(t_{OS(s4mut)})$. If the OS efficacy boundary $v_{mut}(t_{PFS(s4mut)})$ is crossed then no more testing. If not then go to step 5.
5. During the planned final OS analysis, test OS at $\delta_{mut}(t_{OS(s5mut)})$. The study will be stopped at this step regardless of the outcome.

PIK3CA Non-mutant Cohort

- Let $H_{0OSnonmut}$ denote the null hypotheses for testing OS in the PIK3CA non-mutant cohort.
- Let $\delta_{nonmut}(t)$ denote the alpha-spending function for OS.
- Let $s_{1nonmut} < s_{2nonmut} < s_{3nonmut}$ denote the time points for:
  - Planned final analysis of PFS when targeted number of PFS events is expected to be observed (first OS interim analysis)
  - Planned second interim analysis of OS.
  - Planned final analysis of OS
- Let $t_{OS(s1nonmut)}$, $t_{OS(s2nonmut)}$, $t_{OS(s3nonmut)}$ represent information fractions for OS at time points $s_{1nonmut}$, $s_{2nonmut}$, $s_{3nonmut}$ respectively.
- $v_{nonmut}(t_k)$ are the efficacy stopping boundaries for OS, at information fraction t_k.

Using the following testing strategy, the overall type-I error rate can be controlled:

1. During the final PFS analysis, test PFS at 0.005:
   - If the PFS efficacy boundary is crossed then test OS at $\delta_{nonmut}(t_{OS(s1nonmut)})$.
   - If the OS efficacy boundary $v_{nonmut}(t_{OS(s1nonmut)})$ is also crossed then no more testing for OS.
   - If the PFS efficacy boundary is crossed but $v_{nonmut}(t_{OS(s1nonmut)})$ is not crossed then we will continue to test OS as planned (go to step 2).
2. During the second OS interim analysis, test OS at $\delta_{nonmut}(t_{OS(s2nonmut)})$. If the OS efficacy boundary $v_{nonmut}(t_{PFS(s2nonmut)})$ is crossed then no more testing. If not then go to step 3.
3. During the planned final OS analysis, test OS at $\delta_{nonmut}(t_{OS(s3nonmut)})$. The study will be stopped at this step regardless of the outcome.

4.6 Statistical methodology and operating characteristics – PFS futility criteria in the *PIK3CA* mutant cohort

At the time of the futility analysis, the cohort may be stopped for futility if one or both of the following criteria are met:

- The observed (i.e. nominal) p-value > 0.128
- Conditional Probability ($HR_{final} \leq 0.6 | HR_{interim}$) < 0.20

Criterion (2) uses the observed interim data and an assumption regarding the distribution of future unobserved data in the two treatment groups conditioned under the alternative hypothesis (HR_{Ha}). Under the alternative hypothesis, the following formula (Jennison and Turnbull 2000 formula 10.2) can be used to find the futility boundary on the Z-statistic scale (z_t) for criterion (2) that satisfies:

$$\Phi\left[-\frac{1}{\sqrt{1-t}}\left(z_{\alpha'}-\sqrt{t}z_t-\sqrt{1-t}\,\frac{-\log HR_{H_a}}{\sqrt{\left(\frac{a}{D(1-t)}\right)}}\right)\right] < 0.2$$

where:

$a = \frac{(r+1)^2}{r}$ with r:1 randomization ratio for treatment and control,

D: Total number of PFS events,

t: Information fraction at the futility interim analysis (i.e. 0.4)

$z_{\alpha'}$: Final boundary on Z scale,

z_t: Observed value on Z scale at futility interim

The critical value for the Z-statistic at the futility interim analysis that will ensure criterion (2) is satisfied is $|z_t| = 1.489$. The futility boundary in terms of p-value scale is thus calculated as p =0.068.

The operating characteristics of the revised futility criteria are provided in Table 4-1.

Table 4-1 Operating characteristics for PFS futility criteria in the *PIK3CA* mutant cohort

Futility Criteria	Probability to be stopped for futility	
True HR	**Under Criterion 1 only**	**Under Criterion 2**
1.0	86.95%	92.98%
0.6	8.89%	15.31%

Criterion 1: Observed (i.e. nominal) p-value greater than 0.128

Criterion 2: Conditional Probability (HRfinal ≤ 0.6¦ HRinterim) < 0.20

Note: Operating characteristics for criterion 1 performed in East 6.3 with number of simulations = 10,000 and randomization seed =37275.
Operating characteristics for criterion 2 performed in SAS v9.4 with 10,000 simulations and random seed = 111064

4.7 Statistical design and operating characteristics – PFS in the *PIK3CA* non-mutant cohort

A Bayesian double criteria-based design is used to estimate the treatment effect in the *PIK3CA* non-mutant cohort, the methodology and operating characteristics based on simulation are detailed below.

4.7.1 Bayesian methodology for proof of concept criteria

Let θ denote the natural logarithm of the hazard ratio (HR) of PFS (experimental arm vs. control, i.e. $\theta < 0$ indicates efficacy in favor of the experimental arm i.e. alpelisib + fulvestrant) and y_m denote the log(HR) estimated from a Cox proportional Hazards model with treatment as covariate based on *m* observed events and then using asymptotic theory of the log hazard ratio (Schoenfeld 1981):

$$Y_m \sim N(\theta, 4/m)$$

Further assume θ follows a conjugate normal prior distribution, written as

$$\theta \sim N(\theta_0, 4/n_0)$$

where θ_0 is the specified prior mean and the prior variance $4/n_0$, n_0 is the number of events worth of prior information.

This results in a posterior distribution of θ as

$$\theta | y_m \sim N(\phi\, y_m, + (1- \phi)\, \theta_0, 4/(m+n_0))$$

where $\phi = m/(m+n_0)$ and in this study we consider a non-informative prior with $n_0=0$.

Therefore the posterior distribution is of the following form;

$$\theta | y_m \sim N(y_m, 4/m)$$

The cumulative posterior distribution will be used to derive the probability that the true HR is less than 1.

4.7.2 Proof of concept (PoC) criteria

The following PoC criteria, based on analysis of PFS using Cox proportional Hazards model with treatment as covariate, are considered;

- Estimated HR ≤ 0.6, and
- Posterior Probability (HR < 1) $\geq 90\%$

Both criteria need to be met in order to meet primary objective for this part of the study and test PFS in this cohort using a stratified log-rank test at one-sided 0.5% level of significance. The first criterion is met if the estimated HR is 0.6 or less which is the minimum HR of clinical interest. The second criterion provides reasonable evidence that the estimated HR is better than the value of no interest (HR=1) and also guarantees a level of precision for the estimate of HR.

4.7.3 Sample size considerations and simulation details

Based on the assumption that log (HR) is normally distributed, then the minimum number of events to satisfy criterion (b) can be calculated (Schoenfeld 1981) as:

$\#\text{ events} = 4\,(z_{1-\alpha} + Z_{1-\beta})^2 / \theta^2$

Taking, $\theta = \log(0.6)$, and one-sided, $\alpha=0.005$, $\beta=0.5$, then 102 events are required. That is, with at least 102 events if the estimated HR is < 0.6 then criterion (b) in Section 4.7.2 will be met.

Assuming an enrollment rate of 10 patients during the first 6 months (7 with *PIK3CA* non-mutant status), 30 patients up to 12 months (21 with *PIK3CA* non-mutant status) and 50 patients afterwards (35 with *PIK3CA* non-mutant status) and 10% patients will be lost to follow-up, 220 patients will be randomized (110 per arm), in order to observe the required 102 events in the two arms in approximately 18 months (if the observed HR is 0.6 and the median PFS for the control arm is 7.4 months).

The primary analysis to estimate the HR will be performed after approximately 102 PFS events have been observed. If the true HR is 1, the probability (obtained by simulation) to obtain a positive conclusion is 0.005. If the true HR is 0.60 (reflecting the minimum clinically relevant difference), the probability to meet efficacy criteria is 0.491. In addition if the true HR is 0.6 and the PoC is met, the probability (by simulation) to also observe a positive result with formal testing is 0.983.

Table 4-2 Operating characteristics for PoC criteria in *PIK3CA* non-mutant cohort

True HR	True Median PFS alpelisib (months)	Probability to meet PoC	Probability formal testing is positive given PoC is met*
0.3	24.67	0.999	0.999
0.4	18.50	0.975	0.999
0.5	14.80	0.813	0.994
0.6	**12.33**	**0.491**	**0.983**
0.7	10.57	0.220	0.972
0.8	9.25	0.076	0.970
0.9	8.22	0.020	0.956
1.0	7.4	0.005	0.940

Assumes: (1) HR.=0.6, (2) true median PFS for fulvestrant = 7.4 months, (3) protocol planned accrual rates, (4) Analysis after 102 PFS events have been observed

* Probabilities conditional on PoC criteria being met. Formal testing using log-rank test at a one-sided, α=0.005

4.8 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 1000 times, giving rise to 1000 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 4-3.

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

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

Table 4-3 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 assessment (for subjects with complete agreement of local and central sources) will be considered to agree if they occur within ±7 days of each other, aligned with the protocol-specified window for tumor assessments.

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

Bergh J, Jönsson PE, Kerstin Lidbrink E,Trudeau et al (2012) FACT: An Open-Label Randomized Phase III Study of Fulvestrant and Anastrozole in Combination Compared With Anastrozole Alone As First-Line Therapy for Patients With Receptor-Positive Postmenopausal Breast Cancer. J.Clin.Oncol.; 30 (16): 1919-1925.

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.

[redacted]

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

National Cancer Institute (2010) Common Terminology Criteria for Adverse Events (CTCAE) v.4 data files and related documents:http://evs.nci.nih.gov/ftp1/CTCAE/About.html

De Buck SS, Jakab A, Boehm M, Bootle D, Juric D, Quadt C, Goggin TK (2014) Population pharmacokinetics and pharmacodynamics of BYL719, a phosphoinositide 3-kinase antagonist, in adult patients with advanced solid malignancies. Br J Clin Pharmacol; 78(3):543-555

Di Leo A, Jerusalem G, Petruzelka L, et al (2010) Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer.J Clin Oncol. Oct 20;28(30):4594-600. Erratum in: J Clin Oncol.2011 Jun 1;29(16):2293.

Di Leo A, Jerusalem G, Petruzelka L, Torres R, et al (2013) Final Overall Survival: Fulvestrant 500 mg vs 250 mg in the Randomized CONFIRM Trial. JNCI J Natl Cancer Inst 106(1).

Dodd E L, Korn L E, Freidlin B, Gray R, Bhattacharya S. An audit strategy for progression-free survival. Biometrics, Vol 67, Issue 3; 1092-99, 2011

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

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

Haybittle JL (1971). Repeated assessments of results in clinical trials of cancer treatment". British. Journal of Radiology 44 (526),: 793–797

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).

Howell A, Robertson JF, Abram P, et al (2004) Comparison of fulvestrant versus tamoxifen for the treatment of advanced breast cancer in postmenopausal women previously untreated with endocrine therapy: a multinational, double-blind, randomized trial. J Clin Oncol.; 22(9):1605- 1613.

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

Jennison C, Turnbull B (2000) Group sequential methods with applications to clinical trials. Boca Raton, Chapman & Hall/CRC

Johnston SR, Kilburn LS, Ellis P, et al (2013) SoFEA investigators. Fulvestrant plus anastrozole or placebo versus exemestane alone after progression on non-steroidal aromatase inhibitors in postmenopausal patients with hormone-receptor-positive locally advanced or metastatic breast cancer (SoFEA): a composite, multicentre, phase 3 randomised trial. Lancet Oncol.;14(10):989-998.

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

Maurer W and Bretz F (2013) Memory and other properties of multiple test procedures generated by entangled graphs. Statist. Med. 2013, 32 1739–1753.

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 (11): 2101-09.

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

[redacted]

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

Robertson J, Lindemann J, Sally Garnett S, et al (2014). A Good Drug Made Better: The Fulvestrant Dose-Response Story Clin Breast Cancer, Vol. 14, No. 6, 381-9.

Schoenfeld D (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. Biometrika 68, 316-319.