Parexel International

AstraZeneca

D8530C00002

SERENA-2: A Randomised, Open-Label, Parallel-Group, Multicentre Phase 2 Study Comparing the Efficacy and Safety of Oral AZD9833 versus Fulvestrant in Women with Advanced ER-Positive HER2-Negative Breast Cancer

Statistical Analysis Plan

Version: 5.0

Parexel Project Number: 245155

TABLE OF CONTENTS

LIST OF TABLES

REVISION HISTORY

Version No.	Effective Date	Summary of Change(s)
1.0	27 Feb 2020	New document
2.0	08 Oct 2020	• Updated terminology of unblinded team to additional team to link to the blinding maintenance plan. • Updated Section 4.6.1 (disposition of patients) to add listings and summaries on patients impacted with the COVID-19 pandemic. • Updated Section 4.6.2 (protocol deviations) to clarify that some deviations will be checked for analysis purposes will not necessarily be included in the protocol deviations log for this study. • Added to Section 4.6.2 (protocol deviations) that COVID-19 related protocol deviations will be summarised and presented separately. • Added oestrogen receptor 1 mutation at baseline to Section 4.7 (demographic and baseline characteristics). • Clarified the hotspots required for ESR1 mutation in Section 4.10.1.5 (examination of subgroups). • Updated ER percent positivity at baseline categories to <10% and ≥10% in in Section 4.10.1.5 (examination of subgroups). • Added in Section 4.10.3 (progression free survival) that Kaplan-Meier plots will be repeated on the subgroup for liver and/or lung metastasis. • Clarified in Section 4.10.4.3 (change in tumour size) that results at progression and latest RECIST assessment should be included in the determination of best percentage change in tumour size. • Added in Section 4.10.4.4 (overall survival) that proportion alive will only be presented at the 12 month and 24 month milestones. • Added in Section 4.11.1 (extent of exposure) calculations to determine exposure duration in days of fulvestrant. • Added summaries for time to onset and duration of AESIs in Section 4.11.2 (adverse events). • Added higher-level terms to be searched for visual disturbance AEs in Section 4.11.2 (adverse events). • Update CCI [redacted] terminology to CCI [redacted]. • Added S/P-Troponin and S/P-NT-proBNP to Table 6 in Section 4.11.4 (clinical laboratory) as laboratory variables to be included in summaries. These two tests were added as part of protocol version 2.0.

Version No.	Effective Date	Summary of Change(s)
		• Updated Section 4.11.6 (electrocardiograms) to add details and a summary on cardiology review items, to clarify that Holter results will also be presented and to add normal ranges for ECG parameters. • Updated Section 4.11.6 (electrocardiograms) to add categorised shift tables for heart rate and CCI . • Added Section 4.11.8 (echocardiograms) that was added as part of protocol version 2.0. • Added Section 4.11.9 (neurological assessments) that was added as part of protocol version 2.0.
3.0	05 Jul 2021	• Updated throughout based on protocol version 4.0 where recruitment to the 300 mg arm has been closed in December 2020. Clarified that no formal comparisons using the 300 mg arm will be presented but that the 300 mg arm should still be included in all models. • Clarified time window derivations in Section 4.3.2. • Corrected typo in Section 4.10.4.1 to clarify that a best objective response of NE is assigned if death occurs after 17 weeks of randomisation without a valuable RECIST assessment instead of nine weeks as originally stated. • Updated ESR1 mutation definition in Section 4.10.1.5, effectively removing L536Q and P535H. • Clarified in Section 4.10.4.1 that the analysis on CBR_{24} will be based on the subset of patients who were randomised at least 25 weeks prior to the DCO for analysis, or who had their 24 week RECIST assessment. • Added imputation rules for change from in tumour size at week 16 in Section 4.10.4.3.1. • Added COVID-19 related sensitivity analyses for progression free survival in Section 4.10.3.1 and overall survival in Section 4.10.4.4. • Added COVID-19 related adverse event summaries in Section 4.11.2. • Added CCI to ECG outlier analysis in Section 4.11.6. • Removed shift tables to be presented for echocardiogram results in Section 4.11.8. • Added figures for EQ-5D in Section 4.12.3.3.
4.0	01 Mar 2022	• Updated Section 3.2 to state the primary analysis will be conducted when at least 108 events for each pairwise comparison is available, based on updated protocol version 5.0. • Clarified in Section 4.3.2 that an End of Treatment visit for patient reported outcomes should be kept (i.e., should not be windowed).

Version No.	Effective Date	Summary of Change(s)
		• Added that the deviation bias analysis will be performed for overall survival and removed the requirement that 10% of patients need to have deviations impacting on efficacy for the deviation bias analysis to be performed in Section 4.6.2. • Added details on second interim in Section 4.10.1.4 as per updated protocol version 5.0. • Update definition of endocrine status from primary and secondary to sensitive and resistant in Section 4.10.1.5. • Clarified in Section 4.10.1.5 that a subgroup must contain at least 10 patients to be presented in outputs. • Clarified calculation of ER percent positivity in Section 4.10.1.5. • Updated race categories for subgroup analyses to white and other in Section 4.10.1.5. This was done to ensure large enough groups for statistical analyses. • Added subgroup analyses for ESR1 mutation and endocrine treatment status to Section 4.10.3. • Added weighted log-rank test to Section 4.10.3.2. • Added to CCI [redacted] that we expect to see increased survival with increasing dose in Section 4.10.3.2. • Added subgroup analyses for ORR in Section 4.10.4.1. • Clarified in Section 4.10.4.1 that CBR_{24} should be calculated relative to randomisation and not relative to treatment. • Updated preferred terms to search for in identifying CCI [redacted] adverse events in Section 4.11.2. • Clarified normal albumin in Section 4.11.4. • Added conversion for differentials from absolute to percentages and percentages to absolute in Section 4.11.4. • Added a statement in Section 4.12.2, that only listings will be presented if there is an inadequate amount of data available for summaries. • Updated sample size information in Section 4.13 based on updated protocol version 5.0.
5.0	See footnote	• Updated endocrine sensitivity definitions in Section 4.10.1.5. • Added ESR1 mutation subgroup analyses for PFS in Section 4.10.3.

Version No.	Effective Date	Summary of Change(s)
		• Update subset of patients for CBR_{24} in Section 4.10.4.1 to only include patients randomised at least 25 weeks prior to the DCO for the analysis.

LIST OF ABBREVIATIONS

Abbreviation / Acronym	Definition / Expansion
AE	adverse event
AESI	adverse event of special interest
ALT	alanine transaminase
ANCOVA	analysis of covariance
AST	aspartate transaminase
ATC	Anatomical Therapeutic Chemical
AZ	AstraZeneca
BICR	Blinded Independent Central Review
BMI	body mass index
BoR	best objective response
BSR	baseline scaled ratio
CBR_{24}	clinical benefit rate at 24 weeks
CDK4/6	cyclin-dependent kinase 4 and 6
CI	confidence interval
COVID-19	coronavirus disease 2019
CSP	Clinical Study Protocol
CR	complete response
CSR	clinical study report
CT	computed tomography
CCI	[redacted]
CTCAE	Common Terminology Criteria for Adverse Events
ctDNA	circulating tumour deoxyribonucleic acid
CV	coefficient of variation
DCO	data cut-off
DMC	data monitoring committee
DNA	deoxyribonucleic acid
DoR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EORTC	European Organisation for Research and Treatment of Cancer
EORTC QLQ-BR23	EORTC quality of life questionnaire - breast cancer module
EORTC QLQ-C30	EORTC quality of life questionnaire - core questionnaire
EQ-5D-5L	EuroQol 5-Dimension 5-level
ER	oestrogen receptor
ESR1	oestrogen receptor 1
FAS	Full Analysis Set
FFPE	formalin fixed paraffin embedded

Abbreviation / Acronym	Definition / Expansion
gmean	geometric mean
HER2	human epidermal growth receptor 2
HR	hazard ratio
HRQoL	health-related quality of life
CCI [redacted]	CCI [redacted]
IHC	immuno-histochemistry
IM	intramuscular
IWRS	interactive web response system
LD	longest diameter
LLoQ	lower limit of quantification
LSMeans	least square means
LVED	left ventricular end-diastolic
LVEF	left ventricular ejection fraction
LVES	left ventricular end-systolic
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model for repeated measures
MRI	magnetic resonance imaging
NA	not applicable
NC	not calculable
NE	not evaluable
NEI VFQ-25	National Eye Institute 25-Item Visual Function Questionnaire
NQ	non-quantifiable
NTL	non-target lesions
OAE	other significant adverse event
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
CCI [redacted]	[redacted]
PgR	progesterone receptor
PK	pharmacokinetic
PO	oral
PR	partial response
PRO	patient reported outcome
PT	preferred term
CCI [redacted]	[redacted]

Abbreviation / Acronym	Definition / Expansion
CCI	[redacted]
RDI	relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumours
REML	restricted maximum likelihood
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SOC	system organ class
SRC	Safety Review Committee
StD	standard deviation
TEAE	treatment emergent adverse event
TL	target lesion
TNM	tumour, node and metastasis
TTR	time to response
ULN	upper limit of normal
ULoQ	upper limit of quantification
WHO	World Health Organisation
WHODD	World Health Organisation Drug Dictionary

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical elaboration of the statistical analysis of efficacy, safety, pharmacokinetic (PK), pharmacodynamic, and health-related quality of life (HRQoL) data.

The analyses described in this SAP are based upon the following documents:

- Clinical Study Protocol, Version 5.0 (September 15, 2021)

Specifications for tables, figures, and listings are contained in a separate document.

2 STUDY OBJECTIVES AND ENDPOINTS

2.1 Primary Objective

Primary Objective:	Endpoint/Variable:
To determine the clinical efficacy (as assessed by progression-free survival [PFS]) of AZD9833 when compared to fulvestrant in women with advanced oestrogen receptor (ER)-positive human epidermal growth receptor 2 (HER2)-negative breast cancer	PFS assessed by the investigator as defined by Response Evaluation Criteria in Solid Tumours (RECIST) version 1.1

2.2 Secondary Objectives

Secondary Objectives:	Endpoint/Variable:
To determine anti-tumour effect of AZD9833 when compared to fulvestrant in women with advanced ER-positive HER2-negative breast cancer	Based on tumour response assessed by the investigator, as defined by RECIST version 1.1: • Objective response rate (ORR) • Duration of response (DoR) • Best percentage change in tumour size and percentage change in tumour size at 16 weeks
To determine the effect of AZD9833 on survival and clinical benefit when compared to fulvestrant in women with advanced ER-positive HER2-negative breast cancer	• Overall survival (OS) • Clinical benefit rate at 24 weeks (CBR_{24})
To evaluate the PK of AZD9833 in this patient population at steady state	Plasma concentrations of AZD9833 and, if appropriate, metabolite(s) on Day 15 (pre- and post-dose) and Day 29 (pre-dose)

Secondary Objectives:	Endpoint/Variable:
To evaluate the pharmacodynamics of AZD9833 and fulvestrant in a subgroup of patients with advanced ER-positive HER2-negative breast cancer	• Percent change from baseline ER and progesterone receptor (PgR) expression assessed by the manual H-score method • Percent change from baseline in Ki67 labelling index
To evaluate the effect of AZD9833 and fulvestrant on the patients' HRQoL, as assessed by patient-completed HRQoL questionnaires	Changes from baseline in scores of the European Organisation for Research and Treatment of Cancer (EORTC) quality of life questionnaire – core questionnaire (QLQ-C30), EORTC quality of life questionnaire – breast cancer module (QLQ-BR23), National Eye Institute 25-Item Visual Function Questionnaire (NEI VFQ-25), and EuroQol 5-Dimension 5-level (EQ-5D-5L)

2.3 Safety Objective

Safety Objective:	Endpoint/Variable:
To evaluate the safety and tolerability of AZD9833 when compared to fulvestrant in women with advanced ER-positive HER2-negative breast cancer	• Adverse events (AEs)/ serious AEs (SAEs) • Vital signs, electrocardiograms (ECGs), clinical chemistry, haematology, urinalysis parameters

2.4 Exploratory Objectives

Exploratory Objectives:	Endpoint/Variable:
To investigate predictive markers of response and/or acquired resistance to AZD9833 and fulvestrant	Change from baseline in amount and [redacted] circulating tumour deoxyribonucleic acid (ctDNA), [redacted]
To assess the [redacted]	• Subgroup analysis of [redacted] during treatment

Exploratory Objectives:	Endpoint/Variable:
To evaluate the effect of AZD9833 and fulvestrant on the CCI [redacted]	• CCI [redacted] • CCI [redacted] • CCI [redacted]
To examine overall CCI [redacted], CCI [redacted]	CCI [redacted]
Optional CCI [redacted]: To collect and store CCI [redacted] according to each country's local and ethical procedures for future exploratory research into CCI [redacted] that may influence response to treatment	Results from future exploratory research may be reported outside of the clinical study report (CSR).

With the exception of CCI [redacted], CCI [redacted], CCI [redacted], CCI [redacted] CCI [redacted], all exploratory endpoints will be reported outside of the CSR and will therefore not be described further within this SAP.

Population PK analysis and exposure response modelling in support of pharmacodynamic, safety and efficacy endpoints may be performed. The planned analyses will be documented in a separate modelling analysis protocol if required and results will be reported outside of the CSR.

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

This is a randomised, open-label, parallel-group, multicentre Phase 2 study to compare the efficacy and safety of daily oral (PO) AZD9833 versus intramuscular (IM) fulvestrant in women with advanced ER-positive HER2-negative breast cancer. Post-menopausal women with histologically or cytologically confirmed metastatic or loco-regionally recurrent disease before randomisation and fulfilling all the inclusion criteria and none of the exclusion criteria will be included. Randomisation will be stratified according to the prior use of cyclin-dependent kinase 4 and 6 (CDK4/6) inhibitors and the presence of liver and/or lung metastases.

After the screening visit and confirmation of eligibility, approximately 288 patients will be randomly assigned in a 1:1:1:1 ratio to receive one of the following four treatments, or from

December 2020 in a 1:1:1 ratio to the remaining three treatments, consisting of 4-week treatment cycles until disease progression (assessed by the investigator as defined by RECIST version 1.1):

- AZD9833 (75 mg, PO, once daily)
- AZD9833 (150 mg, PO, once daily)
- AZD9833 (300 mg, PO, once daily) (closed to recruitment as of December 2020)
- Fulvestrant (500 mg, IM, Day 1, Day 15, Day 29, and 4-weekly thereafter)

As of December 2020, the Sponsor stopped enrolment to the 300 mg treatment arm. Ongoing patients in the 300 mg treatment arm may continue treatment as planned.

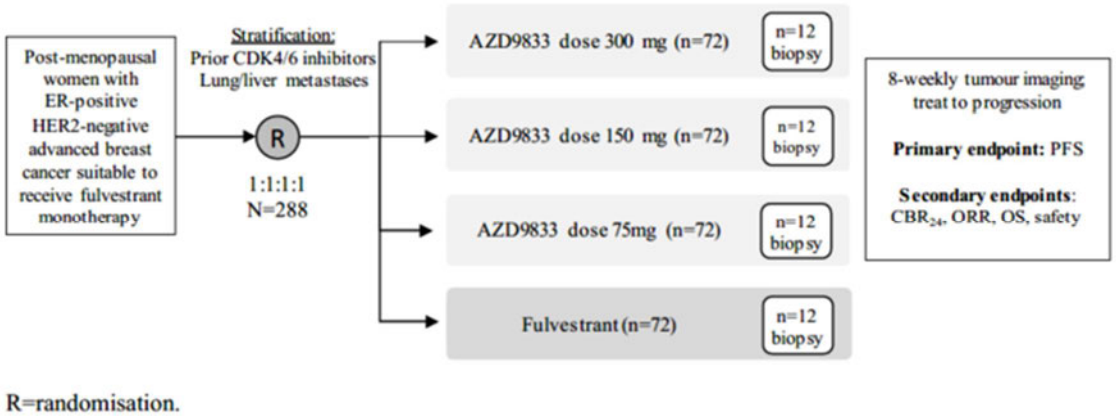
During the treatment period patients will attend visits on:

- Day 1, Day 8, and Day 15 of Cycle 1
- Day 1 of each subsequent cycle until treatment discontinuation

After the treatment period, patients will attend two safety follow-up visits (at the time of treatment discontinuation and 28 days later) and will continue to be followed for survival and progression if they haven't progressed.

The general study design is summarised in Figure 1.

Figure 1 Study Design



Note: As of December 2020, the Sponsor stopped enrolment to the 300 mg treatment arm. Ongoing patients in the 300 mg treatment arm my continue treatment as planned.

3.2 Planned Analyses

All AstraZeneca (AZ) and CCI team members (except the CCI additional team) involved in the study will remain blinded to the aggregate efficacy data until after the database lock for the primary analysis. Following database lock for the primary analysis, all study team members will become unblinded.

Table 1 details the analyses planned during the conduct of this study at each timepoint.

Table 1 Summary of Analyses During Study Conduct

Analysis	Trigger/Timepoint	Data included
Rolling safety update	Approximately every two months up to primary analysis for PFS.	Safety data
Safety Review Committee (SRC) analyses	Approximately every six months up to primary analysis for PFS and thereafter at the discretion of the SRC. Other safety reviews may also be conducted as requested by the SRC.	Safety data
Interim analysis 1	At least [redacted] events for each of the pairwise treatment comparisons between each dose of interest of AZD9833 (75 mg or 150 mg) and fulvestrant	Efficacy data (including PFS, ORR, and best objective response [BoR]). Safety data may be included, depending on how recently the last SRC has been conducted. Safety data will be presented as meeting minutes and conclusions from the last SRC meeting. Full details are included in the data monitoring committee (DMC) charter. Following the decision to stop enrolment to the 300 mg AZD9833 treatment arm, data from this arm will be reported together with the remaining treatment arms.

Analysis	Trigger/Timepoint	Data included
Interim Analysis 2	At least CCI events for each of the pairwise treatment comparisons between each dose of interest of AZD9833 (75 mg or 150 mg) and fulvestrant	Efficacy data (including PFS, ORR, and BoR). Safety data may be included, depending on how recently the last SRC has been conducted. Safety data will be presented as meeting minutes and conclusions from the last SRC meeting. Full details are included in the data monitoring committee (DMC) charter. Following the decision to stop enrolment to the 300 mg AZD9833 treatment arm, data from this arm will be reported together with the remaining treatment arms.
PFS primary analysis (approximately 75% maturity in the best performing dose group)	At least 108 events for each of the pairwise treatment comparisons between each dose of interest of AZD9833 (75 mg and 150 mg) and fulvestrant	All data (including summary OS data available at the time of primary analysis) Following the decision to stop enrolment to the 300 mg AZD9833 treatment arm, data from this arm will be reported together with the remaining treatment arms.
Survival follow-up analysis [a] (CCI and CCI maturity)	Further survival analysis may be conducted at CCI and CCI OS maturity (i.e., once CCI and CCI patients have died respectively)	OS[b] Following the decision to stop enrolment to the 300 mg AZD9833 treatment arm, data from this arm will be reported together with the remaining treatment arms.
Safety update	At clinical database closure	Key safety data (including SAEs)

BoR = best objective response; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; SAEs = serious adverse events; SRC = Safety Review Committee.
[a] The PFS and OS analysis may be combined if there is sufficient maturity in the OS data at the time of the primary PFS analysis.

[b] The safety update may be combined with the OS follow-up analysis if the decision is made to close the study database at this time. A PFS analysis may also be conducted at this time.

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency, integrity and in accordance with standard Parexel procedures.

4.2 General Presentation Considerations

Continuous data will be summarised using descriptive statistics (number of observations, mean, standard deviation [StD], median, 25th and 75th percentiles [where appropriate], minimum and maximum). For log-transformed data it is more appropriate to present geometric mean (gmean), geometric coefficient of variation (CV), median, minimum and maximum. Frequencies and percentages will be used for summarising categorical (discrete) data.

For continuous data, the mean, median and gmean will be rounded to one additional decimal place compared to the original data. The StD and geometric CV will be rounded to two additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data. The maximum number of decimal places reported will be four for any summary statistic. For categorical data, percentages will be rounded to one decimal place. Percentages will not be presented for zero counts. Unless otherwise stated, percentages will be calculated using the number of patients included in the analysis set for that treatment group as denominator.

Confidence intervals (CIs) and p-values, when presented, will generally be constructed at the 2-sided 90% level. CIs will be presented to one additional decimal place compared to the original data. P-values will be presented to three decimal places. P-values less than 0.001 will be presented as "<0.001".

Unless otherwise stated, summaries will be presented by treatment. The treatment comparisons of interest are: AZD9833 75 mg vs fulvestrant and AZD9833 150 mg vs fulvestrant. As enrolment to the 300 mg treatment arm has been stopped in December 2020, no formal comparisons of AZD9833 300 mg vs fulvestrant will be performed. Data from the 300 mg treatment arm will still be included in descriptive summaries. No formal comparisons between the different AZD9833 doses will be conducted. AZD9833 300 mg should still be included in all statistical models and CIs presented where applicable. As no formal statistical analyses are conducted on the AZD9833 300 mg group, no p-values for this group will be presented.

In general, for efficacy endpoints the last observed measurement prior to randomisation will be considered the baseline measurement. For safety, PD, PK, and HRQoL endpoints the last observation prior to the first dose of study treatment will be considered the baseline measurement unless otherwise specified.

For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as evidence that the assessment occurred prior to first dose. Assessments on the day of first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to first dose if such procedures are required by the protocol to be conducted before first dose.

In all summaries, change from baseline will be calculated as the post-treatment value minus the baseline value. Percent change from baseline will be calculated as:

$$\frac{\text{post-baseline value - baseline value}}{\text{baseline value}} \times 100.$$

For any variable subject to log transformation, the back transformed change from baseline, calculated and summarised on the log scale, will be presented as 'baseline scaled ratio' (BSR). Percentage change will then be calculated as (BSR – 1) x 100.

4.3 General Variables

4.3.1 Study Day Definitions

For safety variables, study day 1 is defined as the date of first dose of study treatment (Cycle 1 Day 1).

For visits (or events) that occur on or after first dose of study treatment, study day is defined as (date of visit [event] – date of first dose of study treatment + 1). For visits (or events) that occur prior to first dose, study day is defined as (date of visit [event] – date of first dose of study treatment). There is no study day 0 defined for this study.

For listings (such as for AEs) that include the derivation of "days since last dose", this is defined as (event date – date of last dose) where "date of last dose" is defined as the date of dosing immediately preceding the event occurrence. Events that occur on the same day as the last dose of study treatment will therefore be described as occurring zero days from last dose of study treatment.

4.3.2 Time Windows

Time windows will be defined for any presentations that summarise values by visit. The following conventions will apply:

- The time windows will be exhaustive so that data recorded at any timepoint has the potential to be summarised. Inclusion within the time window will be based on the actual date and not the intended date of the visit.
- All unscheduled visit data will have the potential to be included in the summaries.
- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls halfway between the two visits (the lower limit of the first

post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits, then the upper limit will be taken as the midpoint value minus 1 day. For example, the visit windows for vital signs up to Day 85 for this study are:

- Day 8, visit window 2 – 11
- Day 15, visit window 12 – 21
- Day 29, visit window 22 – 42
- Day 57, visit window 43 – 70
- Day 85, visit window 71 – 98

- Windowing will be done separately for each assessment based on the schedule of events specific to that assessment.
- Should study day be missing (due to partial or missing dates), then visit will be assigned to the nominal visit at which the assessment was recorded, and no windowing will be performed.
- Except for patient reported outcome measures, visit windowing will be conducted up to and including the end of treatment visit. That is, the end of treatment visit will be reassigned a scheduled visit based on the study day the end of treatment visit occurred at. For patient reported outcome measures, the end of treatment visit will be analysed as a separate visit and will not be windowed.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).
- Listings will display all values contributing to a timepoint for a patient.

For visit-based summaries:

- If there is more than one value per patient within a time window then the closest value to the scheduled visit date will be summarised, or the earlier, in the event the values are equidistant from the nominal visit. The listings will highlight the value for the patient that contributed to the summary table, wherever feasible. In summaries of extreme values all post-baseline values collected are used including those collected at unscheduled visits regardless of whether the value is closest to the scheduled visit date, or not.
- To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group, visit data will only be summarised if the number of observations is greater than five in any treatment group for that visit.

4.3.3 Handling of Missing Data

In general, other than for the below described, or where otherwise specified in the particular analysis, missing data will not be imputed and will be treated as missing.

4.3.3.1 Imputations of Partial Dates

Concomitant medication and adverse events start dates

- Missing day: impute with the 1st of the month, unless month and year are the same as month and year of first dose of study treatment, then impute with first dose date.
- Missing day and month: impute with the 1st of January unless the year is the same as the year of first dose of study treatment, then impute with first dose date.

- Completely missing: impute with date of first dose of study treatment, unless the end date suggests it could have started prior to this in which case impute with 1st January of the same year as the end date.

When imputing a start date care should be taken to ensure the start date is sensible, i.e., prior to the end date.

Concomitant medication and adverse events end dates

- Missing day: impute with the last day of the month, unless month and year are the same as month and year of last dose of study treatment, then impute with the last treatment date in that month.
- Missing day and month: impute with the 31st of December unless the year is the same as the year of last dose of study treatment, then impute with the last treatment date in that year.
- Completely missing: assume the event is still ongoing and do not impute any date if it is flagged as "ongoing".

Generally, the imputation of dates is used to decide if an observation is treatment emergent for AEs or concomitant for medications. Flags will be retained in the database indicating where any programmatic imputation has been applied, and in such cases, durations and study days will not be calculated.

4.3.4 Imputation Rules for Laboratory Values Outside of Quantification Range

Values of the form "< x" (i.e., below the lower limit of quantification [LLoQ]) or "> x" (i.e., above the upper limit of quantification [ULoQ]) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings.

4.4 Software

All report outputs will be produced using SAS® version 9.4 in a secure and validated environment.

4.5 Analysis Sets

All patients that sign the informed consent form will be included in the **Enrolled Analysis Set**.

The efficacy and HRQoL summaries and analyses will be based on the **Full Analysis Set (FAS)**, which is based upon the Intention-to-Treat principle, unless otherwise stated for a specific analysis. The FAS is defined as all randomised patients, with treatment groups assigned in accordance with the randomisation, regardless of the actual treatment received. Patients who are randomised, but do not subsequently receive treatment will be included in the FAS. If a patient is allocated the incorrect study treatment as per the study randomisation list, patients will be summarised and analysed 'as randomised' i.e., by randomised treatment group. If a patient is stratified incorrectly, 'randomised stratum' will be used rather than 'actual stratum'.

The safety summaries will be based on the **Safety Analysis Set**. The Safety Analysis Set is defined as all patients who received any amount of study treatment (AZD9833 or fulvestrant), regardless

of whether that was the randomised therapy intended or whether they received therapy without being randomised. If a patient is allocated the incorrect study treatment as per the study randomisation list, patients will be summarised 'as treated' i.e., patients randomised to AZD9833 who received only fulvestrant will be accounted for in the fulvestrant treatment group and patients randomised to fulvestrant who received AZD9833 at the randomisation visit in error, or patients randomised to AZD9833 who received an incorrect AZD9833 dose at the randomisation visit will be accounted for in the treatment group corresponding to the first AZD9833 dose they received. Safety data will not be formally analysed.

PK summaries will be based on the **PK Analysis Set**. The PK Analysis Set is defined as all patients who received at least one dose of AZD9833 per protocol, for whom there is at least one reportable PK concentration. If a patient is allocated the incorrect study treatment as per the study randomisation list, patients will be summarised 'as treated' i.e., patients randomised to AZD9833 who received only fulvestrant will be excluded from the PK Analysis Set and patients randomised to fulvestrant who received AZD9833 at the randomisation visit in error, or patients randomised to AZD9833 who received an incorrect AZD9833 dose at the randomisation visit, will be accounted for in the treatment group corresponding to the first AZD9833 dose they received.

Pharmacodynamic summaries and analyses will be based on the **Pharmacodynamic Analysis Set**. The Pharmacodynamic Analysis Set is defined as:

- for patients who received AZD9833:
  - received the last dose of AZD9833 within 24 hours of on-treatment biopsy
  - where last dose of AZD9833 is based on a minimum of 3 days continuous dosing
- for patients who received fulvestrant:
  - received at least 2 doses of fulvestrant
- had evaluable paired tumour samples[a] by central pathology assessment; and
- had no major protocol deviations that impacted the biomarkers analysis.

[a] A pair, in order to be considered as evaluable (or adequate) and to be included in the analysis set must meet the following criteria:

- Must contain tumour, defined as > 100 tumour cells in both pre- and on-treatment formalin fixed paraffin embedded (FFPE) samples and pas the H&E qualitative assessment as above
- A minimum of 10 (preferably 20) slides of freshly prepared unstained 5 micron sections from the pre-treatment FFPE tumour block should be provided if no block will be provided. For the on-treatment samples FFPE blocks must be provided. If a sample is considered as non-evaluable as a result of this evaluation, no immune-histochemistry (IHC) assessment will be performed.

For the Pharmacodynamic Analysis Set, if a patient is allocated the incorrect study treatment as per the study randomisation list, patients will be summarised 'as treated', i.e., patients randomised to AZD9833 who received only fulvestrant will be accounted for in the fulvestrant treatment group and patients randomised to fulvestrant who received AZD9833 at the randomisation visit in error, or patients randomised to AZD9833 who received an incorrect AZD9833 dose at the randomisation

visit, will be accounted for in the treatment group corresponding to the first AZD9833 dose they received.

Upon database release, protocol deviation and analysis set outputs will be produced and will be sent to AZ for review. Prior to database lock for the primary PFS analysis, an analysis set classification meeting will be arranged to discuss the outputs and to decide which patients and/or patient data will be excluded from certain analyses. Decisions made regarding the exclusion of patients and/or patient data from analyses will be made prior to database hard lock for the primary analysis and will be documented and approved by AZ.

A summary on which analysis set will be used for each outcome variable is provided in Table 2.

Table 2 Summary of Outcome Variables and Analysis Sets

Outcome Variable	Analysis Set
Efficacy Data	
PFS, OS, CBR_{24}, ORR [a], DoR [b], best percentage change in tumour size, percentage change in tumour size at 16 weeks	FAS
Health Related Quality of Life Data	
EORTC QLQ-C30, EORTC QLQ-BR23, NEI VFQ-25, EQ-5D-5L, CCI [redacted]	FAS
Study Population/Demography Data	
Disposition of patients	Enrolled
Demography characteristics	FAS
Baseline and disease characteristics	FAS
Important protocol deviations	FAS
Medical/surgical history	FAS
Previous anti-cancer therapy	FAS
Concomitant medications/procedures	FAS
Subsequent anti-cancer therapy	FAS
PK Data	
PK concentrations	PK
Pharmacodynamic Data	
ER, PgR, Ki67	Pharmacodynamic
Safety Data	
Exposure	Safety
AEs	Safety
Laboratory measurements	Safety
Vital signs	Safety
ECGs	Safety

CRB_{24} = clinical benefit rate at 24 weeks; DoR = duration of response; ER = oestrogen receptor; EORTC = European Organisation for Research and Treatment of Cancer; EORTC QLQ-C30 = EORTC quality of life questionnaire – core questionnaire; EORTC QLQ-BR23 = EORTC quality of life questionnaire – breast cancer module; EQ-5D-5L = EuroQol 5-Dimension 5-level; FAS = Full Analysis Set; NEI VFQ-25 = National Eye Institute 25-Item Visual Function Questionnaire; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; CCI [redacted]; CCI [redacted]

PK = pharmacokinetic.
[a] Patients who are evaluable for the analysis of ORR are those with measurable disease at baseline.
[b] Patients who are evaluable for the analysis of DoR are those who responded in the ORR analysis.

4.6 Study Patients

4.6.1 Disposition of Patients

A clear accounting of the disposition of all patients who enter the study will be provided, from screening to study completion.

Patient disposition will be listed and summarised for the Enrolled Analysis Set. Summaries will include the number and percentage of patients:

- Enrolled (informed consent received)
- Randomised
- Treated
- Ongoing in study at data cut-off (DCO)
- Ongoing treatment at DCO
- Included in each analysis set (refer to Section 4.5)

In addition, the number and percentage of patients who discontinued treatment and who discontinued the study, including a breakdown of the primary reason for discontinuation will be presented for all patients.

The number of patients recruited in each country and each centre will be presented by treatment group and total.

Additional listings and summaries on patients affected by the coronavirus disease 2019 (COVID-19) pandemic will also be presented.

4.6.2 Protocol Deviations

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

The following general protocol deviation categories will be programmatically derived from the electronic case report form (eCRF) data. These deviations will be reviewed and assessed on a case-by-case basis by AZ to determine importance. Deviations considered to be important will be listed and discussed in the CSR as appropriate. All decisions on importance will be made ahead of database lock for the primary PFS analysis and will be documented prior to the primary analysis being conducted.

- Patients who deviate from entry criteria per the Clinical Study Protocol (CSP) (Deviation 1).
- Baseline RECIST scan > 42 days before randomisation (Deviation 2). Note that the 42 days is based upon a 28-day screening period plus two weeks allowance so that only serious violators are identified.
- No baseline RECIST version 1.1 assessment on or before date of randomisation (Deviation 3).
- Received prohibited medications (including other anti-cancer agents) (Deviation 4). Refer to the CSP Section 6.5 for those medications that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to database lock.
- Patients randomised who received their randomised treatment at an incorrect dose or received an alternative study treatment to that which they were randomised to (Deviation 5).

The following will be checked for analysis purposes and will not be included as part of the deviations to be recorded in the protocol deviations log during the study:

- Patients randomized but who did not receive any study treatment (Deviation 6).
- Received AZD9833 and had an on-treatment biopsy more than 24 hours after last AZD9833 dose (Deviation 7).
- Received AZD9833 and had an on-treatment biopsy without a minimum of 3 consecutive dosing days (Deviation 8).
- Pre- or on-treatment FFPE samples show ≤ 100 tumour cells (Deviation 9).
- Did not pass the H&E qualitative assessment (Deviation 10).
- Less than 10 slides of freshly prepared unstained 5 micron sections from pre-treatment FFPE tumour block provided (Deviation 11).
- No on-treatment FFPE blocks provided (Deviation 12).
- Patient vomits on the day of PK sampling and within 8 hours of AZD9833 dosing (Deviation 13). These will be reviewed on a case-by-case basis to determine whether any exclusion from the PK analysis set is deemed necessary.

Patients who received the wrong treatment at any time will be included in the analysis set as described in Section 4.5. During the study, decisions on how to handle errors in treatment dispensing (with regards to continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader and/or statistician.

The important protocol deviations will be listed and summarised by randomised treatment group. Important protocol deviations related to COVID-19 will be summarised separately. Deviation 6 will lead to exclusion from the Safety Analysis Set. None of the other deviations will lead to patients being excluded from the analysis sets described in Section 4.5 (with the exception of the Pharmacodynamic Analysis Set, if the deviation is considered to impact upon pharmacodynamic data). A per-protocol analysis excluding patients with specific important protocol deviations is not planned, however a "deviation bias" sensitivity analysis may be performed on the PFS and OS

endpoints excluding patients with deviations that may affect the efficacy of the trial therapy if patients in any treatment group:

- Did not have the intended disease or indication or
- Did not receive any randomised therapy
- Had any other significant deviation deemed to affect the primary endpoint

The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of database lock for the primary PFS analysis and will be documented prior to the primary analysis being conducted.

In addition to the programmatic determination of deviations above, other study deviations captured from the eCRF module for inclusion/exclusion criteria will be tabulated and listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

A full list of protocol deviations can be found in the study-specific Protocol Deviation Specification.

4.7 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarised for all patients in the FAS by treatment group:

- Demographics (age [years], age group [< 50; ≥ 50-< 65; ≥ 65-< 75; ≥ 75], sex [only female subjects are expected in this study], race and ethnicity)
- Patient characteristics at baseline (height [cm], weight [kg], and body mass index [BMI] [kg/m^2])
- Previous disease-related treatment modalities
- Disease characteristics at baseline (Easter Cooperative Oncology Group [ECOG] performance status, primary tumour location, histology type, tumour grade, overall disease classification, receptor status [ER/PR/HER2], and oestrogen receptor 1 [ESR1] mutation [refer to Section 4.10.1.5 for further details])
- Extent of disease upon entry (metastatic or locally advanced disease)
- Tumour, node and metastasis (TNM) classification at baseline
- Prior use of CDK4/6 inhibitors (yes/no)
- Presence of liver and/or lung metastases (yes/no)

4.8 Medical/Surgical History

Disease related medical history and relevant surgical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA). All disease related medical history (past and current) will be listed and the number and percentage of patients with any disease related medical history will be summarised for the FAS by system organ class (SOC) and preferred term (PT).

All relevant surgical history will be listed and summarised similarly.

4.9 Concomitant and Other Treatments

Information on any treatment that the patient is receiving at the time of enrolment and all concomitant treatments given up to 28 days after discontinuation of study treatment, or objective disease progression (whichever is later), with reasons for the treatment, will be recorded in the eCRF. Thereafter only subsequent regimens of anti-cancer therapy will be recorded in eCRF.

See CSP Section 6.5.5 for allowable anti-cancer therapies. Other anti-cancer therapies, investigational agents, and radiotherapy should not be given while the patient is on study treatment.

Treatments received prior to, concomitantly, or post-treatment will be coded using the Word Health Organisation (WHO) Drug Dictionary (WHODD) Anatomical Therapeutic Chemical (ATC) classification codes. Concomitant medications will be summarised for the FAS by ATC classification codes.

For the purpose of inclusion in prior and/or concomitant medication or therapy summaries, incomplete medication or radiotherapy start and stop dates will be imputed as detailed in Section 4.3.3.1.

Prior medications, concomitant and post-randomised treatment medications are defined based on imputed start and stop dates as follows:

- Prior medications are those taken prior to study treatment with a stop date prior to first dose of study treatment.
- Concomitant medications are those with a stop date on or after the first dose date of study treatment (and could have started prior to or during treatment).
- Post treatment medications are those with a start date after the last dose date of study treatment.

The following summaries will be produced:

- Previous cancer therapies prior to this study
- Disallowed concomitant medications (as identified during physician review described in Section 4.6.2, deviation 4)
- Allowed concomitant medications (to include concomitant bisphosphonate/denosumab therapies)
- Concomitant radiotherapy
- Post-discontinuation cancer therapy

All prior, concomitant and post study treatment medication data (including surgical procedures) will be listed.

Missing coding terms should be listed and summarised as “Not coded”.

4.10 Efficacy Evaluation

4.10.1 Analysis and Data Conventions

The null hypothesis for the treatment comparisons in this study will be that there is no treatment effect (i.e., there is no difference in PFS between patients treated with any dose of interest [75 mg or 150 mg] of AZD9833 and patients treated with fulvestrant). The alternative hypothesis will be that there is a difference. Symbolically, this is expressed as follows:

H_0: PFS hazard ratio $(HR)_{AZD9833/fulvestrant} = 1$
H_1: PFS $HR_{AZD9833/fulvestrant} \neq 1$

A two-sided log-rank test adjusting for prior use of CDK4/6 inhibitors and presence of lung and/or liver metastasis with α=0.1 will be used to test this hypothesis.

AZD9833 75 mg and AZD9833 150 mg will be compared with fulvestrant in a pairwise comparison. No formal comparisons of AZD9833 300 mg against fulvestrant will be performed. Data from the AZD9833 300 mg arm will be summarised and presented where appropriate. It should be noted that AZD9833 300 mg should still be included in all models and where appropriate CIs on AZD9833 300 mg may be presented. No p-values based on AZD9833 300 mg comparisons will be presented.

4.10.1.1 Multi-centre Studies

No adjustments for centre will be performed in this study.

4.10.1.2 Adjustments for Covariates

The primary efficacy analysis will be adjusted for the following stratification factors:

1. Prior use of CDK4/6 inhibitors (yes/no)
2. Presence of lung and/or liver metastasis (yes/no)

The stratification variables in the statistical modelling will be based on the values entered into the interactive web response system (IWRS) at randomisation, even if it is subsequently discovered that these values were incorrect. If considered necessary, a sensitivity analysis may be conducted based on the correct assignment.

4.10.1.3 Multiple Comparisons/Multiplicity

As this is a Phase 2 study, there will be no formal adjustment for multiplicity. See Section 4.10.3.2 for an exploratory analysis of the primary outcome which adjusts for multiplicity.

4.10.1.4 Interim Analysis

An independent AZ DMC (iAZDMC) will review the output from the interim analyses for this study.

Two interim analyses will be conducted at approximately CCI PFS maturity and CCI PFS maturity in the best performing dose group. These interim analyses will only be triggered when the treatment pair (AZD9833 75 mg vs fulvestrant or AZD9833 150 mg vs fulvestrant) with the fewest events hits CCI events for interim analysis 1 and CCI events for interim analysis 2. At the interim analyses, AZD9833 75 mg and AZD9833 150 mg will be compared against fulvestrant to assess futility and efficacy. These dose levels will be tested to assess whether the dose passes the futility hurdle and then whether the dose meets the pre-defined acceleration criteria. All boundaries will be specified in the DMC charter. Data from the AZD9833 300 mg arm will be summarised and presented where appropriate (including hazard ratios, odds ratios, and confidence intervals) without being formally compared against fulvestrant (i.e., no p-values will be presented).

The data will be reviewed by the iAZDMC, who will make a recommendation based on pre-defined decision criteria. The iAZDMC may include a recommendation to continue the study as planned, stop an AZD9833 dose level (or dose levels) for futility, or for future planning activities. The clinical project team within AZ will remain blinded to the aggregate efficacy data and will only be informed of the final decision of the iAZDMC, should there be a requirement to change study conduct.

Further details of the iAZDMC remit will be provided in a separate DMC charter.

Descriptive summaries and the analysis methods for the outputs required for the interim analysis will follow the methodology outlined in this SAP.

If required, further exploratory supportive analyses (e.g., ctDNA data) may be produced to support interpretation of the interim analyses. Further details of such supportive analysis will not be described in this SAP and will be outlined in an exploratory analysis plan.

4.10.1.5 Examination of Subgroups

The following subgroup analyses will be performed for PFS (the primary efficacy variable) by comparing PFS between treatments (for each dose of AZD9833 versus fulvestrant) in the following groups:

- Age at screening (< 65 years versus ≥ 65 years)
  - This will be determined from the date of birth (BIRTHDAT in the DEM module) and date of randomisation (RND_DAT in the CRIT1 module) on the eCRF at screening. As only year of birth is collected in this study, all patients will have an assumed date of birth of 1st Jan [given year]. Patients with a missing age value will be categorised as missing.
- Race (White or Other)
- Baseline ECOG/WHO performance status (0/1)

- Prior use of CDK4/6 inhibitors (yes/no)
- Presence of lung and/or liver metastasis (yes/no)
- Visceral disease (yes/no)
  - Yes: defined as any disease at baseline (from the extent of disease page of the CRF) identified in: adrenal, brain/ central nervous system, oesophagus, liver, lung/ respiratory, pancreas, spleen, cardiovascular, gall bladder, gastrointestinal, genitourinary
  - No: defined as any disease at baseline (from the extent of disease page of the CRF) identified in: peritoneum, ascites, pericardial effusion, pleural effusion, breast, bone and locomotor, lymph nodes, neck skin/ soft tissue, other locally advanced sites
  - Any sites specified under "other metastatic sites" will be reviewed manually to determine their classification
- ESR1 mutation present at baseline (yes/no) from either assays from Guardant Health or Resolution Bioscience
  - Determined based on the 11 most common ESR1 mutations.
    - D538G
    - Y537S
    - Y537N
    - Y537C
    - Y537D
    - L536H
    - L536P
    - L536R
    - S463P
    - V422del/VE422E (note that this is the same mutation, annotated differently by different vendors)
    - E380Q
- Sensitive or resistant to endocrine therapy[a]
  - Resistant: defined as:
    - A relapse while on the first 24 months of adjuvant endocrine therapy
    - Progression within the first 6 months of first-line endocrine monotherapy for advanced disease
    - Progression within the first 12 months of first-line endocrine combination therapy (CDK4/6 inhibitors) for advanced disease
  - Sensitive: defined as:
    - Received adjuvant endocrine therapy for at least 24 months
    - Progression ≥ 6 months after initiating endocrine therapy as monotherapy for advanced disease
    - Progression ≥ 12 months after initiating endocrine combination therapy (CDK4/6 inhibitors) for advanced disease

    Note: If a patient received endocrine therapy in the advanced and adjuvant setting then only the advanced setting will be used to determine endocrine sensitivity.
- ER percent positivity (total of weak '+', moderate '++', strong '+++' staining) at Baseline (<10%, ≥10%)
  - ER percent positivity < 10% defined as H-score < 10 and/or Allred score < 3
  - ER percent positivity ≥ 10% defined as H-score ≥ 30 and/or Allred score ≥ 6

- Results not assigned an ER percent positivity category per above definitions are in a "grey area" and will not be included in subgroup analyses

[a] Endocrine therapies, including CDK4/6 inhibitors will be identified using ATC terms. As far as feasible, the determination of endocrine therapy status (sensitive or resistant) will be determined programmatically, but if deemed necessary may be determined using a manual review by a medically qualified expert.

The subgroup analyses for the stratification factors will be based on the values entered into the IWRS, all other factors will be based on values recorded on the eCRF as indicated above.

For each subgroup level of a factor, the median, HR and 90% CI will be calculated from a single Cox proportional hazards model that contains a term for treatment, the subgroup covariate of interest and the treatment by subgroup interaction term. The treatment effect HR will be obtained for each level of subgroup from this model. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control ties.

These HRs and associated two-sided 90% CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis.

Other baseline variables may also be assessed if there is clinical or biological justification or an imbalance is observed between the treatment arms. The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic factors. If a baseline imbalance is observed between treatment arms, ad-hoc subgroup analyses may be used to investigate any potential impact on the main results.

Any subgroup containing less than 10 events will be suppressed and will not be presented in outputs.

No adjustment to the significance level for testing will be made since the subgroup analysis will be considered exploratory and may only be supportive of the primary analysis of PFS.

4.10.1.5.1 Test for Consistency of Treatment Effect Between Subgroups

The presence of quantitative interactions will be assessed by means of an overall global interaction test for subgroups defined in Section 4.10.1.5 with the exception of age, race and ECOG/WHO performance status. Presence of lung/liver metastasis will also be excluded as this subgroup is highly correlated with visceral disease which will be included in this analysis.

This is performed by comparing the fit of a Cox proportional hazards model including treatment, all covariates, and all covariate-by-treatment interaction terms, with one that excludes the interaction terms, and will be assessed at the two-sided 10% significance level. If there are not more than 10 events per stratum for any covariate (i.e., within each stratum of a treatment*covariate interaction [2 treatments * 2 levels of the covariate = 4 stratum]) a pre-defined pooling strategy should be applied to the covariate. If the pooling strategy does not meet the event criteria, then the covariate-by-treatment interaction term should be omitted from the model. Moreover, if the

covariate does not have more than 10 events per level of covariate then the main effect of the covariate will also be excluded. If the fit of the model is not significantly improved, then it will be concluded that overall the treatment effect is consistent across subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and newly significant interactions re-included until a final model is reached where all included interactions are significant, and all excluded interactions are non-significant. Throughout this process all main effects will be included in the model regardless of whether the corresponding interaction term is still present. This approach will identify the factors that independently alter the treatment effect and prevent identification of multiple correlated interactions.

Any quantitative interactions identified using this procedure will then be tested to rule out any qualitative interactions using the approach of Gail and Simon 1985.

4.10.2 Derivation of RECIST Visit Responses

For all patients, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. It will also be used to determine if, and when a patient has progressed in accordance with RECIST and their BoR to study treatment.

Baseline radiological tumour assessments will be performed no more than 28 days before the start of randomised treatment and ideally as close as possible to the start of study treatment. Tumour assessments are then performed every eight weeks (± 1 week) after randomisation until disease progression. Baseline values recorded after randomisation will not be used as the baseline assessment, although such assessments can be used in the calculation of progressive disease (PD).

If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits. This schedule is to be followed in order to minimise any unintentional bias caused by some patients being assessed at different frequency than other patients.

From the investigator's review of the imaging scans, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. At each visit, patients will be programmatically assigned a RECIST version 1.1 visit response of complete response (CR), partial response (PR), stable disease (SD), or PD, using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions and depending on the status of their disease compared with baseline and previous assessments. If a patient had a tumour assessment that cannot be evaluated, then the patient will be assigned a visit response of not evaluable (NE) (unless there is evidence of progression in which case the response will be assigned as PD).

Refer to Table 5 for the definitions of CR, PR, SD, and PD.

RECIST outcomes (i.e., PFS, ORR, etc.) will be calculated programmatically for the site investigator data from the overall visit responses.

4.10.2.1 Target Lesions (TLs) – site investigator data

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (LD) (except lymph nodes which must have short axis ≥ 15 mm) with computed tomography (CT) or magnetic resonance imaging (MRI) and which is suitable for accurate repeated measurements. A patient can have a maximum of five measurable lesions recorded at baseline with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as TLs. If more than one baseline scan is recorded, then measurements from the one that is closest and prior to randomisation will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

All other lesions (or sites of disease) not recorded as TL should be identified as NTLs at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

For patients who do not have measurable disease at entry (i.e., no TLs), but have non-measurable disease (for this study defined as having at least one lytic or mixed [lytic + sclerotic] bone lesion that is amenable to serial assessment by CT or MRI), evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions. If a patient does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

Table 3 provides the definitions of target lesion visit responses based on RECIST version 1.1.

Table 3 Target Lesion Visit Responses (RECIST 1.1)

Visit Responses	Description
Complete response (CR)	Disappearance of all TLs. Any pathological lymph nodes selected as TLs must have a reduction in short axis to < 10 mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5 mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR, nor sufficient increase to qualify for PD.

Table 3 Target Lesion Visit Responses (RECIST 1.1)

Visit Responses	Description
Not evaluable (NE)	Only relevant in certain situations (i.e., if any of the TLs were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the PD criteria, PD overrides not evaluable as a TL response.
Not applicable (NA)	No TLs are recorded at baseline.

CR = complete response; NA = not applicable; NE = not evaluable; PD = progressive disease; PR = partial response; SD = stable disease; TL = target lesion.

Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to one decimal place before assigning a TL response. For example, 19.95% should be rounded to 20.0%, but 19.94% should be rounded to 19.9%.

Missing TL data

For patients with measurable disease at entry, a visit can only be considered evaluable if all TL measurements have been recorded. However, a visit response of PD should still be assigned if any of the following occurred:

- A new lesion is recorded
- A NTL visit response of PD is recorded
- The sum of TLs sufficiently increased to result in a 20% increase, and an absolute increase of ≥ 5 mm from nadir, even assuming the non-recorded TLs have disappeared.

The nadir (i.e., smallest measurement) can only be taken from assessments where all the TLs had a LD recorded.

Lymph nodes

For lymph nodes, if the size reduces to < 10 mm then these are considered non-pathological. However, a size will still be given, and this size should still be used to determine the TL visit response as normal. In the special case where all lymph nodes are < 10 mm and all other TLs are 0 mm then although the sum may be > 0 mm the calculation of TL response should be over-written as CR.

TL visit responses subsequent to CR

Only CR, PD or NE can follow a CR. If a CR has occurred, then the following rules at the subsequent visits must be applied:

- Step 1: If all lesions meet the CR criteria (i.e., 0 mm or < 10 mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met, i.e., if a lymph node LD increases by 20%, but remains < 10 mm.
- Step 2: If some lesion measurements are missing, but all other lesions meet the CR criteria (i.e., 0 mm or < 10 mm for lymph nodes), then response will be set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.
- Step 3: If not all lesions meet the CR criteria (i.e., a pathological lymph node selected as TL has short axis > 10 mm or the reappearance of previously disappeared lesion) or a new lesion appears, then response will be set to PD.
- Step 4: If after steps 1 – 3 a response can still not be determined; the response will be set to remain as CR.

TL too big to measure

If a TL becomes too big to measure this should be indicated in the database and a size (*'x'*) above which it cannot be accurately measured should be recorded. If using a value of *x* in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team. It is expected that a visit response of PD will remain in most cases.

TL too small to measure

If a TL becomes too small to measure, then this will be indicated as such on the eCRF and a value of 5 mm will be entered into the database and used in TL calculations. However, a smaller value may be used if the radiologist has not indicated 'too small to measure' on the eCRF and has entered a smaller value that can be reliably measured. If a TL response of PD results (at a subsequent visit) then this will be reviewed by the study team.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e., lesions irradiated prior to entry into the study) should be recorded as NTLs and should not form part of the TL assessment.

Any TL (including lymph nodes), which has had an intervention during the study (for example, irradiation/ palliative surgery/ embolisation), should be handled in the following way:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and calculation will be performed in the usual manner. If the visit response is PD, this will remain as a valid response category.
- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with interventions) as missing and if ≤ 1/3 of the TLs have missing

measurements then scale up as described in the 'Scaling' section below. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If, after both steps, PD has not been assigned, then, if appropriate (i.e., ≤ 1/3 of the TLs have missing measurements), the scaled sum of diameters calculated in step 2 should be used, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 mm (or < 10 mm for lymph nodes) and the lesions that have been subject to intervention have a value of 0 mm (or < 10 mm for lymph nodes) recorded. If scaling up is not appropriate due to too few non-missing measurements, then the visit response will be set to NE.

At subsequent visits, the above steps will be repeated to determine the TL and overall visit response. When calculating the previous minimum, lesions with interventions should be treated as missing and scaled up (as per step 2 above).

Once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours.

Scaling (applicable only for irradiated lesions/lesion intervention)

If > 1/3 of TL measurements are missing (because of intervention) then the TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (i.e., if using a value of 0 mm for missing lesions, the sum of diameters has still increased by 20% or more compared to nadir and the sum of TLs has increased by ≥ 5 mm from nadir).

If ≤ 1/3 of the TL measurements are missing (because of intervention) then the results will be scaled up (based on the sizes at the nadir visit to give an estimated sum of the diameters) and this will be used in calculations; this is equivalent to comparing the visit sum of diameters of the non-missing lesions to the nadir sum of diameters excluding the lesions with missing measurements.

Example of scaling:

Lesion 5 is missing at the follow-up visit; the nadir TL sum including lesions 1 – 5 was 74 mm.

The sum of lesions 1 – 4 at the follow-up is 68 mm. The sum of the corresponding lesions at the nadir visit is 62 mm.

Scale up as follows to give an estimated TL sum of 81 mm: 68 x 74 / 62 = 81 mm.

CR will not be allowed as a TL response for visits where there is missing data. Only PR, SD, or PD (or NE) can be assigned as the TL visit response in these cases. However, for visits with ≤ 1/3 lesions assessments not recorded, the scaled-up sum of TLs diameters will be included when defining the nadir value for the assessment of progression.

Lesions that split in two

If a TL splits in two, then the LDs of the split lesions should be summed and reported as the LD for the lesion that split.

Lesions that merge

If two TLs merge, then the LD of the merged lesion should be recorded for one of the TL sizes and the other TL size should be recorded as 0 mm.

Change in method of assessment of TLs

CT, MRI and clinical examination are the only methods of assessment that can be used within a trial, with CT and MRI being the preferred methods and clinical examination only used in special cases. If a change in method of assessment occurs, between CT and MRI this will be considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g., CT changes to clinical examination or vice versa), any affected lesions should be treated as missing.

4.10.2.2 Non-target Lesions (NTLs) and New Lesions – site investigator data

At each visit, the investigator should record an overall assessment of the NTL response. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit.

NTL response will be derived based on the investigator's overall assessment of NTLs as described in Table 4.

Table 4 Non-target Lesion Visit Responses

Visit Response	Description
Complete response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (< 10 mm short axis).
Progressive disease (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases, the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Non-CR/Non-PD	Persistence of one or more NTLs with no evidence of progression.

Table 4 Non-target Lesion Visit Responses

Visit Response	Description
Not evaluable (NE)	Only relevant when one or more of the NTLs where not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For patients without TLs at baseline, this is relevant if any of the NTLs were not assessed at this visit and the progression criteria have not been met.
Not applicable (NA)	Only relevant if there are no NTLs at baseline.

CR = complete response; NA = not applicable; NE = not evaluable; NTL = non-target lesion; PD = progressive disease.

To achieve 'unequivocal progression' based on NTLs, there must be an overall level of substantial worsening in non-target disease such that, even in the presence of SD or PR in TLs, the overall tumour burden has increased sufficiently to merit a determination of disease progression. A modest 'increase' in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

A lesion is identified at a follow-up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

The finding of a new lesion should be unequivocal, i.e., not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumour.

New lesions will be identified via a Yes/No tick box in the eCRF. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses.

A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.

If the question 'Any new lesions since baseline' has not been answered with Yes or No and the new lesion details are blank, this is not evidence that no new lesions are present but should not overtly affect the derivation.

Symptomatic progression is not a descriptor for progression of NTLs: it is a reason for stopping a study therapy and will not be included in any assessment of NTLs.

Patients with 'symptomatic progression' requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumour assessments where possible until objective disease progression is observed.

4.10.2.3 Overall Visit Response – site investigator data

Table 5 defines how previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 5 Overall Visit Responses

Target	Non-target	New Lesions	Overall Visit Response
CR	CR or NA	No (or NE)	**CR**
CR	Non-CR/Non-PD or NE	No (or NE)	**PR**
PR	Non-PD or NE or NA	No (or NE)	**PR**
SD	Non-PD or NE or NA	No (or NE)	**SD**
PD	Any	Any	**PD**
Any	PD	Any	**PD**
Any	Any	Yes	**PD**
NE	Non-PD or NE or NA	No (or NE)	**NE**
NA	CR	No (or NE)	**CR**
NA	Non-CR/Non-PD	No (or NE)	**SD**
NA	NE	No (or NE)	**NE**

CR = complete response; NA = not applicable; NE = not evaluable; PD = progressive disease; PR = partial response; SD = stable disease.

4.10.2.4 Blinded Independent Central Review of Tumour Response

A planned Blinded Independent Central Review (BICR) of all radiological imaging data will be carried out using RECIST version 1.1. Coded copies of all imaging assessments (regardless of modality and including unscheduled visit scans) will be sent to the BICR for central analysis.

The imaging scans will be reviewed by two independent radiologists using RECIST version 1.1 and will be adjudicated, if required (i.e., two reviewers will review the scans and, in case of a disagreement, adjudication is performed by a separate reviewer). The independent reviewers will be blinded to study treatment.

The BICR will define the overall visit response (i.e., the response obtained overall at each visit by assessing TLs, NTLs, and new lesions) and no programmatic derivation of overall visit response is necessary. (For patients with TLs at baseline: CR, PR, SD, PD, NE; for patients with NTLs only: CR, SD, PD, NE). If a patient has a tumour assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision will be used). In the absence of adjudication, the records for all visits for a single reviewer will be

used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to report all BICR RECIST information including dates of progression, visit response, censoring and changes in TL dimensions. PFS will be derived programmatically from this information.

Results of the independent review will not be communicated to investigators, and the management of patients will be based solely upon the results of the RECIST assessment conducted by the investigator.

A BICR of all patients will be performed for the final database lock for the primary analysis of PFS, and may also be conducted for interim analysis 2. The BICR will cover all available scans up to the respective DCO.

Further details of the BICR will be documented in the BICR charter.

4.10.3 Primary Efficacy Variable – Progression-free Survival

PFS is defined as the time from randomisation until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomised therapy or received another anti-cancer therapy prior to progression (i.e., date of PFS event or censoring – date of randomisation + 1). Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their latest evaluable RECIST assessment. However, if the patient progresses or dies immediately after two or more consecutive missed visits, the patient will be censored at the time of the latest evaluable RECIST version 1.1 assessment prior to the two missed visits. NE is not considered a missed visit.

Given the schedule of RECIST assessments (i.e., eight-weekly) two missing visits will equate to 18 weeks, allowing for early and late visits (i.e., 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks).

If the patient has no evaluable visits or does not have baseline data, they will be censored at Day 1 unless they die within 17 weeks (16 weeks + 1 week allowing for a late assessment within the visit window) of baseline.

The PFS time will always be derived based on the scan/assessment dates and not visit dates.

RECIST assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- For BICR assessments, the date of progression will be determined based on the **earliest** of the scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of the reviewer who read baseline first if there is no adjudication for BICR data.
- For investigator assessments, the date of progression will be determined based on the **earliest** of the dates of the component that triggered the progression.

- For both BICR and investigator assessments, when censoring a patient for PFS the patient will be censored at the **latest** of the dates contributing to a particular overall visit assessment.

Note: for both TLs and NTLs only the latest scan date is recorded out of all scans performed at that assessment for the TLs/NTLs respectively.

The treatment status at progression of patients at the time of analysis will be summarised. This will include the number and percentage of patients who were on treatment at the time of progression, the number and percentage of patients who discontinued study treatment prior to progression, the number and percentage of patients who have not progressed and were on treatment or discontinued treatment.

PFS will be analysed using a stratified log-rank test adjusting for the stratification factors described in Section 4.10.1.2 for generation of the p-value (using PROC LIFETEST with a TEST statement).

The HR for each treatment comparison against fulvestrant and their CIs (both 90% and 80% [to support the go/no-go decisions]) will be estimated from a stratified Cox Proportional Hazards model (with ties = Efron and stratification factors described in Section 4.10.1.2) and the CIs calculated using a profile likelihood approach (RISKLIMITS = PL in PROC PHREG).

A Kaplan-Meier plot of PFS will be presented with all treatment arms (including the AZD9833 300 mg arm that was closed for enrolment) overlaid on one plot. Summaries of the number and percentage of patients experiencing a PFS event, and the type of event (progression or death) will be provided along with the median PFS and respective CIs, and the proportion of patients that are alive and progression-free at three months, six months, nine months, 12 months, 18 months and 24 months for each treatment arm, calculated using the Kaplan-Meier technique.

The assumption of proportionality will be assessed. Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time depending covariate (adding a treatment-by-time or treatment-by-ln(time) interaction term) to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect can be described by presenting piecewise HRs calculated over distinct time periods. In such circumstances, the HR from the primary analysis can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found this may be a result of a treatment-by-covariate interaction, which will be investigated.

In addition, the number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if they had not progressed (or died in the absence of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval plus two weeks prior to the DCO date.

Additionally, summary statistics will be given for the number of days from censoring to DCO for all censored patients.

The duration of follow-up will be summarised using median time from randomisation to date of censoring (date last known to have not progressed) in censored (not progressed) patients only, presented by treatment group.

Additionally, summary statistics for the number of weeks between the time of progression and the last evaluable RECIST assessment prior to progression will be presented for each treatment group.

Summaries of the number and percentage of patients who miss two or more consecutive RECIST assessments will be presented for each treatment group. The average days between RECIST assessments per patient will also be summarised for each treatment group.

An additional Kaplan-Meier plot of the time to censoring using the PFS data from the primary analysis and where the censoring indicator of the PFS analysis is reversed will also be presented.

All collected RECIST data will be listed for all randomised patients. In addition, a summary of new lesions (i.e., site of new lesions) will be produced.

Subgroup analysis

In addition to the forest plot described in Section 4.10.1.5 the stratified log-rank test and Kaplan-Meier plots for the PFS analysis described above will be repeated for prior use of CDK4/6 inhibitors (yes/no). This stratified log-rank test will adjust for the presence of lung and/or liver metastasis (yes/no). Similarly, the analyses will be repeated for lung and/or liver metastasis (yes/no) with the log-rank test adjusting for CDK4/6 inhibitors (yes/no). The stratified log-rank test and Kaplan-Meier plots for PFS analysis will in addition also be presented for ESR1 mutation.

Kaplan-Meier plots will also be presented for endocrine therapy status (sensitive or resistant).

4.10.3.1 Primary Efficacy Variable – Sensitivity Analyses

Evaluation-time bias

A sensitivity analysis will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled timepoints. The midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of assessment) will be analysed using a stratified log-rank test, as described in Section 4.10.3. Midpoint values resulting in non-integer values will be rounded down. For patients whose death was treated as a PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). To support this analysis, the mean of patient-level average inter-assessment times will be tabulated for each treatment. This approach will use the investigator RECIST assessment.

Attrition bias

Attrition bias will be assessed by repeating the PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression

immediately following two, or more, non-evaluable tumour assessments will be included. In addition, and within the same sensitivity analysis, patients who take subsequent therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) prior to their last evaluable RECIST assessment or progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy.

Ascertainment bias

Ascertainment bias will be assessed by analysing the BICR data. The PFS analysis will be repeated using the BICR data based upon RECIST.

If there is an important discrepancy between the primary analysis using the investigator data and this sensitivity analysis using BICR data, then the proportion of patients with site but no central confirmation of progression will be summarised; such patients have the potential to induce bias in the central review due to informative censoring. An approach of imputing an event at the next visit in the central review analysis may help inform the most likely HR value (Fleischer et al 2011), but only if an important discrepancy exists. A team review of this data will be performed to determine if an important discrepancy exist.

Disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group.

Deviation bias

Deviation bias may be assessed by repeating the PFS analysis excluding patients with deviations that may affect the efficacy of trial therapy. Refer to Section 4.6.2 for further details.

COVID-19

A sensitivity analysis will be conducted to assess the potential impact of COVID-19 related deaths on PFS. That is, patients who had a PFS event due to death where the primary or secondary cause of death was COVID-19 infection or COVID-19 infection was reported as a fatal AE, will be censored on the last available RECIST 1.1 assessment prior to COVID-19 infection related death.

4.10.3.2 Primary Efficacy Variable – Exploratory Analyses

CCI [redacted]

[redacted]

[redacted]

accounts for the correlation between the various test statistics due to the common control arm CCI

[redacted]

External evidence

To further increase precision of the estimation of the reduction in the risk of progression or death for AZD9833 vs. fulvestrant, a historical control fulvestrant arm may be added using data from other clinical trials or real-world data. If implemented, the analysis will be fully prespecified in a SAP addendum.

Weighted log-rank

Should a lack of proportional hazards be observed, a weighted log-rank test will be performed to test for early, middle or late differences between the PFS curves. The most appropriate ranking will be determined based on the observed PFS curves.

4.10.4 Secondary Efficacy Variables

4.10.4.1 Objective Response Rate

ORR is defined as the percentage of patients with at least one investigator assessed visit response of CR or PR prior to any evidence of progression and will be based on a subset of the FAS with measurable disease at baseline per site investigator. A response does not need to be confirmed to be included in the calculation of ORR.

Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue randomised treatment without progression, receive a subsequent anti-cancer therapy (for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) and then respond will not be included as responders in the ORR.

The ORR will be based on the investigator RECIST data and using all scans regardless of whether they were scheduled or not. The ORR will be compared between AZD9833 (each dose level) and fulvestrant using a logistic regression model (including all three AZD9833 doses) adjusting for the stratification factors described in Section 4.10.1.2. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour AZD9833) together with its associated profile likelihood 90% CI (e.g., using the option 'LRCI' in SAS PROC GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model).

If there are not enough responses for a meaningful analysis using logistic regression, then a Cochran-Mantel-Haenszel test will be presented.

Summaries will be produced that present the number and percentage of patients with a tumour response (CR/PR) (including frequencies of confirmed response) based upon the number of patients with measurable disease at baseline per investigator. Patients with measurable disease at baseline without post-baseline RECIST scans will be presented as 'NE' in the summaries.

For the purpose of summarising confirmed response, confirmed response is defined as a response of CR/PR recorded at one visit and confirmed by repeat imaging not less than four weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit.

Subgroup Analysis

The logistic regression for ORR described above will be repeated for CDK4/6 inhibitors (yes/no) and for endocrine status (sensitive/resistant). The model will include the covariate of interest and the treatment by subgroup interaction term. The treatment effect will be obtained for each level of subgroup from this model.

Best objective response

BoR is calculated based on the overall visit responses from each RECIST assessment, described in Section 4.10.2.3. It is the best response a patient has had following randomisation, but prior to starting any subsequent cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categories of BoR will be based on RECIST using the following response categories: CR, PR, SD, PD, and NE.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least eight weeks minus 1 week, i.e., 49 days (to allow for an early assessment within the assessment window), after

randomisation. For CR/PR, the initial overall visit assessment that showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

BoR will be determined programmatically based on RECIST from the programmatically calculated overall visit response using all investigator assessed data up until the first progression event. The denominators for each case will be consistent with those used in the ORR analysis, i.e., those with measurable disease.

For patients whose progression event is death, BoR will be calculated based upon all evaluable RECIST assessment prior to death.

For patients who die with no evaluable RECIST assessment, if the death occurs ≤ 17 weeks (i.e., 16 weeks + 1 week to allow for a late assessment within the assessment window) after randomisation, then BoR will be assigned the progression (PD) category. For patients who die with no evaluable RECIST assessment, if death occurs > 17 weeks after randomisation, then BoR will be assigned to the NE category. For patients who die with no evaluable RECIST assessment that started subsequent cancer therapy, then BoR will be assigned to the NE category.

A patient will be classified as a responder if the RECIST criteria for a CR or PR are satisfied at any time following randomisation, prior to RECIST progression and prior to starting any subsequent cancer therapy.

For each treatment arm the BoR will be summarised by number of patients and percentage for each category (CR, PR, SD, PD, and NE). No formal statistical analyses are planned for BoR.

For each treatment arm, the number of patients that progressed after having SD for 7 weeks, 15 weeks and 23 weeks respectively will be presented.

Clinical benefit rate (disease control rate) at 24 weeks

The CBR_{24} is defined as the percentage of patients who have a BoR of CR or PR in the first 25 weeks (to allow for a late assessment within the assessment window) or who have SD (without subsequent cancer therapy) for at least 23 weeks after randomisation (to allow for an early assessment within the assessment window). For patients with no result between 23 to 25 weeks, SD can be assumed if the last result prior to 23 weeks and the first result after 25 weeks are both SD. If the last result prior to 23 weeks was SD and the first result after 25 weeks is not SD and there was no BoR of CR or PR in the first 25 weeks, then no CBR_{24} result will be defined for this patient. The CBR_{24} will be defined based on the investigator's assessment of RECIST.

The CBR_{24} will be summarised by treatment group and analysed using a logistic regression model similar to the analysis described for ORR. Analysis on CBR_{24} will be based on the subset of patients who were randomised at least 25 weeks prior to the DCO for analysis.

4.10.4.2 Duration of response

DoR will be defined as the time from the date of first documented response until the date of documented progression or death in the absence of disease progression (i.e., date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response will be defined as the latest of the dates contributing towards the first response of PR or CR.

If a patient does not progress following a response, then their DoR will use the PFS censoring time.

Descriptive data will be provided for the DoR in responding patients, including the associated Kaplan-Meier curves, medians (utilising Kaplan-Meier methodology) and corresponding CIs (without any formal comparison or p-value attached).

Time to response

Time to response (TTR) is defined as the time from the date of randomisation until the date of first documented response. The date of first documented response will coincide with that used for DoR. TTR will not be defined for those patients who do not have a documented response.

The TTR will be summarised (i.e., number of patients [%] based upon the number of responders) by the scheduled assessment timepoint that the response was first observed. Additional descriptive summary statistics (i.e., median and quartiles) will also be presented (utilising Kaplan-Meier methodology) and will include corresponding CIs (without any formal comparison or p-value attached).

4.10.4.3 Change in tumour size

One of the secondary outcome variables for this study is percentage change from baseline in the TL tumour size at 16 weeks. This is based on RECIST TL measurements taken at baseline and at 16 weeks (Day 1 Week 17). Tumour size is the sum of the LDs of the TLs (or short axis measurements for lymph nodes). TLs are measurable tumour lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to randomisation. The percentage change in TL tumour size at 16 weeks will be obtained for each patient taking the difference between the sum of the TLs at 16 weeks and the sum of the TLs at baseline divided by the sum of the TLs at baseline times 100 (i.e., [16 weeks – baseline]/baseline * 100).

Patients who progress before 16 weeks should have had a tumour assessment performed at the time of progression prior to treatment discontinuation. For these patients, the tumour size from their latest progression assessment will be used instead of the assessment at 16 weeks.

The absolute values, change in TL tumour size from baseline and percentage change in TL tumour size from baseline will be summarised using descriptive statistics and presented at each timepoint and by randomised treatment group. The percentage change in 16-week TL tumour size from baseline will also be summarised and presented by randomised treatment group.

The effect of AZD9833 on percentage change in TL tumour size will be estimated from one analysis of covariance (ANCOVA) model including treatment group, a term for the percentage change in week 16 value, the stratification factors described in Section 4.10.1.2, a covariate for baseline TL tumour size and a covariate for the time from the baseline scan to randomisation. The number of patients, unadjusted mean and least square means (LSMeans) for each treatment group will be presented, together with the differences in LSMeans versus fulvestrant, 90% CIs and corresponding p-values (p-values will not be presented for the comparison of AZD9833 300 mg vs. fulvestrant).

A histogram of the residuals from fitting an ANCOVA model (including all relevant covariates including treatment) to the data (note: this includes both actual and imputed data) will be produced and used to assess whether it is appropriate to analyse the data using percentage change from baseline at week 16. The decision regarding normality needs to be made after unblinding to aggregate efficacy data when the model can be fitted to the full dataset. If the week 16 TL tumour size data follow a normal distribution, these data will be analysed as described previously. If, however, if it is judged the data do not adequately follow a normal distribution, then the use of log-transformed data or a non-parametric approach will be used.

If a log transformation of the data is required, the change in TL tumour size at week 16 will be assessed as the ln of the BSR of the week 16 over the baseline TL tumour size measurement for each patient (see Section 4.2 for detailed description of this variable). Zero TL sums will be substituted with 5 mm for the purposes of this calculation. The effect of AZD9833 on change in TL tumour size will be estimated from an ANCOVA model including a term for the covariate for baseline tumour size (ln transformed), the stratification factors described in Section 4.10.1.2, and a covariate for the time from baseline scan to randomisation. The ratio of the geometric LSMeans (i.e., back-transformed LSMean difference) with corresponding CI and p-values will also be presented. In order to aid interpretation, the gmean BSR (and corresponding CI) will also be presented as a percentage change from baseline (for example a geometric LSMean of 0.8 would correspond to a 20% reduction).

If the primary analysis is completed using percentage change from baseline or the log ratio, a sensitivity analysis will also be performed using a non-parametric method. If a non-parametric method is used either as a primary or sensitivity analysis, an ANCOVA model on the ranked percentage change in TL tumour size will be used, including a covariate for baseline TL tumour size, a covariate for the time from the baseline scan to randomisation and the stratification factors described in Section 4.10.1.2. The ranking will be of the analysis data set following appropriate imputation. The patient with the greatest reduction in TL tumour size will be assigned the lowest rank with smaller changes and increases in tumour sizes taking increasing ranks. Deaths prior to the end of the window used to select week 16 data will be assigned the highest rank (i.e., those patients for whom a value equal to the maximum percentage change in tumour size was imputed because the patient died). The p-values from this ANCOVA will be presented together with the Hodges Lehmann estimate of the median differences and corresponding 90% CIs will be derived. The median percentage change and range will be presented for each treatment group together with the number of patients and percentage of patients in each treatment group whose 16-week data was imputed in the non-parametric analysis.

Tumour size will also be presented graphically using waterfall plots for each treatment group, to present each patient's week 16 percentage change in TL tumour size as a separate bar with the bars ordered from the largest increase to the largest decrease. A reference line at the -30% change in TL tumour size level will be added to the plots, which correspond with the definition of a 'partial response'. All progressions will be marked with a '●' or designated with patterns or colours for ORR categories. Flagged progressions on the percentage change in TL tumour size at a particular timepoint will be based upon the NTL or new lesion progression at that timepoint. The scale in these plots will be fixed to be from -100 to 100 to avoid presenting extreme values. Values that are capped as a result of this restriction to the scale will be marked with '#'. Values are ordered in descending order with the imputations due to death appearing first followed by a gap followed by all other patients. Imputed values will be clearly marked with '*' and patients with imputation where there was a death or evidence of progression will have different shading to each other and the other patients to make it clear that these are different. A similar plot will be presented with all patients included. Different colours for each treatment group will be used for this combined plot.

Additionally, 'spider' plots will be produced for each treatment group. This depicts each patient's percentage change in TL tumour size as a line over time and progression due to NTLs or new lesions will be indicated (imputed data will not be included in the plots).

Best percentage change in tumour size

The absolute change and percentage change from baseline in the sum of tumour size at each assessment will be calculated. The best change in tumour size (i.e., depth of response) is the largest decrease from baseline or the smallest increase from baseline in the absence of a reduction and will include all assessments prior to the earliest of death in the absence of progression, any evidence of progression, the start of subsequent anti-cancer therapy (including radiotherapy), or the latest evaluable RECIST assessment if the patient has not died, progressed or started subsequent anti-cancer therapy. Change in tumour size at progression or the latest evaluable RECIST assessment (as applicable) should be included in the determination of best percentage change in tumour size.

If best percentage change cannot be calculated due to missing data (including if the patient has no TLs at baseline), a value of +20% will be imputed as best percentage change from baseline in the following situations (otherwise best percentage change will be left as missing):

- If a patient has no post-baseline assessment and has died
- If a patient has new lesions or progression of NTLs or TLs
- If a patient has withdrawn due to PD and has no evaluable TL data before or at PD

The best percentage change in TL tumour size from baseline will be summarised and presented by randomised treatment group. The number and percentage of patients in each treatment group whose best percentage change data is imputed will also be presented.

Best percentage change will further be presented graphically using a waterfall plot, similarly to what has been described above for the change in tumour size.

4.10.4.3.1 Missing data imputation methods

For patients who have less than or equal to one-third of TLs missing (because of intervention) at week 16, assessment data from missing lesions will be scaled up proportionally to the sum of the corresponding lesions at baseline to give an estimated sum of diameters as described in Section 4.10.2.1 (Scaling).

Whenever TL tumours size data for the week 16 visit (or visit at which progression was documented if before week 16) is available then this should be used in the analysis. Windowing will be applied as described in Section 4.3.2; therefore, any RECIST scan performed within the allowable window will be used for that visit.

If after applying the above considerations to the missing data, there is still missing TL tumour size measurement data at week 16, the imputation process outlined below will be followed for each individual patient where data is missing.

- If there is no observed tumour size measurement data at week 16, but there is TL tumour size measurement data collected at a visit prior to week 16 or the first scheduled visit after week 16, use all of the available data up to and including the first scheduled visit after week 16 (i.e., baseline and all visits up to and including the first scheduled visit after week 16) to fit a linear regression to the individual patient's baseline and follow-up assessment(s). Note that actual day of the measurement rather than planned day should be used in fitting this model. This model can be used to generate an estimated value for TL tumour size measurement at week 16 and hence impute a change from baseline at week 16.
- If there is no observed tumour size measurement data at week 16, but there is evidence of progression for the individual prior to the end of the time window used to select week 16 data, where evidence of progression is defined as progression of NTLs, the appearance of new lesions or as determined by an investigator (i.e., investigator's opinion of response recorded on the eCRF is PD at that assessment or study treatment was discontinued for progression in the assessment window), and there is at least 5 patients with non-missing TL tumour size measurements who have also progressed at this timepoint, then impute a change from baseline at week 16 as the median percentage change from patients with non-missing tumour size who also have progressed. If the patient already has an imputed value from the first bullet above, then use the maximum value of this calculated median or the imputed value from the bullet above. However, if there are less than 5 patients with non-missing tumour size who have also progressed then impute a change from baseline at week 16 as 20%. If the patient has an imputed value from the first bullet above, use the maximum of 20% or the imputed value from the first bullet above.
- If there is no evidence of progression for the individual, use the imputed value from the first bullet above if data is available. If there is no evidence of progression for the individual and no observed TL tumour size measurement data is collected at a visit prior to week 16 or the first scheduled visit after week 16, assume that the data is completely missing at random. The patient will be excluded from the analysis.
- If it is known that the patient has died prior to the end of the time window used to select week 16 data, impute a change from baseline at week 16 as the maximum of the observed or imputed (as in the second bullet) percentage change reported in the study for week 16.

If after TL imputation and applying a window around the week 16 visit there remains more than 10% of missing TL tumour size measurement data, a non-parametric method will be used to analyse change in tumour size. Patients who have died will be assigned the worst rank.

No imputations will be performed if a patient has no post-baseline measurements.

All imputed data will be derived within the reporting dataset with a corresponding flag against the imputed value to show that the value has been programmatically derived.

Imputed data will not be used for spider plots (since these show the whole profile for a patient as well as progression indicators for NTLs/new lesions) but will be used in the formal statistical analyses and waterfall plots. Imputed data for change in tumour size will not be included in descriptive summaries but will be included for descriptive summaries on best percentage change.

4.10.4.4 Overall survival

OS is defined as the time from the date of randomisation until death due to any cause regardless of whether the patient withdraws from randomised therapy or received another anti-cancer therapy (i.e., date of death or censoring – date of randomisation + 1). Any patient not known to have died at the time of analysis will be censored at the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Survival follow-up phone calls will be made in the week following the DCO for the analysis, and if patients are confirmed to be alive or if the death date is after the DCO date these patients will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the final OS analysis should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly available death registries. In the event that the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

For any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it is necessary to use all relevant eCRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment (since SURVIVE module is only completed for patients off treatment if a survival sweep is performed). The last date for each individual patient is defined as the latest among the following dates recorded in the eCRF:

- AE start and stop dates
- Admission and discharge dates of hospitalization
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST eCRF

- Start and stop dates of alternative anti-cancer treatment
- Date last known alive on survival status eCRF
- End of study date

If a patient is known to have died where only a partial death date is available, then the date of death will be imputed as the latest of the last date known to be alive + 1 from the database and the death date using the available information provided:

- For missing day only – using the 1st of the month
- For missing day and month – using the 1st of January

If there is evidence of death, but the date is entirely missing, it will be treated as missing, i.e., censored at the last know alive date.

The OS data will be analysed at the time of the primary analysis of PFS and will use the same methods as outlined for PFS (refer to Section 4.10.3), adjusting for the same set of covariates (provided there are enough events available for a meaningful analysis [> 20% maturity in OS]). The proportion of patients alive will only be presented for the 12 and 24 months milestones. If there are not enough events for a meaningful analysis, descriptive summaries will be provided. Further survival analyses may be conducted after CCI and CCI of patients have died.

A sensitivity analysis for OS will examine the censoring patterns to rule out attrition bias with regards to the primary treatment comparisons, achieved by a Kaplan-Meier plot of time to censoring where the censoring indicator of OS is reversed.

A further sensitivity analysis will be conducted to assess the potential impact of COVID-19 related deaths on OS. That is, patients who had a death event where the primary or secondary cause of death was COVID-19 infection or COVID-19 infection was reported as a fatal AE, will be censored at the date of their COVID-19 infection related death.

The number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if there is no indication that the subject has died, but there is no survival status (either from survival follow-up phone calls or last date known to be alive based on data from eCRF) available in the 10 weeks prior to the DCO.

In addition, duration of follow-up will be summarised using medians:

- In all patients: Time from randomisation to the date of death (i.e., OS) or the date of censoring (date last known to be alive) for censored patients regardless of treatment arm.

4.11 Safety Evaluation

All safety summaries will be based upon the Safety Analysis Set and presented by actual treatment group.

4.11.1 Extent of Exposure

Extent of exposure for fulvestrant will be defined in terms of the number of treatment cycles received. A cycle corresponds to a period of 28 days. If a cycle is prolonged due to toxicity, this should be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

Relative dose intensity (RDI) will be calculated as appropriate. RDI is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation and is helpful to summarise dose reductions. Relative dose intensity will be defined as follows:

- $\text{RDI}(\%)=\frac{d}{D}\times 100$, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would have been delivered if there were no modification to dose or schedule.

Intended (total) and actual exposure will be calculated for AZD9833 and fulvestrant.

Exposure (i.e., duration of treatment) will be defined as follows:

Intended (total) exposure of AZD9833 = min(last dose date where dose > 0 mg, date of death, date of DCO) – first dose date + 1.

Intended (total) exposure of fulvestrant = 28 x (number of cycles received).

Actual exposure of study treatment = intended (total) exposure of study treatment – total duration of dose interruptions, where intended exposure will be calculated as above, and a dose interruption is defined as any length of time where the patient has not taken any of the planned daily dose.

The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Missed and forgotten doses will be recorded on the DOSE module as a dose interruption with the reason recorded as “Subject forgot to take dose”. These missed or forgotten doses will not be included as dose interruptions in summary tables, but the information will appear in the listing for dosing. However, these missed and forgotten doses will be considered in the derivation of actual exposure.

If a patient permanently discontinues study treatment during a dose interruption, then the date of last administration of study medication recorded on the DOSDISC eCRF will be used in the programming.

Total dose of AZD9833 and fulvestrant will be calculated as the sum of all doses received during the study. Average daily dose will then be defined as [Total dose]/[Actual exposure].

The duration of an interruption will be calculated as [date/time treatment resumed – date/time treatment stopped]. Overall duration of interruption will be determined as the sum of all individual interruptions during the study for a particular patient.

The following summaries will be produced:

- Summary of treatment cycles received
- Summary of interruptions (including duration of interruptions) and reductions of study treatment
- Summary of RDI
- Summary of intended and actual exposure of AZD9833
- Summary of total dose and average daily dose

4.11.2 Adverse Events

AEs and SAEs will be collected throughout the study from date of informed consent until the end of the safety follow-up period (28 days after last dose of AZD9833 and 56 days after last dose of fulvestrant). MedDRA (using the latest or current MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute of Common Terminology Criteria for AEs (CTCAE) version 5.0.

Events will be defined as treatment-emergent AEs (TEAEs) if they onset or worsen (by investigator report of a change in CTCAE grade), during the treatment period as defined in the CSP or during the 28-day safety follow-up period. Any events in this period that occur after a patient has received further therapy for cancer (following discontinuation of study treatment) will be flagged in the data listings.

During the evaluation of the AE data prior to lock, an AZ medically qualified expert will review the list of AEs that were not reported as SAEs and AEs leading to discontinuation of study treatment. Based on the expert's judgment, significant AEs of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered other significant AEs (OAEs) and reported as such in the CSR. A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs.

Examples of these could be marked haematological and other laboratory abnormalities, and certain events that lead to intervention (other than those already classified as serious), dose reduction or significant additional treatment.

Some clinical concepts (including some selected individual preferred terms and higher-level terms) are considered to be "AEs of special interest" (AESIs). AESIs represent pre-specified risks that are considered to be of importance to a clinical development program.

These AESIs, if applicable, will be identified as a list of categories provided by the patient safety team.

Reviews will take place prior to database lock to determine whether any AE should be classified as AESIs. The review will identify which higher-level terms, and which preferred terms should contribute to each AESI.

In general, all AE summary tables include only TEAEs. AEs occurring prior to dosing or starting more than 28 days after discontinuation of study drug will be flagged in listings and will not be included in any summaries.

All reported AEs will be listed along with the actual treatment received at the time of onset, date of onset, date of resolution (if AE is resolved), investigator's assessment of CTCAE grade, relationship to study treatment, action taken and outcome. Frequencies and percentages of patients reporting each preferred term will be presented (i.e., multiple events per patient will not be accounted for, except for event level summaries).

Summary information (the number and percentage of patients by treatment) by MedDRA SOC and PT will be tabulated for:

- All TEAEs
- All TEAEs causally related to study treatment
- TEAEs of CTCAE grade 3 or higher
- TEAEs of CTCAE grade 3 or higher, causally related to treatment
- TEAEs with outcome of death
- TEAEs with outcome of death, causally related to treatment
- TEAE leading to dose reduction and interruption (separately)
- All treatment-emergent SAEs
- All treatment-emergent SAEs causally related to study treatment
- TEAEs leading to discontinuation of treatment
- TEAEs leading to discontinuation of treatment, causally related to treatment
- Treatment-emergent SAEs leading to discontinuation of treatment
- Treatment-emergent SAEs leading to discontinuation of treatment, causally related to treatment
- Treatment-emergent OAEs
- Treatment-emergent AESIs (if applicable)

An overall summary of the number and percentage of patients in each of the above categories will be presented, as well as an overall summary of number of events in each of the above categories. In addition, a truncated TEAE table of most common AEs, showing all events that occur in at least 5% of patients overall will be summarized by PT, by decreasing frequency. For SRC purposes, an additional table of most common AEs, showing events that occur in at least 10% of patients, will also be presented.

In addition, an event level summary will be presented for all TEAEs by PT.

Summaries of the number and percentage of patients with TEAEs will also be produced by maximum reported CTCAE grades, SOC, and PT and of most common TEAEs of CTCAE grade 3 or higher by SOC and PT.

The following additional summaries will be presented should more than 2% of the patients treated in this study have COVID-19 infection:

- All TEAEs, excluding TEAEs associated with COVID-19 infection
- All TEAEs associated with COVID-19 infection
- TEAEs with outcome of death, excluding TEAEs associated with COVID-19 infection
- TEAEs associated with COVID-19 infection with outcome of death
- TEAEs leading to discontinuation of study intervention, excluding TEAEs associated with COVID-19 infection
- TEAEs associated with COVID-19 infection leading to discontinuation of study intervention

The above categories will also be included in the overall AE summaries.

Further COVID-19 related summaries related to TEAEs may be added if required.

All AESI PTs, if applicable, searched for in this study will be presented. In addition, summaries of treatment-emergent AESIs will be presented by maximum reported CTCAE grade, by AE outcome, including time of resolution (on-treatment, follow-up or post-treatment), whether treatment was received (yes/no) and action taken.

The number and percentage of patients per categorised number of AESI events will also be summarised where categorised number of events will be: ≤ 5, 5 - ≤ 10, 10 - ≤ 20 and 20 - ≤ 50. The final categories will be data driven and will be reviewed to ensure they are applicable, so the above categories are only a guideline.

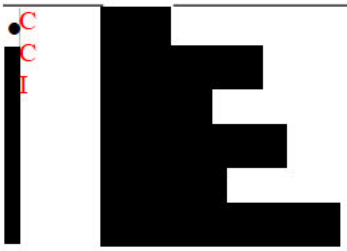
Descriptive statistics for time to onset and duration of first AESI for patients experiencing AESIs will be presented. Time to onset will be derived as [AESI start date – Treatment start date] while duration will be derived as [AESI end date – AESI start date]. If the AESI is ongoing at the respective DCO, the DCO date will imputed as the end date for duration calculations. Time to onset and duration will be presented in days.

In addition, TEAEs with outcome death, treatment-emergent SAEs, TEAEs leading to discontinuation of treatment and treatment-emergent OAEs will be listed separately.

CCI [redacted]

CCI [redacted] AEs will be presented in detail during the study and will follow the same outputs as produced for AESIs (as described above). CCI [redacted] AEs will be identified using the following MedDRA preferred terms:

- Visual perseveration



4.11.3 Deaths

A summary of deaths will be provided with number and percentage of patients, categorised as:

- Related to disease under investigation only
- AE outcome = death only
- Both related to disease under investigation and with AE outcome = death
- AE with outcome = death ≥ 28 days after last treatment
- Other deaths

A corresponding listing will also be produced.

4.11.4 Clinical Laboratory Evaluation

All local laboratory results collected will be listed.

Summaries for safety laboratory will only include the parameters specified in Table 6.

Table 6 Laboratory Safety Variables

Haematology (whole blood)	**Clinical chemistry (serum or plasma)**
B-Haemoglobin	S/P-Creatinine
B-Leukocyte count	S/P-Bilirubin, total
B-Haematocrit	S/P-Conjugated bilirubin
B-Red blood cell count	S/P-Unconjugated bilirubin
B-Leukocyte differential count (absolute count)	S/P-Alkaline phosphatase
Neutrophils	S/P-Aspartate transaminase
Lymphocytes	S/P-Alanine transaminase
Monocytes	S/P-Creatine kinase
Basophils	S/P-Albumin
Eosinophils	S/P-Calcium, total
B-Platelet count	S/P-Potassium
	S/P-Sodium
Coagulation Parameters	S/P-Glucose
Activated partial thromboplastin time	S/P-Magnesium
International normalised ratio	S/P-Phosphate

Table 6 Laboratory Safety Variables

Haematology (whole blood)	**Clinical chemistry (serum or plasma)**
	S/P-Urea nitrogen
Urinalysis (dipstick)	S/P-Protein, total
U-Glucose	S/P-Troponin
U-Protein	S/P-NT-proBNP
U-Blood	

Corrected calcium will be calculated as:

Corrected calcium (mg/dL) = 0.8 × (normal albumin - patients albumin) + total serum calcium

or

Corrected calcium (mmol/L) = 0.02 × (normal albumin - patients albumin) + total serum calcium

depending on the unit used. Normal albumin will be considered as 4 g/dL.

Differential counts (Neutrophils, Lymphocytes, Monocytes, Basophils, and Eosinophils) will be presented as absolute values and percentages. If the results are available only in one of these two units, the other unit will be derived using the following relationship between absolute and percentage results:

$$\text{Differential (\%)} = \frac{\text{Differential (absolute unit)}}{\text{Total White Blood Count (absolute unit)}} \times 100 \text{ or}$$

$$\text{Differential (absolute unit)} = \frac{\text{Differential (\%)}}{100} \times \text{Total White Blood Count (Absolute unit).}$$

All values will be classified as low (below range), normal (within range), or high (above range) based on local laboratory reference ranges. Results will be converted to standard units and graded with CTCAE version 5.0.

If the same parameter is found as measured in serum and in plasma, then the summaries will not distinguish between them (e.g., values from plasma Albumin and serum Albumin will be summarised under Albumin). If the same parameter is found as measured in serum and in plasma within the same patient, which would be a rare case, then the change from baseline will only be calculated for those post-baseline values using the same source, i.e., only within plasma or serum. If one patient has multiple toxicity grades, because they are derived separately from serum and plasma then the maximum value of the two will be considered.

For all continuous laboratory assessments, absolute value, change from baseline and percentage change from baseline will be summarised using descriptive statistics at each scheduled assessment time by actual treatment group.

For clinical chemistry and haematology, shift tables will present movements from baseline to worst value on-treatment (defined from start of treatment to 28 days following the last dose of study treatment) according to reference range classification. CTCAE grade changes from baseline to the maximum grade on-treatment will also be provided. Corresponding shift tables ("Negative", "Trace", "Positive", "0", "+", "++", "+++") will be produced for urinalysis. In addition, the number of patients with ≥ 2 CTCAE grade changes and CTCAE grade changes to 3 or 4 will be summarised by actual treatment group for clinical chemistry and haematology parameters.

Plots for both maximum post-baseline alanine transaminase (ALT) and aspartate transaminase (AST) versus the maximum post-baseline total bilirubin (expressed as multiples of their upper limit of normal [ULN] reference range) will be produced with reference lines at 3 x ULN for ALT and AST and 2 x ULN for total bilirubin. Box plots of absolute values and change from baseline values for all haematology and clinical chemistry parameters will also be presented.

Liver biochemistry test results over time for patients who show elevated ALT or AST (≥ 3 x ULN) and elevated bilirubin (≥ 2 x ULN) (elevated results do not need to be present at the same visit) or ALT or AST of ≥ 5 x ULN, will be tabulated and plotted.

4.11.5 Vital Signs (Pulse and Blood Pressure) and Weight

All vital signs data collected will be listed.

Absolute values, change from baseline and percentage change from baseline for pulse, systolic and diastolic blood pressure, body temperature and weight will be summarised by actual treatment group and visit.

A shift table of baseline to maximum and minimum value on treatment for blood pressure and pulse will also be presented using the normal ranges in Table 7.

Table 7 Vital Sign Normal Ranges

Vital Sign	Outside AZ defined reference range lower limit if	Outside AZ defined reference range upper limit if	Treatment emergent decrease if	Treatment emergent increase if
Systolic blood pressure (mmHg)	< 100	> 160	< -30	> 30
Diastolic blood pressure (mmHg)	< 60	> 100	< -15	> 15
Pulse (bpm)	< 40	> 100	< -20	> 20

In addition, box plots of absolute values and change from baseline in blood pressure, pulse, body temperature and weight will also be presented.

4.11.6 Electrocardiogram

All ECG (including Holter) data received will be presented in data listings.

The following ECG parameters will be summarised (absolute values, change from baseline and percentage change from baseline) by visit and actual treatment group: heart rate, QT interval corrected for heart rate using Fridericia's formula (QTcF), RR, PR, QRS, and QT. The average of the three individual tracings will be used in summaries. However, the individual tracings will be displayed in the data listings. For SRC purposes, QT interval corrected for heart rate using Bazett's formula (QTcB) will also be presented.

Overall evaluation on of ECG (including Holter) results will be summarised descriptively.

Box plots of absolute values and change from baseline in ECG parameters over time will be presented.

Shift tables from baseline to maximum and minimum value on-treatment will be presented using the normal ranges in Table 8.

Table 8 ECG Normal Ranges

ECG Parameter	Outside AZ defined reference range lower limit if	Outside AZ defined reference range upper limit if
Heart rate (bpm)	< 40	> 100
RR (msec)	< 600	> 1200
PR (msec)	< 120	> 200
QRS (msec)	< 60	> 109
QT (msec), QTcF (msec) and QTcB (msec)	< 320	> 450

Additional shift tables will be presented using categorised results as specified below:

- Heart rate baseline to minimum value during treatment with categories: ≥ 60 bpm, 50 – 59 bpm, 45 – 49 bpm, 40 – 44 bpm, < 40 bpm.
- CCI baseline to maximum value during treatment with categories: CCI , CCI

CCI (defined as values following study treatment that are greater than CCI or increases from baseline greater than CCI) will be summarised using cumulative counts and percentages under the following categories:

- Absolute value CCI
- Absolute value CCI
- Absolute value CCI
- Change from baseline CCI

- Change from baseline [redacted]
- Absolute value [redacted] and change from baseline [redacted]
- Absolute value [redacted] and change from baseline [redacted]

The number and percentage of patients having any cardiology review items as set out in Table 9 will also be presented.

Table 9 Cardiology Review Items

Category	Review Items		
Rhythm	Artificial pacemaker	Atrial fibrillation	Atrial flutter
	Ectopic supraventricular rhythm	Idioventricular rhythm	Junctional rhythm
	Junctional tachycardia	Sinus pauses	Sinus bradycardia
	Supraventricular tachycardia	Sinus tachycardia	Ventricular fibrillation
	Atrial pacing	Atrial tachycardia	Other
Ectopy	Atrial bigeminy	Atrial premature complexes	Frequent atrial premature complexes (>3)
	Ventricular premature complexes	Frequent ventricular premature complexes (>2)	Ventricular bigeminy
	Ventricular couplets	Ventricular trigeminy	Non-sustained ventricular tachycardia
	Premature junctional complexes	Junctional escape complexes	Other
Conduction	2:1 AV block	AV Mobitz I	AV Mobitz II
	Complete heart block	First degree AV block	Incomplete left bundle branch block
	Incomplete right bundle branch block	Intraventricular conduction defect	Left anterior hemiblock
	Left bundle branch block	Left posterior hemiblock	Prolonged QTc
	Right bundle branch block	Wolff-Parkinson-White	Other
ST Segment	Depressed	Elevated	Other
T Waves	Biphasic	Flat	Inverted
	Notched	Other	
U Waves	Abnormal	T-U Fusion	Other
Myocardial Infarction	Lateral MI 1, L, V5, V6	Septal MI V1, V2, (V3)	Inferior MI (2), 3, F
	Antero Septal MI V1-V4	Extensive Anterior MI 1, L, V1-V6	High lateral MI 1, AVL
	Anterior MI V3, V4	Antero Lateral MI V3-V6	Posterior MI
	Other		
Morphology	Left atrial abnormality	Left ventricular hypertrophy	Low voltage
	Right atrial abnormality	Right ventricular hypertrophy	Brugada syndrome
	Other		

Similarly, cardiology review items will also be presented for Holter ECGs. Cardiology review items for Holter ECGs are:

- Frequent VPCs (number in hours or >30 in a hour)
- Nonsustained ventricular tachycardia (number of episodes)
- Sustained ventricular tachycardia

- Torsade de Pointes
- Ventricular fibrillation or ventricular flutter
- Frequent short episodes of nonsustained supraventricular tachycardia
- Atrial fibrillation
- Atrial flutter
- Mobitz (Wenckebach) 2nd degree AV block
- 2:1 AV block
- High grade AV block
- Mobitz II 2nd degree AV block
- Complete heart block
- Pause >3.0 seconds (longest pause in seconds)
- Average heart rate <40 for any one hour
- Marked sinus bradycardia
- Intermittent ectopic atrial rhythm
- Intermittent junctional rhythm
- Other

4.11.7 Performance Status

The ECOG/WHO performance status will be listed and summarised as frequency counts by actual treatment group and visit.

4.11.8 Echocardiogram

All echocardiogram data received will be presented in data listings.

The following echocardiogram parameters will be summarised (absolute values, change from baseline and percentage change from baseline) by visit and actual treatment group:

- Left ventricular volumes:
  - Left ventricular end-systolic volume (mL)
  - Left ventricular end-diastolic volume (mL)
  - Stroke volume (mL)
- Left ventricular systolic function:
  - Left ventricular ejection fraction (%)
  - Cardiac output (L/min)
- Left ventricular diastolic function:
  - Ratio of Doppler derived E to tissue Doppler derived e' for myocardial velocity (MV/TD E/e')
  - Left atrial volume index (L/min/m2)
  - Lateral early diastolic myocardial velocity (cm/s)
  - Septal early diastolic myocardial velocity (cm/s)
  - Peak tricuspid regurgitation velocity (cm/s)

Overall evaluation of echocardiogram results will be summarised descriptively.

Box plots of absolute values and change from baseline in echocardiogram parameters over time will be presented.

4.11.9 Neurological Assessment

Results for neurological assessment will be listed and summarised as frequency counts by actual treatment group and visit.

4.11.10 Safety Review Committee

A review of the safety data will be conducted by the SRC approximately every six months up to the primary analysis for PFS, and thereafter at the discretion of the SRC. Other safety reviews outside of this may be conducted either at the request of the SRC, based on the emerging profile of the treatment (either within or outside of this clinical trial) or at the Sponsor's discretion. At each review the SRC will review the accumulated safety data available at that point in time. Based on the SRC's review of the data, the SRC will make a recommendation to the AZ study team. This may include to continue the study as planned, to modify the study, or to consider dropping one or more doses from the study based on the tolerability profile.

The SRC will consist of:

- AZ physician who will chair the committee
- AZ safety physician
- Two external investigator clinicians who are independent of this study

The full remit, membership and conduct of the SRC will be described in a separate SRC charter, including outputs to be produced.

Descriptive summaries and plots produced for the SRC will follows the guidelines and methodology detailed in this SAP.

4.12 Other Analyses

4.12.1 Pharmacokinetics

PK concentration data will be collected as per protocol.

Pre- and post-dose concentrations of AZD9833 at steady state will be summarised by AZD9833 dose level using standard summary statistics for PK concentrations (gmean, geometric CV, gmean ± StD, arithmetic mean, StD, minimum, maximum and n). PK concentration data will be listed for each patient in the PK analysis set. If applicable, AZD9833 metabolites will be summarised appropriately.

The gmean will be calculated as e^{μ} where μ is the mean of the data on the logarithmic scale.

The geometric CV will be calculated as $100 \times \sqrt{\exp s^2 - 1}$ where s is the StD of the data on the logarithmic scale.

gmean ± StD will use the StD derived as $\exp(\mu + s)$ where μ and s has been defined as above.

Non-quantifiable (NQ) values of plasma concentrations will be handled as follows:

- If, at a given timepoint, 50% or less of the plasma concentrations are NQ, the gmean, geometric CV, gmean ± StD, arithmetic mean and StD will be calculated by substituting the LLoQ for values which are NQ.
- If more than 50%, but not all, of the values are NQ, the gmean, geometric CV, gmean ± StD, arithmetic mean and StD will be reported as not calculable (NC).
- If all the concentrations are NQ, the gmean and arithmetic mean will be reported as NQ and the geometric CV, gmean ± StD and StD as NC.
- The number of values above LLoQ will be reported for each timepoint along with the total number of collected values.

If data are available for less than three patients, no summary statistics other than minimum, maximum and n will be presented.

Box plots of concentration data by sample time will be displayed.

4.12.2 Pharmacodynamics

Pharmacodynamic data will be collected as per protocol.

The following secondary pharmacodynamics biomarker variables will be calculated:

- Change and percent change from baseline in ER expression as assessed by the manual H-score method
- Change and percent change from baseline in PgR expression as assessed by the manual H-score method
- Change and percent change from baseline in Ki67 index as assessed by manual method.

Biopsy samples will be sectioned and assessed for tumour content, and IHC staining for ER, PgR and Ki67 protein biomarkers will be carried out at AZ and/or contracted external good clinical practice accredited laboratory. Sections will be scored manually for ER, PgR and Ki67 determination in tumour by an academic/ contract research organisation and/or AZ pathologist/ AZ contracted independent pathologist.

ER and PgR expression will be determined by the percentage of positively stained tumour nuclei in each staining category, i.e., negative '-/-', weak '+', moderate '++', strong '+++'. The sum of

the percentages of these 4 staining categories will be 100%. Then, ER and PgR expression will be assessed using the following manual H-score method, which yields to a manual H-score range of 0 to 300:

$$\text{H-score} = (1 \times \%\text{ of } +) + (2 \times \%\text{ of } ++) + (3 \times \%\text{ of } +++)$$

When a biopsy contains less than 100 tumour cells, then the ER and PgR H-score is not reported, and NE will be presented.

The Ki67 labelling index will be assessed and expressed as the average percentage of tumour cells with counted positive nuclei, following the International Ki67 in Breast Cancer Working Group Recommendations. Ki67 percentage will be scored by increments of 1%. When a biopsy counted contains less than 100 tumour cells, then Ki67 percentage is not reported and NE will be presented.

The secondary endpoints; change and percentage change from baseline in ER and PgR expressions as assessed by the manual H-score method and change and percentage change in Ki67 labelling index will be summarised appropriately. An ANCOVA model will be fitted to the percentage change from baseline for each endpoint with baseline value, day of biopsy, stratification factors as described in Section 4.10.1.2, and treatment group included in the model. Least squares (LS) estimates of the treatment effect will be calculated together with 90% CIs.

If data does not adequately follow a normal distribution, then the use of the natural log transformed data (i.e., ratio) will replace the untransformed analysis as the primary approach. In case of ANCOVA on natural log transformed data, the estimated treatment effect and its 90% CIs will be back transformed to the original scale. Values equal to zero on the original scale will be set to the LLOQ for the natural log transformation.

Normality distribution of data will be tested by means of a Shapiro-Wilk test. In case of a two-sided p-value > 0.20 there will be evidence that the data tested are normally distributed. Consultation of plots and additional statistics could help to assess the severity of the deviations from normality.

If necessary, a nonparametric method will be performed using the Hettmansperger and McKean nonparametric ANCOVA (Nakonezny et al 2007).

It is expected that the distribution of Ki67 will not be normally distributed (Robertson et al 2013), hence, Ki67 index data will be naturally log transformed before being analysed.

Analyses may be repeated for the subgroup of patients with baseline % ER positivity > 10%.

Only listings will be presented if an inadequate amount of data is available for summaries.

4.12.3 Patient Reported Outcome (PRO) Measures

Patient reported outcome (PRO) measures will be used to examine the impact of treatment on symptoms, functioning, and HRQoL and aid in understanding the benefit/risk evaluation from the patient's perspective.

PROs will be assessed using the following measures:

- EORTC QLQ-C30
- EORTC QLQ-BR23
- NEI VFQ-25
- EQ-5D-5L
- CCI
- CCI
- CCI
- CCI

4.12.3.1 EORTC QLQ-C30 and EORTC QLQ-BR23

The EORTC-QLQ-C30 was developed to assess HRQoL, functioning, and symptoms in cancer clinical trials.

The EORTC QLQ-C30 consists of 30 questions, which can be grouped to produce five multi-item functional scales, three multi-item symptom scales, six individual items (five items assessing additional symptoms commonly reported by cancer patients and one item on the financial impact of the disease) and a two-item global measure of health status/quality of life:

- Functional scales:
  - Physical functioning
  - Role functioning
  - Emotional functioning
  - Cognitive functioning
  - Social functioning
- Multi-item symptom scales:
  - Fatigue
  - Nausea and vomiting
  - Pain
- Individual items:
  - Dyspnoea
  - Insomnia
  - Appetitive loss
  - Constipation
  - Diarrhoea
  - Financial difficulties
- Global health status/quality of life

The EORTC QLQ-BR23 is a breast cancer specific module used in conjunction with the core QLQ-C30 to assess breast cancer-specific HRQoL. It comprises 23 questions which can be grouped into five multi-item scores (body image, sexual functioning, arm symptoms, breast symptoms, and

systemic therapy side effects) and single items on sexual enjoyment, upset by hair loss and future perspective. This instrument can be combined in four functional scales and four symptom scales:

- Functional scales:
  - Body image
  - Sexual functioning
  - Sexual enjoyment
  - Future perspective
- Symptom scales
  - Systemic therapy side effects
  - Breast symptoms
  - Arm symptoms
  - Upset by hair loss

An outcome variable consisting of a score from 0 to 100 will be derived for each of the functional scales, symptom scales, individual items and global health status/ QoL in the EORTC QLQ-C30 and for each of the functional and symptom scales for the EORTC QLQ-BR23 according to the EORTC QLQ-C30 Scoring Manual and EORTC QLQ-BR23 instructions.

Higher scores on the global health status/ QoL and functioning scales indicate better health status/functioning. Higher scores on the symptom scales indicate greater symptom burden.

For all scales, if at least half the components of a scale are present for a timepoint then the score will be calculated, otherwise the score will be set to missing.

The principle for scoring the scales of the QLQ-C30 and QLQ-BR23 are:

- Estimate the average of the items that contribute to the scale (raw score)
- Use a linear transformation to standardise the raw score, so that the scores range from 0 to 100

In practical terms, if terms $I_1, I_2, \ldots, I_n$ are included in a scale, then:

$$RS = \frac{\sum_{i=1}^{n} I_i}{n}$$

where RS denotes the raw score.

All functional scales are then calculated as $S = \left\{1 - \frac{RS - 1}{range}\right\} \times 100$ and symptom scales, items and global health status as $S = \left\{\frac{RS - 1}{range}\right\} \times 100$.

Range is defined as the difference between the maximum possible value of RS and minimum possible value of RS.

The global health status/ QoL, functional scales and symptom scales, including the items included in each of these scales are presented in Table 10 and Table 11.

Table 10 Scoring the QLQ-C30

	Scale	Number of Items	Item Range [a]	Item Numbers
Global Health Status				
Global Health Status	QL2	2	6	29, 30
Functional Scales				
Physical functioning	PF2	5	3	1 to 5
Role functioning	RF2	2	3	6, 7
Emotional functioning	EF	4	3	21 to 24
Cognitive functioning	CF	2	3	20, 25
Social functioning	SF	2	3	26, 27
Symptom Scales/ Items				
Fatigue	FA	3	3	10, 12, 18
Nausea and vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17
Financial difficulties	FI	1	3	28

[a] Item range is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving range = 3.

Table 11 Scoring the BR-23

	Scale	Number of Items	Item Range [a]	Item Numbers
Functional Scales				
Body image	BRBI	4	3	9 to 12
Sexual functioning [d]	BRSEF	2	3	14, 15
Sexual enjoyment [b,d]	BRSEE	1	3	16
Future perspective	BRFU	1	3	13
Symptom Scales/ Items				
Systemic therapy side effects	BRST	7	3	1 to 4, 6, 7, 8

	Scale	Number of Items	Item Range [a]	Item Numbers
Breast symptoms	BRBS	4	3	20 to 23
Arm symptoms	BRAS	3	3	17 to 19
Upset by hair loss [c]	BRHL	1	3	5

[a] Item range is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving range = 3.
[b] Not applicable if item 15 is "not at all".
[c] Not applicable if item 4 is "not at all".
[d] Items for these scales are scored positively (i.e., "very much" is best) and therefore use the same scoring methodology as for symptom scales

For all EORTC QLQ-C30 and BR23 scores, a clinical meaningful change or difference in score will be defined as a change of at least 10 points (Osoba et al. 1998). Specifically, a clinically meaningful improvement in a symptom will be defined as a decrease in the score from baseline of ≥ 10, whereas a clinical meaningful deterioration will be defined as an increase in the score from baseline of ≥ 10. In contrast, a clinical meaningful improvement in a functional scale or global health status/ QoL score will be defined as an increase in the score from baseline of ≥ 10, while a clinically meaningful deterioration in a functional scale or global health status/ QoL will be defined as a decrease in the score from baseline of ≥ 10.

Descriptive statistics for subscale scores, including changes from baseline will be presented by visit and treatment group. Plots of mean change from baseline in total and subscale scores over time will also be presented with indicators for the number of patients at each visit. In addition, the proportion of patients experiencing improvement, no change or worsening from baseline will be presented by visit and treatment group. Improvement, no change and worsening will be defined as follows for all subscales:

Score	Change from baseline	Visit response
Symptom scales/items	≥+10	Worsened
	≤-10	Improved
	Otherwise	No change
Functional scales and global health status/QoL	≥+10	Improved
	≤-10	Worsened
	Otherwise	No change

QoL Quality of life

Change from baseline in the EORTC QLQ-C30 global health status/ QoL, physical functioning and role functioning scores will be analysed using a mixed model for repeated measures (MMRM). Additionally, mean change from baseline averaged across visits up to six months will be examined for each treatment. Other scales or items may be analysed similarly if deemed necessary.

It is acknowledged that patients will discontinue treatment at different timepoints during the study and this is an important time with regards to PRO data collection. To account for this and in order to include discontinuation and safety follow-up visits, PRO data will be assigned to visits as discussed in Section 4.3.2.

The MMRM model will include patient as a random effect, treatment, windowed visit, treatment-by-visit interaction and stratification factors as described in Section 4.10.1.2 as exploratory variables and the baseline PRO score as a covariate. Treatment, visit, treatment-by-visit interaction, and the stratification factors will be fixed effects in the model. Restricted maximum likelihood (REML) estimation will be used. An overall adjusted mean estimate will be derived that will estimate the average treatment effect over visits giving each visit equal weight. Additionally, estimates will also be presented for each visit. For this overall treatment comparison, adjusted mean estimates per treatment group and corresponding 90% CIs will be presented along with an estimate of the treatment difference and its 90% CI. No p-values will be presented. The treatment-by-visit interaction will remain in the model regardless of significance.

An unstructured covariance matrix will be used to model the within-patient error and the Kenward-Roger approximation will be used to estimate the degree of freedom. If the fit of the unstructured covariance matrix fails to converge, the following covariance structures will be tried in order until convergence is reached: toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, and autoregressive. If there are still issues with the fit of the model or estimation of the treatment effects, patient will be treated as a fixed effect.

4.12.3.2 NEI VFQ-25

The NEI VFQ-25 consists of 25 vision-targeted questions representing 11 vision-related subscales, plus an additional single-item general health rating question. The subscales (including the general health rating question) include:

- Global vision rating
- Difficulty with near vision activities
- Difficulty with distance vision activities
- Limitations in social functioning due to vision
- Role limitations due to vision
- Dependency on others due to vision
- Mental health symptoms due to vision
- Driving difficulties
- Limitations with peripheral vision
- Limitations with colour vision
- Ocular pain
- General health rating question

An outcome variable consisting of a score from 0 to 100 will be derived for each of the subscales as per the NEI VFQ-25 scoring manual.

Scoring of the NEI VFQ-25 will be done as follows:

- Original results from the survey will be recoded following the scoring rules in Table 12.
- Once items are recoded, items within each subscale are averaged to create the 11 sub-scale scores and the score for the general health rating.

Table 12 Recoding NEI VFQ-25 Items

Item Numbers	Original response category	Recoded value
1, 3, 4, 15c [a]	1	100
	2	75
	3	50
	4	25
	5	0
2	1	100
	2	80
	3	60
	4	40
	5	20
	6	0
5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 16a	1	100
	2	75
	3	50
	4	25
	5	0
	6 [b]	*
17, 18, 19, 20, 21, 22, 23, 24, 25	1	0
	2	25
	3	50
	4	75
	5	100

[a] Item 15c has four-response levels, but is expanded to five levels using item 15b.
Note: If 15b = 1, then 15c should be recoded to "0". If 15b = 2 or 3, then 15 c should be recoded to "missing".
[b] Response choice "6" indicates that the person does not perform the activity because of non-vision related problems. If this choice is selected, the item is coded as "missing".

The grouping of individual NEI VFQ-25 scores to generate subscales are detailed in Table 13.

Table 13 Subscales of NEI VFQ-25

Scale	Number of Items	Item Numbers
General health	1	1
General vision	1	2
Ocular pain	2	4, 19
Near activities	3	5, 6, 7
Distance activities	3	8, 9, 14

Table 13 Subscales of NEI VFQ-25

Scale	Number of Items	Item Numbers
Vision specific:		
Social functioning	2	11, 13,
Mental health	4	3, 21, 22, 25,
Role difficulties	2	17, 18
Dependency	3	20, 23, 24
Driving	3	15c, 16, 16a
Colour vision	1	12
Peripheral vision	1	10

Higher scores represent a better possible outcome with lower scores worse outcomes.

Descriptive statistics for subscale scores, including changes from baseline will be presented by visit and treatment group. Plots of mean change from baseline in subscale scores over time will also be presented with indicators for the number of patients at each visit.

Missing values and MMRM analyses (if deemed necessary) of NEI VFQ-25 will be done using similar methodology as described in Section 4.12.3.1.

4.12.3.3 EQ-5D-5L

The EQ-5D-5L will be used to explore the impact of treatment and disease state on health state utility for use in economic analyses. The questionnaire comprises six questions that cover 5 dimensions of health:

- Mobility
- Self-care
- Usual activities
- Pain/discomfort
- Anxiety/depression

For each dimension, patients select which statement describes their health on that day from a possible five options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems). A unique EQ-5D health state, termed the EQ-5D-5L profile, is reported as a five-digit code with a possible 3,125 health states. For example, state 11111 indicates no problems on any of the five dimensions.

Patients will also assess their health using the EQ-VAS which ranges from 0 (worst imaginable health) to 100 (best imaginable health).

The EQ-5D profile will be converted into a weighted health state utility value, termed the EQ-5D index, by applying a country-specific equation to the EQ-5D-5L profile that represents the comparative value of health states. The equation is based on national valuation sets elicited from the general population and the base case will be the United Kingdom perspective. Where a

valuation set has not been published, the EQ-5D-5L profile will be converted to the EQ-5D index using a crosswalk algorithm (van Hout et al. 2012).

Results from the EQ-5D-5L will be summarised descriptively by visit, including a shift table from baseline to worst on treatment for each dimension of health. Worst on treatment is defined as the highest numerical state for each dimension scored while on treatment (e.g., 2 is worse than 1). Plots of mean change from baseline in EQ-VAS and health state utility value over time will also be presented with indicators for the number of patients at each visit.

The EQ-5D index values and EQ-VAS will in addition be analysed with a MMRM model as described in Section 4.12.3.1.

CCI

The CCI over the past week. The item is rated using a CCI scale from CCI CCI

The CCI is a CCI assessing the patient's CCI since the start of study treatment. The item is rated using a CCI from CCI to CCI

The CCI is a CCI assessing the CCI . The item is rated using a CCI from CCI to CCI CCI .

The CCI is a CCI assessing the CCI The CCI items assess: CCI Items are rated on CCI or CCI scales.

CCI and CCI will be summarised descriptively by visit. No formal statistical comparisons will be conducted on these results.

4.12.3.5 PRO Compliance

Summary measures of overall compliance and compliance over time will be derived for each PRO respectively. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.
- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g., a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under PRO follow-up at the specified

assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.

- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that is non-missing.
- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all timepoints, divided by the total number of questionnaires expected to be received across all timepoints multiplied by 100.
- Overall patient compliance rate is defined for each randomised treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by 100.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire at the timepoint (as defined above), divided by the number of patients still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires (per definition above), divided by the number of received questionnaires.

4.13 Determination of Sample Size

A sample size of approximately 288 patients, randomised in equal proportions to the four treatment groups will be required to observe a total of at least 108 PFS events for each pairwise comparison against fulvestrant. As of December 2020, and the decision to stop enrolment to the AZD9833 300 mg arm, the resulting total sample size across the three enrolling treatment arms is approximately 216, with approximately an additional 20 patients from the AZD9833 300 mg arm enrolled up to recruitment closure of this arm. The primary analysis will only be triggered when a minimum of 108 events is observed for the pairwise comparison of 75 mg and 150 mg AZD9833 dose versus fulvestrant.

A HR of 0.59 for each pairwise treatment comparison versus fulvestrant is of interest. Under the assumption that a 5-month median PFS will be observed on fulvestrant, that is equivalent to a 3.5-month increase in median PFS over fulvestrant. A minimum of 108 events for the pairwise comparison of each AZD9833 dose of interest (75 mg and 150 mg) versus fulvestrant will provide 86% power at the 2-sided 10% significance level if the assume treatment effect is HR=0.59.

The primary endpoint will be based on the investigator assessment of progression. A sensitivity analysis based on the BICR of scans will be conducted. Due to the different methods of declaring progression, it is expected that the number of events declared via the two methods will be different.

Sample size within subgroups of patients with/without prior use of CDK4/6 inhibitors

Patients will be stratified for prior use of CDK4/6 inhibitors. Assuming the cap limit is maintained, the study will randomize between 32 and up to a maximum of 40 patients per stratum in each treatment group. Under this scenario, approximately equal proportions (i.e., 36 patients per stratum in each treatment group) are expected.

Controlling the sample size in this way will allow for approximately 54 events for the pairwise comparisons of interest at the primary analysis (under the assumption that the event rates are the same in the subgroups of patients with or without prior use of CDK4/6 inhibitor). Fifty-four events for the pairwise comparison of each AZD9833 dose versus fulvestrant will provide 86% power at the 1-sided 20% significance level if the assumed true treatment effects are HR=0.59.

4.14 Changes in the Conduct of the Study or Planned Analysis

The requirement to have post-dose data available for inclusion into the safety analysis set (CSP Section 9.3) has been removed for analysis purposes. The reason for removal is to ensure that all safety events and data are accounted for and the fact that all patients who took any dose were at risk of having a safety event.

5 REFERENCES

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45(2):228-47

Fleisher F, Gaschler-Markefski B, Bluhmki E. How is retrospective independent review influenced by investigator-introduced informative censoring: A quantitative approach. Stat Med. 2011; 30(29):3373-86

Fleming TR and Harrington DP. Counting Process and Survival Analysis New York, NY: John Wiley & Sons (1991).

Gail M, Simon R. Testing for qualitative interactions between treatment effects and subject subsets. Biometrics. 1985; 41:361-72

Nakonezny PA, Shull RD. Hettmansperger and Mckean linear model aligned rank test for the single covariate and one-way ANCOVA case (SAS). Journal of Modern Applied Statistical Methods. 2007; 6:336-340

Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality of life scores. J Clin. Oncol. 1998; 16:139-144

Robertson JF, Dixon JM, Sibbering DM, Jahan A, Ellis IO, Channon E, et al. A randomized trial to assess the biological activity of short-term (pre-surgical) fulvestrant 500 mg plus anastrozole versus fulvestrant 500 mg alone or anastrozole alone on primary breast cancer. Breast Cancer Res 20136; 15R18.

Sun X, Chen C. Comparison of Finkelstein's method with the conventional approach for interval-censored data analysis. Stat Biopharm Res. 2010; 2:97-108

van Hout B, Janssen MF, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012; 15(5):708-15

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d8530c00002-sap-v5		
Document Title:	Statistical Analysis Plan Version 5	
Document ID:	Doc ID-004235579	
Version Label:	5.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
18-Aug-2022 07:48 UTC	PPD	Content Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.