MONETTE: A Randomised, Open-Label, Phase 2 Study of Ceralasertib Monotherapy and Ceralasertib plus Durvalumab in Patients with Unresectable or Advanced Melanoma and Primary or Secondary Resistance to PD-(L)1 Inhibition

ClinicalTrials.gov Identifier: NCT05061134

Statistical Analysis Plan: Final 4.0, 12 Mar 2024

STATISTICAL ANALYSIS PLAN	
Study Code	D533AC00001
Edition Number	4.0
Date	12-Mar-2024

MONETTE: A Randomised, Open-Label, Phase 2 Study of Ceralasertib Monotherapy and Ceralasertib plus Durvalumab in Patients with Unresectable or Advanced Melanoma and Primary or Secondary Resistance to PD-(L)1 Inhibition

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or Specialized Term	Definition
ADA	Anti-Drug Antibody
AE	Adverse event
AESI	Adverse Event of Special Interest
ALT	Alanine Aminotransferase
ARID1A	AT-rich interactive domain-containing Protein 1A
AST	Aspartate Aminotransferase
ATM	Ataxia Telangiectasia Mutated
ATR	Ataxia Telangiectasia and Rad3-Related Protein
BICR	Blinded Independent Central Review
BILI	Total Bilirubin
BMI	Body mass index
BOR	Best Objective Response
BRAF	B-Rapidly Accelerated Fibrosarcoma Gene
BSR	Baseline Scaled Ratio
CI	Confidence Interval
COVID-19	Coronavirus Disease 2019
CR	Complete Response
CRF	Case Report Form
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for AEs
CV	Coefficient of Variation
DCO	Data Cut-Off
DCR	Disease Control Rate
DoR	Duration of Response
DRM	Data Review Meeting
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
CCI [redacted]	CCI [redacted]
CCI [redacted]	CCI [redacted]
CCI [redacted]	CCI [redacted]
CCI [redacted]	CCI [redacted]

GCP	Good Clinical Practice
gCV	Geometric Coefficient of Variance
Gmean	Geometric Mean
gSD	Geometric Standard Deviation
HR	Hazard Ratio
CCI	CCI
IO	Immune-Oncology
ICF	Informed Consent Form
IDMC	Independent Data Monitoring Committee
IL	Item Library
IP	Investigational Product
IPD	Important Protocol Deviation
KIT	Proto-oncogene c-KIT
LDH	Lactic Acid Dehydrogenase
LLOQ	Lower Limit of Quantification
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary for Regulatory Activities
max	Maximum
min	Minimum
MMRM	Mixed Model for Repeated Measures
MSI	Activating Mutations and Microsatellite Instability
MUGA	Multigated Acquisition
NA	Not Applicable
NE	Not Evaluable
NF1	Neurofibromatosis type 1
NTL	Non-Target Lesion
NR	Not Reportable
NRAS	NRAS proto-oncogene, GTPase
NS	No Sample
OAE	Other significant adverse event
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease
PD-(L)1	Programmed Cell Death 1 Ligand
PFS	Progression Free Survival
CCI	CCI

CCI	CCI
CCI	CCI
PK	Pharmacokinetics
PR	Partial Response
PRO	Patient Reported Outcome
CCI	CCI
PT	Preferred Term
RDI	Relative Dose Intensity
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable Disease
SE	Standard Error
SMQ	Standardized MedDRA Queries
SOC	System Organ Class
Std Dev	Standard Deviation
TL	Target Lesion
TLF	Tables Listings Figures
TACE	Transarterial Chemoembolization
TARE	Transarterial Radioembolization
CCI	CCI
TEAE	Treatment Emergent Adverse Events
CCI	CCI
TOC	Table of Contents
TTR	Time to Objective Response
ULN	Upper Limit of Normal

AMENDMENT HISTORY

CATEGORY **Change refers to:**	**Date**	**Description of change**	**In line with CSP?**	**Rationale**
N/A	Click or tap to enter a date.	Initial approved SAP	N/A	N/A
Typo and missing author signature	7/27/2022	Fixed typo in 4.2.8.2.2 and added missing author signature which was missing in ed 1.0	yes	See description.
Update to interim analyses details	6/16/2023	Section 3.3 updated to cover additional supportive analysis which will be included for the interim analyses	N/A	See description.
Biopsy sub-study analysis	11/15/2023	Details of the biopsy sub-study have been removed from this SAP	yes	The biopsy sub-study will be covered in a separate SAP
Layout	11/15/2023	SAP layout updated for consistency with AZ Late Oncology TA SAP	Yes	Updated as Study now covered by AZ Late Oncology team
Timing of Analyses	11/15/2023	Descriptions of triggers for analyses to be performed updated	Yes	To reflect what was required for each analysis
Visit window information	11/15/2023	Table outlining visit windows removed The rule for calculating visit windows updated where there is an even number of days between visits	Yes	Inconsistency withing the SAP
Analysis Set definitions	11/15/2023	Added definitions of analysis sets used for interim analyses and corrected definitions for others	Yes	To endure all analysis sets used for tables, listings and figures were clearly defined within SAP

Important Protocol Deviations	11/15/2023	Added categories presented under AZ standard IPD table shell	Yes	updated to show how individual IPDs could be mapped to standard table
Definition of Best Objective Response added	11/15/2023	Added section defining endpoint and analysis to be performed	Yes	This is needed for determining Objective Response Rate
Definition of Disease Control Rate endpoint added	11/15/2023	Added section defining endpoint and analysis to be performed	Yes	This was interim analysis endpoint defined in CSP but previously missing from SAP
Exposure	11/15/2023	Duration of exposure and total actual exposure calculations updated Summary of dose omissions removed	Yes	Updated to be consistent with approach used on other ceralasertib studies
<Adverse Event - delete> Missing date imputations	11/15/2023	Imputation rules for missing dates updated for consistency with AZ Late Oncology TA SAP Imputation for missing end dates of ceralacertib added	Yes	Updated as Study now covered by AZ Late Oncology team
Analysis of biomarker endpoints	11/15/2023	Methods of summarising/reporting data updated	Yes	Updated following internal AZ discussion

4.3, 4.4.4, 4.4.1	03/12/2024	Removal of subgroup analyses for TL size, reduction of sensitivity analyses, making biomarker outputs optional, CCI [redacted] from exploratory analysis, imputation for completely unknown death date, subgroup analyses reduced.	yes	Reduction of outputs

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of the main study of D533AC00001 supporting the clinical study report (CSR). The reader is referred to the clinical study protocol (CSP) and the case report form (CRF) for details of study conduct and data collection. This statistical analysis plan (SAP) is based on Amendment *2* of the CSP dated 14 March 2022. In the event of future amendments to the protocol, this SAP may be modified to account for changes relevant to the statistical analysis.

A separate SAP will be created for the biopsy sub-study

2 CHANGES TO PROTOCOL PLANNED ANALYSES

Not applicable

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

The details of the analyses planned during the conduct of this study are outlined below and summarised in the following table.

Table 1: Summary of planned analyses

Analysis	Trigger	Scope
Interim Analysis[CCI]	Once[CCI] subjects are treated in the main study and have approximately[CCI] weeks of follow-up	Safety data, efficacy data excluding overall survival (OS), Study Conduct until DCO
Interim Analysis[CCI]	Once[CCI] subjects are treated in the main study and have approximately[CCI] weeks of follow-up	Safety data, efficacy data excluding OS, Study Conduct until DCO
Primary Analysis	Once all randomised subjects have had a minimum of[CCI] months of follow-up after start of study intervention or discontinued from study, whichever occurs first.	All available data

Analysis	Trigger	Scope
	May be postponed to ensure a minimum of [redacted] months median duration of follow-up.	
Repeated Primary Analysis	It is anticipated at least approximately [redacted]% of all randomised subjects will have progressed or died at the time of the primary analysis. However, if this is not the case, a later analysis for Progression Free Survival (PFS) may be conducted once approximately [redacted]% of all randomised subjects have progressed or died.	Optional Analysis
[redacted] Analysis	Once approximately [redacted]% of all randomised subjects have died or [redacted] years from when the last subject starts treatment, whichever is earliest.	All available data

DCO: Data cut-off

Details of the interim analyses are given in Section 5.

The primary analysis of the ORR is conducted once all randomised subjects have had a minimum of [redacted] months of follow-up after start of study treatment or have discontinued from the study, whichever occurs first. It is anticipated that the median duration of follow up for DoR is at least [redacted] months at this point, however if not, then the primary analysis may be postponed to ensure a minimum of [redacted] months median duration of follow up. At this primary analysis the CSR is written.

It is anticipated at least approximately [redacted]% of all randomised subjects have progressed or died at the time of the primary analysis. However, if this is not the case, a later analysis for Progression Free Survival (PFS) may be conducted once approximately [redacted]% of all randomised subjects have progressed or died. If later analyses are required an appendix to the SAP will be written.

A [redacted] analysis is conducted once approximately [redacted]% of all randomised subjects have died, or [redacted] years from when the last subject starts treatment, whichever is earliest. If the [redacted] analysis is performed [redacted] years from when the last subject starts treatment due to less deaths than assumed and [redacted]% of PFS events have not been reached, then only the [redacted] analysis will be performed i.e. the primary analysis will not be repeated after the [redacted] analysis. The primary and [redacted] analyses may be combined if the required PFS and OS events are due to be reached at a similar time.

3.2 General Considerations

The below-mentioned general principles will be followed throughout the study:

- Summary tables are produced by treatment group. Total columns are produced for non-efficacy tables.
- Descriptive statistics will be used for all variables, as appropriate.
    - Continuous variables will be summarised by the number of observations, mean, standard deviation, median, upper, and lower quartiles where indicated, minimum and maximum. For log-transformed data (ie, pharmacokinetic [PK] concentration data) it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum, and maximum.
    - If data are available for less than 3 subjects, no summary statistics other than minimum, maximum and number of observations are presented.
    - Categorical variables will be summarised by frequency counts and percentages for each category. Unless otherwise stated, percentages will be calculated based on the analysis set population total, by treatment group and by timepoint as appropriate.
- For continuous data, the mean, standard deviation and median are rounded to one additional decimal place compared to the original data. Minimum and maximum are displayed with the same accuracy as the original data.
- Time-to-event variables will be presented using the Kaplan-Meier (KM) methodology, including median time calculated from the KM curves.
- For categorical data, percentages are rounded to one decimal place with the exception of 100% which is presented as a whole number.
- SAS® version 9.4 (or higher) and/or other validated software (as appropriate) will be used for the analyses. In case R is used, the R build version and the version of each used R package will be documented.

3.2.1 Sample Size Determination

Confirmed objective response rate is the primary endpoint in this study. The primary analysis will be conducted once all randomized patients have had a minimum of [CCI] months of follow-up after start of study treatment or have discontinued from the study, whichever occurs first. It is anticipated that the median duration of follow up will be at least [CCI] months at this point, however if not, then the primary analysis may be postponed to ensure a minimum of [CCI] months median duration of follow up. At this primary analysis the CSR will be written.

Each treatment arm is experimental and thus will be analyzed separately for ORR. The study will recruit approximately 100 patients in the ceralasertib and durvalumab combination therapy arm. With 100 patients the expected [CCI] width of the [CCI]% CI for ORR will be up to approximately [CCI]% when the proportion of patients with an objective response is in the range [CCI]% to [CCI]%. For example, ORR [CCI]%, [CCI]% CI [CCI]%, [CCI]%. The study will recruit approximately 50 patients in the ceralasertib monotherapy arm. With 50 patients, the expected [CCI] width of the [CCI]% CI for ORR will be up to approximately [CCI]% when the proportion of patients with an objective response is in the range [CCI]% to [CCI]%. For example, ORR [CCI]%, [CCI]% CI [CCI]%, [CCI]%. The sample size of the ceralasertib monotherapy arm may be expanded up to a total of approximately 100 patients. This would occur in the event an efficacy signal is observed, and subject to a protocol amendment.

There is no standard of care after anti-PD-(L)1 immunotherapy for the intended patientpopulation. Objective response rate for chemotherapy treatment options for patient second line and beyond, post-ipilimumab, are previously reported to be 4% (95% CI 2%, 9%) and 10% (95% CI 5%, 16%). There is no prospective clinical trial data on response rates forchemotherapy specifically in the post CPI-resistant setting. An ORR of [CCI]% is considered an upper limit of expected response rate for alternative available treatment options for theintended patient population in this study. Per the proposed Phase 2 study design, with100 evaluable patients in the ceralasertib and durvalumab combination therapy arm, successwill be achieved if a minimum ORR of [CCI]% is observed, for which the [CCI]% exact CI wouldbe ([CCI]%, [CCI]%).

A secondary objective is to compare the ORR between the two treatment arms. With150 patients (100 patients randomized to ceralasertib and durvalumab combination therapyarm and 50 patients randomized to ceralasertib monotherapy), there will be at least [CCI]% power at a [CCI] of [CCI] to detect a difference in ORR, assuming an ORR for ceralasertib plus durvalumab of [CCI]% and an ORR for ceralasertib monotherapy of [CCI]%.

It is anticipated at least approximately [CCI]% of all randomized patients will have progressed or died at the time of the primary analysis. However, if this is not the case, a later analysis for PFS may be conducted once approximately [CCI]% of all randomized patients have progressed or died.

A [CCI] analysis will be conducted once approximately [CCI]% of all randomized patients have

died, or [redacted] years from when the last patient starts treatment, whichever is earliest.

3.2.2 Study Treatment

The main study is a Phase 2, randomised, open-label, multicentre, international study assessing the efficacy and safety of ceralasertib and ceralasertib plus durvalumab in subjects with unresectable or advanced melanoma and primary or secondary resistance to programmed death-ligand 1 [PD-(L)1] inhibition. Approximately 150 subjects are randomised, at a 2:1 randomisation ratio (approximately 100 subjects in Arm 1 and approximately 50 subjects in Arm 2) to the following intervention groups:

1. Arm 1: Ceralasertib 240 mg BD oral tablets, from Days 1 to 7 plus durvalumab 1500 mg Day 8, IV infusion, Q28D, until progressive disease, (confirmed RECIST 1.1-defined radiological progression), unacceptable toxicity, withdrawal of consent, or if a study treatment discontinuation criteria is met.

2. Arm 2: Ceralasertib 240 mg BD oral tablets, from Days 1 to 7, Q28D, until progressive disease, (confirmed RECIST 1.1-defined radiological progression), unacceptable toxicity, withdrawal of consent, or if a study treatment discontinuation criterion is met.

Crossover within the study is not be permitted. There is no defined maximum treatment duration for the treatment arms as the subjects receive study treatment until progressive disease (RECIST 1.1-defined radiological progression), unacceptable toxicity, withdrawal of consent, or if a study treatment discontinuation criterion is met.

Randomisation into the treatment arms is stratified by:

• Resistance to prior immune-oncology treatment (primary/early relapse in adjuvant setting vs secondary resistance).

• Baseline lactate dehydrogenase (LDH) expression (below and equal to the upper limit of normal vs above upper limit of normal).

3.2.3 Baseline

In general, the last observed measurement prior to first dose of study intervention will be considered the baseline measurement. For efficacy variables, baseline is defined as the last observed measurement prior to randomisation

For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, serves as sufficient evidence that the assessment occurred prior to first dose. Assessments on the day of the first dose where neither time nor a nominal predose indicator are captured is considered prior to the first dose if such procedures are

required by the protocol to be conducted before the first dose. If no value exists before the first dose/administration, then the baseline value is treated as missing..

If two measurements are equally eligible to assess patient status at baseline (eg, two assessments both on the same date with no time recorded), the average is used as the baseline value. For non-numeric laboratory tests (ie, some of the urinalysis parameters)where taking the average is not possible, the best value (value closest to none/normal/negative) is used as baseline as this is most conservative. In the scenario where there are two eligible baseline assessments recorded on the same day, one with time recorded and the other without time recorded, the one with the time recorded is selected as baseline. Where safety data will be summarised over time, time on study is calculated in relation to date of first treatment.

In all summaries change from baseline variables are calculated as the post-treatment value minus the value at baseline. The percentage change from baseline is calculated as (post-baseline value - baseline value)/(baseline value) x 100. For any endpoint subjected to log transformation, the change from baseline calculated and summarised on the log scale are back-transformed and presented as a 'baseline scaled ratio' (BSR). Percentage change is then calculated as (BSR - 1) × 100.

3.2.4 On-treatment

For the purposes of summarizing safety data assessed at visits, in addition to baseline data, only on-treatment data are included in the summary tables. On-treatment data is defined as data on/after the first dose of study treatment and with assessment date up to and including the date of last study treatment + 30 days (safety follow-up) for ceralasertib monotherapy or the date of last study treatment + 90 days (safety follow-up) for ceralasertib and durvalumab combination, and prior to the start of any subsequent anti-cancer therapy.

The on-treatment period therefore comprises of the treatment period (first administration of any study treatment to and including last day of administration of any study treatment) and the safety follow-up period (30 days after last administration for ceralasertib monotherapy or 90 days after last administration for ceralasertib and durvalumab combination), and prior to the start of any subsequent anti-cancer therapy.

3.2.5 Visit Window

For safety, time windows are defined for any presentations that summarise values by visit. The following conventions apply:

- For safety data study day reference is date of first dose of study treatment as Day 1, for PK the reference is the time of study treatment administration on the day PK blood

samples are taken, for efficacy and PRO data study day references date of randomisation as Day 1.

- The time windows are exhaustive so that data recorded at any timepoint has the potential to be summarised. Inclusion within the time window are based on the actual date and not the intended date of the visit.
- All unscheduled visit data have the potential to be included in the summaries.
- The window for the visits following baseline are constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit is Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as scheduled visit day + (duration between scheduled visits)/2. (i.e .don't apply the minus 1 day).
- Subjects may delay durvalumab dosing under certain circumstances.
    - Dosing may be delayed per the Dosing Modification and Toxicity Management Guidelines, due to either an immune or a non-immune-related AE.
    - If dosing must be delayed for reasons other than treatment-related toxicity, dosing will resume as soon as feasible.
    - Dosing intervals of subsequent cycles may be shortened as clinically feasible in order to gradually align treatment cycles with the schedule of tumour efficacy (RECIST) and PRO assessments.
- Visit windowing is done separately for each assessment based on the schedule of events specific to that assessment.
- Should Study Day be missing (due to partial dates), then visit is assigned to the nominal visit at which that assessment was recorded, and no windowing will be performed for that specific assessment.
- Visit windowing is conducted up to and including the "Last dose of study treatment" visit. That is, the "Last dose of study treatment" visit is reassigned to a scheduled visit based on the study day that visit occurred at.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment are used (regardless of where it falls in an interval).
- Listings display all values contributing to a time point for a subject.
- For visit based summaries, if there is more than one value per subject within a time window then the closest value to the scheduled visit date are summarised, or the earlier, in the event the values are equidistant from the nominal visit date. The listings highlight the value for the subject 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 or not the value is close to the scheduled visit date.
- For summaries at a subject level, all values are included, regardless of whether they appear in a corresponding visit based summary, when deriving a subject level statistic such as maximum.

3.2.6 Handling of Unscheduled Visits

Unscheduled visits are included in the method of assigning data to scheduled visits described in the rules in Section 3.2.5above. Unscheduled visits are not included as a separate visit in the summary tables.

For summaries at patient level, such as of extreme values, all post-baseline values collected are used to derive a patient level statistic including those collected at unscheduled visits and regardless of whether they appear in the corresponding visit-based summary.

3.2.7 Multiplicity/Multiple Comparisons

No adjustments for multiplicity are planned.

3.2.8 Missing Dates

Generally, the imputation of dates is used to decide if an observation is treatment emergent for adverse events (AEs) or concomitant medications. Flags are retained in the database indicating where any programmatic imputation has been applied, and in such cases, any imputed dates should not be used to calculate durations, where the results would be less accurate.

The following are the guidelines used when partial dates are detected in the study:

- For missing diagnostic dates (e.g. disease diagnosis), if day and/or month are missing use 01 and/or Jan. If year is missing, put the complete date to missing.
- For missing AE and concomitant medication start dates, the following is applied:
  a. Missing day–- impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date.
  b. Missing day and month–- impute 1st January unless year is the same as first dose date then impute first dose date.
  c. Completely missing–- impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date.
  d. Imputed start date should be no later than the end date.
- For missing AE and concomitant medication end dates, the following is applied:
  a. Missing day–- impute the last day of the month unless month is the same as month of study discontinuation, then impute as study discontinuation date.
  b. Missing day and month–- impute 31st December unless year is the same as year of study discontinuation then impute study discontinuation date.

c. Completely Missing – If an AE/medication has a completely missing end date then it is treated as ongoing, unless this is for a prior anti-cancer medication then impute the date of informed consent.

- If a subject are known to have died where only a partial death date is available then the date of death is 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:
    a. For Missing day only – using the 1st of the month.
    b. For Missing day and Month – using the 1st of January.

For patients with a treatment start date for ceralasertib, the following rules are applied for missing treatment end dates for ceralasertib:

- If start date of durvalumab is present, then impute the missing ceralasertib end date as start date of durvalumab.
- Otherwise if start date of durvalumab is missing, then impute the missing ceralasertib end date as min(start date of ceralasertib + 6 days, end of study date, DCO, date of investigator decision to stop ceralasertib).

For partial subsequent anti-cancer therapy dates, the following rules will be applied for missing start dates:

- Missing day: If the month is the same as treatment end date then impute to the day after treatment, otherwise first day of the month.
- Missing day and month: If year is the same as treatment end date then impute to the day after treatment, otherwise 1st January of the same year as anti-cancer therapy date.

For time to event endpoints, dropouts and missing data are handled according to the censoring rules detailed within the relevant sections for the endpoint.

Other rules for handling missing data are described under the derivation rules for that particular variable.

3.2.9 Global/Country Situation

Impact of global or country situation is captured in the eCRF. Impact in terms of

- Disposition (discontinued IMP or withdrew study due to global/country situation)
- Disruption (visit impact, drug impacted)

- Important Protocol Deviations (IPDs) (Subjects with at least one IPD related to global/country situation)

will be summarized descriptively based on the FAS.

3.2.10 Derivations of RECIST 1.1 Visit Responses

For all subjects, the RECIST tumour response data is used to determine each subject's visit response according to RECIST version 1.1. It is also used to determine if and when a subject has progressed in accordance with RECIST and their best objective response to study treatment.

Baseline radiological tumour assessments should be performed no more than 21 days before the start of study treatment and ideally as close as possible to the start of study treatment. Tumour assessment is conducted every 8 weeks (±1 week) after the start of treatment (Cycle 1 Day 1) up to 18 months, then every 12 weeks (±1 week) until objective disease progression as per RECIST 1.1, irrespective of treatment decisions. Treatment start date will only be used to determine tumour assessment schedule, date of randomisation will be used for analysis.

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

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

Please refer to Table 2 for the definitions of CR, PR, SD and PD.

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

If there are patients with non-measurable disease or no evidence of disease assessed at baseline RECIST has been modified to allow the assessment of progression due to new lesions in subjects with no evidence of disease at baseline.

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 subject 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 target lesions (TLs). If more than one baseline scan is recorded then measurements from the one that is closest and prior to randomisation/first dose is used to define the baseline sum of TLs. It is 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, is selected.

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

Note: For subjects who do not have measurable disease at entry (i.e. no TLs) but have non-measurable disease, evaluation of overall visit responses are based on the overall NTL assessment and the absence/presence of new lesions (see section below for further details). If a subject does not have measurable disease at baseline then the TL visit response is Not Applicable (NA).

For subjects with no disease at baseline (i.e. no TLs and no NTLs), evaluation of overall visit responses is based on absence/presence of new lesions. If no TLs and no NTLs are recorded at a visit, both the TL and NTL visit response is recorded as NA and the overall visit response is no evidence of disease (NED). If a new lesion is observed then the overall visit response is PD.

Table 2 TL 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 <10mm.
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 ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.

Visit Responses	Description
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.
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). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
Not Applicable (NA)	No TLs are recorded at baseline.

Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum are rounded to one d.p. before assigning a TL response. For example 19.95% is rounded to 20.0% but 19.94% is rounded to 19.9%

Missing TL data

For a visit to be evaluable then all TL measurements are recorded. However, a visit response of PD is still assigned if any of the following occurred

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

Note: the nadir can only be taken from assessments where all the TLs had a LD recorded.

If the TL visit response is not recorded as PD, then the TL visit response is NE.

If there is at least one TL measurement missing and a visit response of PD cannot be assigned, the visit response is NE.

If all TL measurements are missing then the TL visit response is NE. Overall visit response is also NE, unless there is a progression of non-TLs or new lesions, in which case the response is PD.

Lymph nodes

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

- If all lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response is set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node short axis increases by 20% but remains < 10mm.

- If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response is set to NE irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node short axis increases by 20% but remains < 10mm.

- If not all lesions are missing, and those that are non-missing do not meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis ≥ 10mm or the reappearance of previously disappeared lesion), then response is set to PD.

- If all lesions are missing the response is set to NE.

TL too big to measure

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

TL too small to measure

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

Irradiated lesions/lesion intervention

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

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolization but note this does not include protocol specified biopsies), are handled in the following way. Once a lesion has had intervention then it is treated as having intervention for the remainder of the study noting that an intervention most likely shrinks the size of tumours:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) are summed and the calculation is performed in the usual manner. If the visit response is PD, this remains as a valid response category.

- Step 2: If there is no evidence of progression after step 1, the lesion diameter (for those lesions with intervention) will be treated 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 patients will be assigned a TL response of PD.

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

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

Lesions that split in two

If a TL splits in two, then the LDs of the split lesions are 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 are recorded for one of the TL sizes and the other TL size is recorded as 0mm.

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 is 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 are treated as missing.

Non-target lesions (NTLs) and new lesions – site investigator data.

At each visit, the investigator is to 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 is derived based on the investigator's overall assessment of NTLs as follows:

Table 3 NTL Visit Responses

Visit Responses	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.
Not evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For subjects 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.

To achieve 'unequivocal progression' on the basis of 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 are also recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and indicates 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 are identified via a Yes/No tick box. The absence and presence of new lesions at each visit are listed alongside the TL and NTL visit responses.

A new lesion indicates progression so the overall visit response is 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 study therapy and is not included in any assessment of NTLs.

Subjects 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.

Overall visit response – site investigator data

Table 4 defines how the previously defined TL and NTL visit responses are combined with new lesion information to give an overall visit response. Confirmation of progression is not required.

Table 4 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**

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
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)	**Non-CR/Non-PD**
NA	NE	No (or NE)	**NE**
NA	NA	No (or NE)	**NED**

4 STATISTICAL ANALYSIS

This section provides information on definition, derivation and analysis/data presentation per domain.

4.1 Analysis Sets

The analysis sets for this study are described in Table 5.

Table 5 Analysis Sets

Analysis set	Description	Endpoint/Output
Enrolled analysis set	All subjects who sign the informed consent form (ICF).	Disposition
Safety analysis set	All subjects receiving at least 1 dose of study treatment. Subjects are summarised according to the actual treatment received.	Exposure Safety ORR BOR TTR DoR EORTC IL PGI-TT PK concentration listings Percentage Change in Target Lesion Tumour size Best Percentage Change in Target Lesion Tumour size
BICR Modified safety analysis set	All subjects receiving at least 1 dose of study treatment who have measurable disease at baseline	ORR (interim) BOR (interim) DCR (interim)

Analysis set	Description	Endpoint/Output
	by BICR and who have first dose at least 9 weeks prior to data cut off.	
Investigator Modified safety analysis set	All subjects receiving at least 1 dose of study treatment who have measurable disease at baseline by Investigator and who have first dose at least 9 weeks prior to data cut off.	ORR (interim) BOR (interim) DCR (interim)
Full analysis set	All subjects who are randomised in the study. The FAS is used for all the efficacy analyses including PROs. Treatment arms are summarized on the basis of randomised study treatment, regardless of the treatment actually received. Subjects who were randomised but did not subsequently receive study treatment are included in the analysis in the treatment arm to which they were randomised.	Baseline and demography ORR (for comparative analysis) PFS OS PROs
Pharmacokinetics (PK) analysis set	All dosed subjects with reportable ceralasertib or durvalumab plasma concentrations. Subjects are summarised according to the treatment received.*	PK concentrations
Pharmacodynamic analysis set	All subjects who receive at least 1 dose of study treatment with at least 1 reportable post-baseline pharmacodynamic measurement. Subjects are summarised according to the treatment received.*	Pharmacodynamic endpoints
BICR Response evaluable set	All dosed subjects who have measurable disease at baseline by BICR	ORR (sensitivity analysis)
Investigator Response evaluable set	All dosed subjects who have measurable disease at baseline by Investigator	ORR (sensitivity analysis)
BICR Interim response evaluable	All dosed subjects who have measurable disease at baseline by BICR assessment and who have first dose at least 17 weeks prior to data cut off.	ORR (interim) BOR (interim) DCR (interim)
Investigator Interim response evaluable	All dosed subjects who have measurable disease at baseline by Investogator and who have first dose at least 17 weeks prior to data cut off.	ORR (interim) BOR (interim) DCR (interim)

ORR: Objective Response Rate, DCR: Disease Control Rate, TTR: Time to Response, DoR: Duration of Response, PFS: Progression Free Survival, OS: Overall Survival, BOR: Best Objective Response; BIRC: Blinded Independent Central Review

TTR and DoR are reported for the subset of subjects with confirmed objective response.

*Individual PK concentration data for any subjects who are excluded from the descriptive summary tables and/or figures are included in the listings and are flagged with an appropriate footnote.

Subject's data are flagged as being unevaluable in case no signed ICF is available. Subjects who are not randomised due to violation of one or more protocol defined inclusion or exclusion criteria are handled as screening failures.

4.1.1.1 Presentation

The analysis sets are summarised by treatment arm and for all subjects combined. (non-efficacy summaries). Any exclusions from analysis sets are listed.

4.2 Study Population

The domain study population covers subject disposition, analysis sets, protocol deviations, demographics, baseline characteristics, medical history, disease characteristics, prior and concomitant medication, and study drug compliance.

4.2.1 Subject Disposition and Completion Status

4.2.1.1 Definitions and Derivations

Subjects screened is defined as informed consent received.

A subject is considered to have completed the study if they have completed all phases of the study including the last visit or the last scheduled procedure. Details can be found in section 4.4 of the CSP.

4.2.1.2 Presentation

Subject disposition including screen failures and reason for screen failure is summarised and listed based on all subjects screened by treatment arm and for all subjects combined as defined by the current relevant tables, figures, listings (TFL) standards. The number and percentage of subjects for the following are summarised if applicable:

- Subjects screened;
- Screen failures;
- Subjects randomised;
- Subjects randomised, did not receive study treatment
- Subjects started treatment
- Subjects who received ceralasertib and subjects who did not receive ceralasertib;
- Subjects who received durvalumab and subjects who did not receive durvalumab (for combination arm);
- Subjects ongoing study treatment at data cut-off;
- Subjects who discontinued treatment (including reason),
- Subjects ongoing study at data cut-off;
- Subjects who terminated study.

Summaries on disposition due to global/country situation due to a pandemic are added to the disposition table if applicable. The number and percentage of subjects for the following summaries are added on if applicable:

- Subjects who discontinued treatment due to global/country situation;
- Subjects who withdrew from study due to global/country situation.

The study disruptions due to the global/country situation is also summarised as a separate table.

The number and percentage of subjects with confirmed or suspected Coronavirus Disease 2019 (COVID-19) infection during the course of the study are presented separately, including details on COVID-19 related interruptions impacting on visits, and study drug administration. Discontinuation of study intervention and/or withdrawal from study due to COVID-19 are presented separately.

Listings of subjects affected by the COVID-19 pandemic are presented detailing any affect and impact on the study. Issues reported in the Clinical Trial Management System are considered for presentation in listings as well.

4.2.2 Protocol Deviations

4.2.2.1 Definitions and Derivations

Important protocol deviations (IPDs) are defined as protocol deviations that may significantly affect the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being. IPDs are defined in detail in the Protocol Deviation Specification Document., but will generally include, but are not limited to:

- Key Inclusion criteria
- Key Exclusion criteria
- Discontinuation criteria for study product met but subject not withdrawn from study treatment
- Investigational Product (IP) Deviation
    - Subjects deviating from the prescribed dosing regimen.
    - Dose modifications (not related due to either an immune or a non-immune-related AE according to Dosing Modification and Toxicity Management Guidelines)
- Excluded medication taken
    - Received prohibited concomitant medications or therapies (including other anticancer agents).
- Deviations related to study procedure
    - Baseline Response Evaluation Criteria in Solid Tumours (RECIST) version 1.1 scan > 28 days before randomisation..
    - No baseline RECIST 1.1 assessment on or before date of randomisation.

- Other important protocol deviations
    - Deviation from Good Clinical Practice (GCP) as determined by medical review

If a deviation is serious enough to have a potential impact on the primary analysis, sensitivity analyses may be performed. A list of IPDs according to the Protocol Deviation Specification Document is compiled and new potential IPDs are classified by the global lead physician and an unblinded biostatistician trailed by AZ study physician and AZ clinical operations review approximately every 4-5 weeks. The Protocol Deviation Specification Document and the final IPD list will be finalised prior to database lock.
Presentation

The incidence of important protocol deviations (IPDs) are summarised for the full analysis set.

- Number of subjects with at least 1 important protocol deviation further split by reason related to;
- Number of subjects with at least 1 pandemic related important protocol deviation further split by reason related to;
- Number of subjects with at least 1 important protocol deviation, excluding pandemic related IPDs further split by reason related to.

A listing is provided with the important protocol deviation details.

4.2.3 Demographics

4.2.3.1 Definitions and Derivations

Age group is defined as a categorical variable with levels <18, ≥ 18,<50, ≥ 50, < 65, ≥ 65, < 75 and ≥ 75 years. Each race category counts subjects who selected only that category.

4.2.3.2 Presentation

Demographics are summarised and listed based on the FAS by treatment arm and for all subjects combined as defined by the current relevant TFL standards. The following are summarised: age (years) at baseline, age group, sex, race and ethnicity. Additionally, the number and percentage of subjects recruited in each country and each centre are presented.

4.2.4 Baseline Characteristics

4.2.4.1 Definitions and Derivations

Body Mass Index is calculated as weight/height^2 (kg/m^2).

4.2.4.2 Presentation

Baseline characteristics are listed and summarised according to Section 3.2 for the FAS by treatment arm and for all subjects combined as defined by the current relevant TFL

standards. The following are summarised: weight (kg), height (cm), body mass index (kg/m^2).

4.2.5 Disease Characteristics

4.2.5.1 Definitions and Derivations

Not applicable.

4.2.5.2 Presentation

Disease characteristics at baseline are listed and summarised in terms of absolute counts and percentages affected of subjects for the FASby treatment arm and for all subjects combined as defined by the current relevant TFL standards.

The following are summarised:

- Extend of disease at study entry (metastatic, locally advanced disease, both metastatic and locally advanced, location)
- Disease characteristics at baseline (Eastern Cooperative Oncology Group [ECOG] performance status, primary tumour location, primary tumour type, number of metastases, histology type, primary tumour stage, regional lymph nodes stage, distant metastases stage, tumour grade, Stage/AJCC stage, resistance to prior IO treatment [IVRS], and baseline LDH level [IVRS])
- Mutation status at study entry (Proto-oncogene c-KIT [KIT], neurofibromatosis type 1 [NF1], ataxia telangiectasia and Rad3-related protein [ATR], ataxia telangiectasia mutated [ATM], AT-rich interactive domain-containing protein 1A [ARID1A], B-Rapidly Accelerated Fibrosarcoma gene [BRAF] V600 Mutation and subtype, NRAS proto-oncogene, GTPase [NRAS] Activating Mutations and microsatellite instability [MSI]-High Genetic or Immunohistochemistry)

Recurrence of earlier cancer is depicted in a listing.

4.2.6 Medical History and Concomitant Disease

4.2.6.1 Definitions and Derivations

Medical and surgical history is coded using Medical Dictionary for Regulatory Activities (MedDRA).

4.2.6.2 Presentation

Medical history and concomitant disease are listed and summarised for the FAS by treatment arm and for all subjects combined as defined by the current relevant TFL standards.

A separate summary of medical history for subjects who had confirmed or suspected COVID-19 infection during the study is presented. If less than five events appear and less than 2% of randomized subjects are affected the summary will not be visualised.

Additionally, the following is summarised in terms of numbers and percentage of subjects who had:

- Previous disease related treatment modalities (cytotoxic chemotherapy, targeted therapy, antiangiogenic therapy, taxane chemotherapy, radiopharmaceuticals, platinum chemotherapy, experimental therapy, transarterial chemoembolization [TACE], transarterial radioembolization [TARE], immunotherapy, other, number of previous chemo regimens, number of prior lines of systemic therapy, reason for therapy failure of previous cancer therapy, and treatment status of previous cancer therapy)

- Previous Anti-PDL1 therapies (Therapy class, treatment status, Reason for Therapy discontinuation, type of resistance of prior Anti-PDL1, name of prior Anti-PDL1 agent)

Summaries for the number and percentage of subjects wo had a certain number of prior regimens and response assessment results are produced.

The number and percentage of subjects who had prior cancer therapies are summarised by ATC classification and generic drug name.

Results from the Hepatitis B and C and HIV tests at screening are depicted in a listing.

4.2.7 Prior and Concomitant Medications

4.2.7.1 Definitions and Derivations

For the purpose of inclusion in prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates are imputed as detailed in Section 3.2.8.

Prior medications, concomitant 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 the first dose of study treatment.
- Concomitant medications are those with a stop date on or after the first dose date of study treatment, and must have started prior to or during treatment so there is at least one day in common with the study treatment.

4.2.7.2 Presentation

The following summaries are provided in terms of absolute counts and percentages affected of subjects for the FAS o by treatment arm and for all subjects combined as defined by the current relevant TFL standards using ATC classification code and the generic term coded by standard drug dictionary WHODrug Global B3 Mar 2022 or later:

- summary of prior medications or therapies
- summary of concomitant medications or therapies

Additionally, the following is summarised in terms of numbers and percentage of subjects who had:

- Previous Anti-PDL1 therapies (Therapy class, treatment status, Reason for Therapy discontinuation, type of resistance of prior Anti-PDL1, name of prior Anti-PDL1 agent)

Also, all important recorded information for previous Anti-PDL1 therapies will be visualised in a listing.

4.3 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints including sensitivity and supportive analyses.

Efficacy analyses, except for OS, are based on BICR assessments. All images are collected centrally whether scheduled or unscheduled. The imaging scans are reviewed by 2 independent radiologists using RECIST 1.1 and are adjudicated, if required. For each post-baseline visit scan, the BICR defines the overall visit response data (complete response [CR], partial response [PR], stable disease [SD], progressive disease [PD], or not evaluable [NE] depending on the status of their disease compared with baseline and previous scans by assessing TLs, non-target lesions (NTLs) and new lesions) and the relevant scan dates for each time point (i.e., for visits where response or progression is/is not identified). Subjects who have disease progression on study treatment based on progression of non-target disease, may also require submission of additional scans/images/photographs e.g., brain scan or photographs of cutaneous lesions. If a subject has had a tumour assessment that cannot be evaluated, then the subject is assigned a visit response of not evaluable (NE) (unless there is evidence of progression, in which case the response is assigned as PD). No programmatic derivation of visit response is necessary. The date of progression is provided (by each reviewer) based on the earliest of the scan dates of the component that triggered the progression. Endpoints ORR, DoR, BOR, TTR, DCR, and PFS are then derived from the scan dates and overall visit responses. Confirmation of progression is not required.

Details on the adjudication process can be found in the CSP section 8.1.1.2.

As per the sensitivity analyses, efficacy analysis, except for OS, are also based on programmatic application of RECIST 1.1. (Eisenhauer et al, 2009) to investigator assessed tumour measurements. All RECIST 1.1 assessments, whether scheduled or unscheduled, are included in the calculations. This is also regardless of whether a subject discontinues study treatment or receives another anti cancer therapy. At each visit, subjects are

programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, PD, or NE depending on the status of their disease compared with baseline and previous assessments.

Baseline should be assessed within the 21 days prior to randomisation. The tumour response endpoints (ORR, DoR, BOR, TTR, PFS, and change in TL tumour size) are then derived from the scan dates and overall visit responses. Programmatic derivation guidance used for the application of RECIST 1.1 are provided in Section 3.2.10 which is to determine disease response.

Another sensitivity analysis for ORR based on the dosed subjects who have measurable disease at baseline by BICR (response evaluable set) may be conducted.

RECIST data, overall visit response and best objective response are listed.

All efficacy analyses are presented by treatment group.

Subgroup Analyses

The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic and/or predictive factors.

The subgroup analyses described below are applicable for the primary endpoint (ORR) and for selected secondary endpoints (DoR, PFS and OS)

Subgroups analyses are conducted in the following subgroups:

- Resistance to prior IO treatment (primary / early relapse on adjuvant versus acquired)
- Baseline LDH (<= ULN, >ULN but <2x ULN, >= 2x ULN)
- Sex (male versus female)
- Age at randomisation (<65 years versus >= 65 years of age)
- Melanoma subtype (cutaneous versus acral versus mucosal)
- Presence of liver metastases (yes versus no)
- Presence of brain metastases (yes versus no)
- BRAF Mutation Status (Mutation detected versus Mutation not detected/Not done/unknown)
- Targeted Therapy as Previous Cancer Therapy Class (yes versus no)

Other baseline variables may also be assessed if there is clinical justification or an imbalance is observed between the treatment arms.

The subgroup analyses for the stratification factors are based on the values entered into the IWRS; all other factors are based on the values recorded on the eCRF, or from the third-party vendor data.

If there are too few responders available for a meaningful analysis of a particular subgroup comparison (it is not considered appropriate to present analyses where there are less than 10 responders across both treatment groups in a subgroup comparison), the subgroup analysis will not be performed. In this case, only descriptive summaries will be provided. The same process will be applied to the secondary endpoints PFS and OS if there are less than 20 events across both treatment groups in a subgroup comparison.

4.3.1 Primary Endpoint: Objective Response Rate

The primary efficacy endpoint is ORR by BICR.

4.3.1.1 Definition

ORR is defined as the proportion of subjects who have a complete response (CR), or partial response (PR), as determined by BICR per RECIST 1.1, prior to any evidence of progression (as defined by RECIST 1.1), that is confirmed at least 4 weeks later. ORR based on BICR is defined as the percentage of subjects with at least two post-baseline BICR-assessed visit response of CR or PR, with the denominator defined as the number of subjects in the Safety analysis set. For the sensitivity analysis, the subset of subjects in the Response evaluable set is used. For the comparative analysis of ORR between treatment arms the FAS is used instead (see Section 4.3.1.6).

The ORR analysis by BICR will be repeated for investigator data.

4.3.1.2 Derivations

A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit. Data obtained up until progression, or last evaluable assessment in the absence of progression, are included in the assessment of ORR. Also only data obtained before the start of subsequent anticancer treatment (excluding radiotherapy) are included. Therefore both visits contributing to a confirmed response must be prior to progression and prior to subsequent anticancer treatment.

In the case where a subject has two non-consecutive visit responses of PR, then, as long as the time between the 2 visits of PR is greater than 4 weeks and there is no PD between the PR visits, the subject is defined as a responder. Similarly, if a subject has visit responses of CR, NE, CR, then, as long as the time between the 2 visits of CR is greater than 4 weeks, then a best response of CR is assigned.

4.3.1.3 Handling of Dropouts and Missing Data

If any of the randomised subjects at the time of the respective analysis (primary, CCI) have no measurable disease at baseline as per BICR, then an additional analysis of ORR will be conducted for the Response evaluable set.

4.3.1.4 Primary Analysis of Primary Endpoint

Summaries are produced that present the number and percentage of subjects with a confirmed tumour response (CR/PR). The ORR is presented for each treatment arm non-comparatively with a two-sided 95% CI using the Clopper-Pearson (exact probability) method. Subjects that have missing overall response assessments at all visits are considered as non-responders, and are therefore counted in the denominator of ORR. The main analysis of ORR is based on the Safety analysis set.

4.3.1.5 Sensitivity Analyses of the Primary Endpoint

As a sensitivity analysis, ORR is analysed according to Section 4.3.1.4 but using the investigator assessment per RECIST 1.1.

4.3.1.6 Secondary measure of interest of the Primary Endpoint: odds ratio of the ORR

ORR is compared between the treatment arms using logistic regression adjusting for PD-(L)1 resistance type and LDH per randomisation stratification as covariates. The results of the analysis are presented in terms of an odds ratio for study treatments together with associated 95% profile likelihood CI and p-value (based on twice the change in loglikelihood resulting from the addition of a treatment factor to the model). This comparative analysis is provided using the full analysis set.

4.3.1.7 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3 .

4.3.2 Best Objective Response

4.3.2.1 Definition

Best objective response (BoR) will be calculated based on the overall visit responses from each RECIST assessment, described in section 3.2.10. 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.

4.3.2.2 Derivations

BoR is based on RECIST and will be determined using BICR data and investigator data until the earliest of the first progression event (including death)/last evaluable assessment in

the absence of RECIST progression or start of any subsequent cancer therapy. The denominators are consistent with those used in the ORR analysis.

BoR will be based on RECIST using the following response categories and in the following order:

- Response with categories: CR (confirmed), PR (confirmed)
- Non-response with categories:
    - SD with categories unconfirmed CR or PR, SD (>=16 weeks) and Non-CR/Non-PD (>=16 weeks)
    - NE with categories SD (<16 weeks), Non-CR/Non-PD (<16 weeks) and no baseline RECIST or post baseline RECIST assessments in the absence of death.
    - Progression with categories: RECIST progressions and death in the absence of RECIST progression

For determination of a best response of SD and Non-CR/Non-PD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD/ Non-CR/Non-PD should be recorded at least 16 weeks minus 1 week, i.e. at least 15 weeks (to allow for an early assessment within the assessment window), after randomisation. For CR/PR, the initial overall visit assessment that showed a response should use the latest of the dates contributing towards a particular overall visit assessment.

4.3.2.3 Analysis of Best Objective Response

The number of patients with a response and non-response along with each of the response/non-response categories will be summarised.

4.3.2.4 Sensitivity Analyses

As a sensitivity analysis, BOR is analysed according to 4.3.5.4 but using BICR and investigator data.

4.3.2.5 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3.

4.3.3 Disease Control Rate

4.3.3.1 Definition

Disease control rate (DCR) at 8 weeks is defined as the percentage of patients who have at least one visit response of CR or PR prior to PD without subsequent cancer therapy or have demonstrated SD or NTL response of Non-CR/Non-PD in the absence of any new lesions for patients with non-measurable disease for at least 7 weeks (i.e. 8 weeks - 1 week to allow

for an early assessment within the assessment window) after randomisation and prior to PD without subsequent cancer therapy.

4.3.3.2 Derivations

Duration of SD (weeks) is defined as: (date last evaluable assessment of SD in the absence of progression prior to subsequent cancer therapy - randomisation date +1)/7

Patients without a post-baseline tumour assessment will be considered to have no Disease Control.

4.3.4 Analysis of Disease Control Rate

DCR will be analysed at Interim Analysis 2 only (see Section 5). The DCR is presented for each treatment arm non-comparatively with a two-sided 80% CI using the Clopper-Pearson (exact probability) method. Subjects that have missing overall response assessments at all visits are considered as non-responders and are therefore counted in the denominator of DCR. The analysis of DCR is based on the Interim Response Evaluable Set.

4.3.5 Secondary Endpoint: Duration of Response

DoR is a secondary efficacy endpoint.

4.3.5.1 Definition

DoR is defined as the time from the date of first documented confirmed objective response until the date of documented progression per RECIST 1.1 as assessed by BICR or death due to any cause.

4.3.5.2 Derivations

DoR (months) = (date of PFS event [progression/death] or censoring – date of first documented confirmed objective response + 1) / (365.25/12)

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 is defined as the latest of the dates contributing towards the first visit response of PR or CR. If a subject does not progress following a response, then their DoR is censored on the PFS censoring date. Only subjects who have achieved objective response (CR or PR) are evaluated for DoR.

4.3.5.3 Handling of Dropouts and Missing Data

DoR is calculated for the subset of subjects with confirmed objective response (CR or PR) only. Subjects without confirmed objective response are excluded from the analysis of DoR. Dropouts and missing data will be handled according to the censoring rules detailed in Section 4.3.9.2,

4.3.5.4 Primary Analysis of Secondary Endpoint

Only subjects who have achieved confirmed objective response (CR or PR) are included in the summaries of DoR. The median DoR as well as landmark estimates at 6, 9, 12, 15, 18 months of DoR and two-sided 95% CI are estimated using the Kaplan-Meier method. The analysis of DoR is based on the safety analysis set.

Swimmer plots that clearly show the profile of each subject who responds are produced.

4.3.5.5 Sensitivity Analyses of the Secondary Endpoint

As a sensitivity analysis, DoR is analysed according to 4.3.5.4 but using the investigator assessment per RECIST 1.1.

4.3.5.6 Supplementary Analyses of the Secondary Endpoint

Another sensitivity analysis for DoR is performed censoring DoR at the date of the last progression-free disease assessment prior to initiation of subsequent anticancer treatment. The rationale for this sensitivity analysis is the same as outlined in 4.3.9.5.

4.3.5.7 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3.

4.3.6 Secondary Endpoint: Time to Response

Time to Response (TTR) is a secondary efficacy endpoint. The analysis of TTR is based on the safety analysis set.

4.3.6.1 Definition

TTR is defined as the time from randomisation until the date of first documented objective response, which is subsequently confirmed per RECIST 1.1 as assessed by BICR.

4.3.6.2 Derivations

TTR (months) = (date of first confirmed objective response – date of randomisation + 1) / *(*365.25/12*)*.

4.3.6.3 Handling of Dropouts and Missing Data

Subjects without confirmed objective response, are treated as missing for the calculation of TTR. Only subjects who have achieved confirmed objective response (CR or PR) are evaluated for TTR.

4.3.6.4 Primary Analysis of Secondary Endpoint

The TTR is summarised (i.e., number of subjects [%] based on the number of responders for each treatment arm) by the scheduled assessment timepoint that the response was first observed. The number of subjects with response at different disease assessment timepoints is provided. Additionally, descriptive summary statistics are presented in addition.

The median TTR and two-sided 95% CI is assessed using the Kaplan-Meier method (without censoring because all subjects have events).

4.3.6.5 Sensitivity Analyses of the Secondary Endpoint

As a sensitivity analysis, TTR is analysed according to 4.3.6.4 but using the investigator assessment per RECIST 1.1.

4.3.6.6 Supplementary Analyses of the Secondary Endpoint

Not applicable

4.3.6.7 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3.

4.3.7 Secondary Endpoint: Percentage Change in target lesion tumour size

Percentage Change in target lesion tumour size is a secondary efficacy endpoint.

4.3.7.1 Definition

An outcome endpoint for this study is percentage change from baseline in TL tumour size. This is based on RECIST TL measurements taken at baseline and each post-baseline assessment. Tumour size is the sum of the longest diameters (or short axis measurements for lymph nodes) of the TLs. Baseline for RECIST is defined to be the last evaluable assessment prior to randomisation.

The TL tumour size and percentage change from baseline in the sum of TL tumour size at each assessment are calculated.

4.3.7.2 Derivations

Whenever TL tumour size data for the week 16 visit (note: or visit at which progression was documented if before week 16) is available then this is used in the analysis. A windowing rule is applied and follows the protocol allowed visit window, therefore RECIST scan performed within ±1 week of the protocol scheduled visit is used for that visit. Subjects who progress before week 16 (plus 1 week allowing for a late assessment within the visit window) should have had a tumour assessment performed at the time of progression prior to study treatment discontinuation. The TL tumour size from their latest progression assessment is used instead of the week 16 assessment for these subjects

The percentage change from baseline in TL tumour size is obtained for each subject by taking the difference between the sum of the TLs at week 16 (or earlier if the patient progressed) and the sum of the target lesions at baseline divided by the sum of the TLs at baseline times 100 (i.e. [week 16 – baseline] / baseline × 100).

4.3.7.3 Primary Analysis of Secondary Endpoint

The TL tumour size and percentage change in TL tumour size from baseline is summarised using descriptive statistics and presented at each timepoint for the Safety set. Descriptive statistics for the percentage change from baseline at week 16 are presented, together with the number of subjects with a baseline and at least 1 post-baseline assessment.

The percentage change from baseline in TL tumour size is presented graphically using waterfall plots, with the bars ordered from the largest increase to the largest decrease. Separate waterfall plots are used to present each subject's week 16 percentage change in TL tumour size as a separate bar. A reference line at the –30% change in TL tumour size level is added to the plots, which corresponds with the definition of 'partial response'. All progressions are marked with a '●' or designated with patterns or colours for ORR categories. Flagged progressions on the percentage change in TL tumour size at week 16 are based upon NTL or new lesion progression at that timepoint and flagged progressions on the best percentage change are based upon NTL or new lesion progression at the same timepoint as the best percentage change. The scale in these plots is 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 are marked with '#'. Values are ordered in descending order with the imputations due to death appearing first followed by a gap followed by all other subjects. These plots will be produced for the Safety Set.

4.3.7.4 Sensitivity Analyses of the Secondary Endpoint

As a sensitivity analysis, percentage change in TL tumour size is analysed in addition as described in 4.3.7.3 but using the investigator assessment per RECIST 1.1.

4.3.7.5 Supplementary Analyses of the Secondary Endpoint

Not applicable.

4.3.7.6 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3.

4.3.8 Secondary Endpoint: Best Percentage Change in target lesion tumour size

Best Percentage Change in target lesion tumour size is a secondary efficacy endpoint.

4.3.8.1 Definition

The TL tumour size and percentage change from baseline in the sum of TL tumour size at each assessment are calculated. The best absolute change in TL tumour size from baseline (i.e. depth of response) is the largest decrease from baseline or the smallest increase from baseline in the absence of reduction and includes all assessments:

- up to and including the first visit at which the overall visit response is PD,
- prior to death in the absence of progression,
- prior to the start of subsequent anti-cancer therapy (excluding radiotherapy);
- or up to and including the last evaluable RECIST assessment if the subject has not died, progressed or started subsequent anti-cancer therapy.

Each post-baseline disease assessment for a subject that meets the following conditions are included: all target lesions identified at baseline have measurements recorded at the current visit (i.e. they cannot be not done or not evaluable). If a lesion is recorded as ≤ 5 mm then this will be used in the calculations.

4.3.8.2 Derivations

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 the best percentage change from baseline in the following situations (otherwise best percentage change is 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

4.3.8.3 Handling of Dropouts and Missing Data

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

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

4.3.8.4 Primary Analysis of Secondary Endpoint

Descriptive statistics for the best percentage change from baseline are presented, together with the number of subjects with a baseline and at least 1 post-baseline assessment.

The best percentage change from baseline in TL tumour size is presented graphically using waterfall plots, with the bars ordered from the largest increase to the largest decrease. Separate waterfall plots are used to present each subject's best percentage change in TL tumour size as a separate bar. A reference line at the –30% change in TL tumour size level is added to the plots, which corresponds with the definition of 'partial response'. All progressions are marked with a '●' or designated with patterns or colours for ORR

categories. Flagged progressions on the best percentage change in TL tumour size at week 16 are based upon NTL or new lesion progression at that timepoint and flagged progressions on the best percentage change are based upon NTL or new lesion progression at the same timepoint as the best percentage change. The scale in these plots is 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 are marked with '#'. Values are ordered in descending order with the imputations due to death appearing first followed by a gap followed by all other subjects. These plots will be produced for the Safety set.

4.3.8.5 Sensitivity Analyses of the Secondary Endpoint

As a sensitivity analysis, best percentage change in TL tumour size is analysed in addition as described in 4.3.7.3 but using the investigator assessment per RECIST 1.1.

4.3.8.6 Supplementary Analyses of the Secondary Endpoint

Not applicable.

4.3.8.7 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3.

4.3.9 Secondary Endpoint: Progression Free Survival

PFS is a secondary efficacy endpoint.

4.3.9.1 Definition

PFS is defined as the time from randomisation until the date of first documented disease progression as assessed by BICR or death (by any cause in the absence of disease progression), regardless of whether the subject withdraws from study therapy or receives another anti-cancer therapy prior to progression.

4.3.9.2 Derivations

PFS (months) = (date of PFS event (progression/death) or censoring – date of randomisation + 1) / (365.25/12)

Subjects who have not progressed or died at the time of analysis are censored at the time of the latest date of assessment from their last evaluable disease assessment. However, if the subject progresses or dies immediately after two or more consecutive missed visits, the subject is censored at the time of the latest evaluable disease assessment prior to the two missed visits. (Note: a not evaluable visit is not considered as a missed visit).

For the time period where the 8-weekly schedule applies, two or more consecutive missing visits resulting in ≥ 18 weeks after the last evaluable post-baseline disease assessment, 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 two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from 8-weekly to 12-weekly this equates to 22 weeks (i.e. the average of 8 and 12 weeks which gives 10 weeks and then apply the same rationale, hence 2 x 10 weeks + 1 week for early assessment + 1 week for late assessment = 22 weeks).

When the scheduling changes to 12-weekly assessments, two missing visits equate to 26 weeks (2 x 12 weeks + 1 week for early assessment + 1 week for late assessment = 26 weeks).

If subjects have no evaluable disease assessments post-baseline or do not have baseline tumour assessment data they will be censored at Day 1 unless they die within two visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window) when the death date qualifies as a PFS event.

A summary of censoring rules and the date of PD/death or censoring are given in Table 6. Note that censoring overrides event in certain specified cases.

Table 6 Summary of Censoring Rules for PFS

Situation	Date of PD/Death or Censoring	PFS Outcome
Documented PD or death in the absence of progression	Date of earliest documentation of PD or date of death in the absence of progression	Event
Either no tumour assessment at baseline or no evaluable assessments post-baseline AND death prior to second scheduled post-baseline disease assessment	Date of death	Event
Either no tumour assessment at baseline or no evaluable assessments post-baseline AND no death prior to second scheduled post-baseline disease assessment	Date of randomisation	Censored
PD or death (in the absence of progression) immediately after ≥ 2 consecutive missed disease assessments as per the protocol specified assessment schedule	Last evaluable progression-free disease assessment prior to missed assessments	Censored
On-going with neither PD nor death at the time of analysis or lost to follow-up or withdrawn consent	Date of last evaluable disease assessment	Censored
Initiation of subsequent anticancer treatment prior to PD or death.	Date of the last progression-free disease assessment prior to initiation of subsequent anticancer treatment	Censored for sensitivity analysis only
COVID-19 related deaths	Date of last evaluable RECIST 1.1 assessment prior to COVID-19 infection related to death	Censored for sensitivity analysis only

PD = progressive disease; PFS = progression-free survival

The PFS time is always derived based on scan/assessment dates not visit dates.

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

- For investigator assessments, the date of progression is determined based on the earliest RECIST assessment/scan dates of the component that indicates progression.
- For BICR assessments, the date of progression is determined based on the earliest dates of the component that triggered the progression on the first set of scans that indicates progression for the adjudicated reviewer selecting disease progression per RECIST 1.1, or of either reviewer where both reviewers select disease progression per RECIST 1.1 as a time point response, and there is no adjudication for central review (BICR) data.

- When censoring a subject for PFS the subject is censored at the latest of the dates contributing to a particular overall visit assessment.

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

The proportion of subjects alive and progression-free at 3, 6, 9, 12 and 18 months from randomisation is defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1) at these specific months.

Potential duration of follow-up for PFS, applicable only for PFS censored subjects, is defined as follows:

Potential duration of follow-up for PFS in censored subjects (months) = (date of PFS censoring – date of randomisation + 1) / (365.25/12).

4.3.9.3 Handling of Dropouts and Missing Data

See Section 4.3.9.2.

4.3.9.4 Primary Analysis of Secondary Endpoint

The analysis of PFS is based on the Full analysis set. The number and percentage of subjects experiencing a PFS event (broken down by type of event/censoring) are provided along with the median PFS and its two-sided 95% CI estimated using the Kaplan-Meier method. Kaplan-Meier plots of PFS are presented by treatment arm. Moreover, the landmark estimates of PFS at 3, 6, 9, and 12 months from randomisation are summarised by treatment arm.

The treatment status at progression of subjects at the time of analysis are summarised. This includes the number (%) of subjects who were on treatment at the time of progression, the number (%) of subjects who discontinued study treatment prior to progression, the number (%) of subjects who have not progressed and were on treatment or discontinued treatment.

PFS is analysed using a log-rank test stratified by the randomisation stratification factors (baseline LDH and IO resistance). The Hazard Ratio (HR) together with its 95% CI and p-value is presented. The HR and CI are estimated from a stratified Cox proportional hazard model with ties=Efron, and the CI is calculated using a profile likelihood approach.

A summary of 'potential' duration of follow-up for PFS is included using median (range). This is presented for censored subjects (including all types of PFS censoring).

4.3.9.5 Sensitivity Analysis of Secondary Endpoint

As a sensitivity analysis, PFS is analysed also using the investigator assessment per RECIST 1.1.

Another sensitivity analysis for PFS is performed censoring PFS at the date of the last progression-free disease assessment prior to initiation of subsequent anticancer treatment. The rationale for this sensitivity analysis is that it allows to indirectly estimate the impact of initiating a subsequent cancer treatment on the PFS. For example, a subject that is progression-free at the date of the last assessment prior to subsequent anticancer treatment, starts subsequent anticancer treatment and progresses on the following assessment is no longer contributing on the PFS for the whole time period until progression since that subject will be censored prior to that.

A further sensitivity is conducted to assess the potential of impact on COVID-19 related deaths on PFS. That is, the subjects 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, are censored at the last evaluable RECIST 1.1 assessment prior to COVID-19 infection related to death. If less than five events appear and less than 2% of enrolled subjects are affected the sensitivity analysis will not be carried out.

4.3.9.6 Supplementary Analyses of the Secondary Endpoint - Assessment bias

The following summaries of the number of disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group:

- Total number of disagreements
- Number of disagreements within 2 weeks
- Progression date >2 weeks earlier by central review than by investigator (does not include instances where progression by central review but not by investigator)

- Progression date >=2 weeks earlier by investigator than by central review (does not include instances where progression by investigator but not by central review)
- Investigator but not central review
- Central review but not investigator

In addition, the following will also be summarised:

- number or patients where there is no progression for both central review and investigator assessed,
- early discrepancy rate defined as the number of investigator declared progressions before central review as a proportion of all investigator progressions and corresponding difference in rates between treatment groups
- late discrepancy rate defined as the number of investigator declared progressions after central review as a proportion of all discrepancies and corresponding difference in rates between treatment groups

4.3.9.7 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3 . The results of subgroup analyses will be presented on a forest plot including the HR and 95% CI, along with the results of the overall analysis.

4.3.10 Secondary Endpoint: Overall Survival

OS is a secondary efficacy endpoint.

4.3.10.1 Definition

OS is defined as the time from the date of randomisation until death due to any cause regardless of whether the subject withdraws from study therapy or receives another anti-cancer therapy.

4.3.10.2 Derivations

OS (months) = (date of death or censoring – date of randomisation + 1) / (365.25/12)

Any subject not known to have died at the time of analysis is censored based on the last recorded date on which the subject was known to be alive.

Note: Survival calls are made in the week following the date of DCO for the analysis, and if subjects are confirmed to be alive or if the death date is post the DCO date these subjects are censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" subjects at the time of any formal OS analysis should be obtained by the site personnel by checking the subject's notes, hospital records, contacting the subject's general practitioner and checking publicly-available death registries. In the event that the

subject has actively withdrawn consent to the processing of their personal data, the vital status of the subject can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note: For any other OS analysis performed , in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the subject was known to be alive for those subjects still on treatment (since the *SURVIVE* module is only completed for subjects off treatment if a survival sweep is not performed). The last date for each individual subject is defined as the latest among the following dates recorded on the CRFs:

- AE start, stop and change in severity dates
- Admission and discharge dates of hospitalisation
- Study treatment date
- End of treatment date
- Concomitant medication start and stop dates
- Laboratory test dates
- Date of vital signs
- Date of ECGs
- Physical examination dates
- Cardiovascular data
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anticancer treatment
- Date last known alive on survival status CRF
- End of study date (excluding patients lost to follow-up)
- Patient Reported Outcomes (PRO) questionnaire dates

The proportion of subjects alive at 6, 9, 12, and 18 months is defined as the Kaplan-Meier estimate of OS at that particular month.

"Potential" duration of follow-up for OS, applicable only for OS censored subjects, is defined as follows:
Potential duration of follow-up for OS (months) = (date of OS censoring [date last known to be alive] – date of randomisation + 1) / (365.25/12).

4.3.10.3 Handling of Dropouts and Missing Data

If a subject 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:

a. For Missing day only – using the 1st of the month.

b. For Missing day and Month – using the 1st of January.

4.3.10.4 If death has been recorded but the date is entirely missing, then date of death will be imputed as the date the patient was last known to be alive +1.Primary Analysis of Secondary Endpoint

The analysis of OS is based on the Full analysis set. The number and percentage of subjects experiencing an OS event and Kaplan-Meier plots of OS are presented. Additionally, summaries of the numbers and percentages of subjects who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent are provided. The median OS and two-sided *95*% CI are estimated using the Kaplan-Meier method.

Moreover, the landmark estimates of OS at 6, 9, 12, and 18 months from randomisation are summarised by treatment arm with corresponding 95% CI.

A summary of the potential duration of follow-up for OS is included using median (range). This is presented for censored subjects.

OS is analysed using a log-rank test stratified by the randomisation stratification factors (baseline LDH and IO resistance). The HR together with its 95% CI and p-value is presented. The HR and CI are estimated from a stratified Cox proportional hazard model with ties=Efron, and the CI are calculated using a profile likelihood approach.

4.3.10.5 Sensitivity Analyses of the Secondary Endpoint

A sensitivity analysis is conducted to assess the potential impact of COVID-19 related deaths on OS. That is, subjects who had a death event where the primary or secondary cause was COVID-19 infection or COVID-19 infection was reported as a fatal AE, are censored at the date of their COVID-19 infection related death. If less than five events appear and less than 2% of enrolled subjects are affected the sensitivity analysis will not be carried out.

4.3.10.6 Supplementary Analyses of the Secondary Endpoint

Not applicable

4.3.10.7 Subgroup Analyses

For a detailed description of the Subgroup Analyses, see section 4.3 **Subgroup Analyses**. The results of subgroup analyses will be presented on a forest plot including the HR and 95% CI, along with the results of the overall analysis.

4.3.11 Secondary Endpoint: PK of ceralasertib

PK of ceralasertib alone and when in combination with durvalumab is a secondary endpoint.

See section 4.5

4.3.12 Safety Analysis

The domain safety covers exposure, adverse events, clinical laboratory, vital signs, and electrocardiogram (ECG).

Tables are provided for the safety set, listings are provided for all subjects or the safety set depending on the availability of data.

4.3.12.1 Exposure

4.3.12.1.1 Definitions and Derivations

Duration of exposure

- Duration of exposure to ceralasertib = min(date of last dose where dose > 0 mg +◦21, date of death, date of DCO) – first dose* date +◦1
- Duration of exposure to durvalumab = min(date of last dose where dose > 0 mg +◦20 date of death, date of DCO) – first dose* date +◦1

*First dose refers to earliest dose of either treatment

Actual duration of exposure (weeks)

- Actual exposure to ceralasertib [months] = total exposure to ceralasertib – total duration of dose interruptions and dose delays
- Actual exposure to durvalumab [months] = total exposure to durvalumab – total duration of dose delays

The total duration of dose delays is defined as any length of time in addition to the planned cycle length where the patient has not taken/received any of the planned doses.

- Definition of dose delay for ceralasertib: if the number of days between Day 1 of a cycle and Day 1 of the previous cycle is > 28 days, then the number of additional days above 28 days is the duration of delay.
- Definition of dose delay for durvalumab: for Cycle 1, if the number of days between the first administration of durvalumab on Day 8 and the first administration of ceralasertib on Day 1 is > 7 days, then the number of additional days above 7 days is the duration of the delay. For all subsequent cycles, if the number of days between Day◦8 of a cycle and Day 8 of the previous cycle is > 28 days, then the number of additional days above 28 days is the duration of delay.

The total duration of dose interruption is defined as any length of time during the cycle where the patient has not taken/received any of the planned doses.

- Definition of dose interruptions for ceralasertib: the total number of full-day treatment interruptions, including missed and forgotten doses. Half-day interruptions are not considered in the actual treatment exposure derivation

Note: In case the last days of cycle doses of ceralasertib are not taken (eg, Day 7) and no further medication will be taken, this is an interruption but will not be included in the total duration of dose interruptions as the total treatment exposure already includes this reduction via "date of last dose where dose > 0 mg".

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

Missed and forgotten doses should be recorded as a dose interruption with the reason recorded as "Patient forgot to take dose". These missed or forgotten doses are not included as dose interruptions in the summary tables, but the information appears in the listing for dosing. However, these missed and forgotten doses are considered in the derivation of actual exposure.

Dose intensity of study treatment(s) is addressed by considering relative dose intensity (RDI), where RDI is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. More specifically, RDI is defined as follows:

  - RDI = 100% × d/D, where d is the actual cumulative dose delivered up to min(date of treatment discontinuation, date of death, data of DCO) and D is the intended cumulative dose up to the day of treatment discontinuation. D is the total dose that would be delivered, if there were no modification to dose or schedule.

Intended cumulative dose is calculated by summing the individual doses that should have been received up to and including the last day of study participation according to the planned dose and schedule.

This is given by the intended dose multiplied with the number of administrations that were planned to occur. These are different per treatment component.

- Intended cumulative dose of ceralasertib [mg] = daily dose × number of dosing days =◦daily dose × [7 dosing days per number of complete cycles + dosing days of incomplete cycle]:
  - If mod_{28}(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib + 1) ≤ 7 then:

- Intended cumulative dose of ceralasertib =◦240◦mg × 2 × [7 × floor((min(date of treatment discontinuation, date of DCO) – date of first dose of ceralaserib +◦1)/28) + mod_{28}(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib + 1)]

- Else if mod_{28}(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib + 1) > 7 then:
  - Intended cumulative dose of ceralasertib = 240 mg × 2 × [7 × floor(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib +◦1)/28) + 7]

- Intended cumulative dose of durvalumab [mg] = dose × number of complete cycles =◦1500◦mg × ceil(min(date of treatment discontinuation, date of DCO) – date of first dose of durvalumab + 1)/28)

Note: ceil() refers to rounding up and floor() refers to rounding down to the next integer. mod_{28}(n) is the number n modulo 28.

Actual cumulative dose is calculated by summing all administered doses:

- ceralasertib dose = Actual dose recorded x actual frequency x duration
- durvalumab dose = Actual dose recorded

4.3.12.1.2 Presentation

Duration of exposure to Investigational Products (IP)(s) in weeks and cycles are summarised by descriptive statistics and by frequency for each treatment arm. Dose intensity is summarised by descriptive statistics. Exposure to IP(s) i.e. total amount of study drug received is listed for all subjects. Exposure swimmer plot(s) are produced, with a line presented for each subject to display relevant exposure and disposition details.

Dosing deviations for IP(s) are summarised with reasons for deviations for the following categories: delays, reductions, and interruptions. Dosing delays are derived based on the scheduled dosing dates using the previous dose given as reference. The number and percentage of subjects with dosing delays and total dose delays per subject are summarised.

4.3.12.2 Adverse Events

4.3.12.2.1 Definitions and Derivations

The Medical Dictionary for Regulatory Activities (MedDRA) (version 25 or higher) is used to code the AEs. AEs are graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE) (version 5.0).

Treatment emergent adverse events (TEAEs) are all AEs which onset or worsen in severity following the first administration of IP within the duration of the treatment period, up to

and including 30 days after the last dose of study treatment for ceralasertib monotherapy or 90 days after last dose of study treatment for ceralasertib and durvalumab combination, and prior to the start of any subsequent anti-cancer therapy,. Worsening in severity is determined by comparison with the pre-treatment CTCAE grade of the AE recorded closest to the start of dosing.

AEs with a missing start time or where time is not collected, which occur on the same day as first IP administration, are reported as treatment emergent.

For rules on missing or partial dates, see Section 3.4.6.

SAEs

A serious adverse event (SAE) is any AE that:

- Results in death
- Is immediately life-threatening
- Requires inpatient hospitalisation or prolongation of existing hospitalisation
- Results in persistent or significant disability/incapacity.
- Is a congenital anomaly/birth defect in offspring of the subject
- Is an important medical event that may jeopardize the subject.

Adverse Events of Special Interest (AESI)

Adverse events of special interests are events of scientific and medical interest specific to the further understanding of ceralasertib and durvalumab safety profile and require close monitoring and rapid communication by the investigators to AstraZeneca. An AESI can be serious or non-serious. All AESIs will be recorded in the eCRF. Additionally, AESI will be identified programmatically using a definition file maintained by Astra Zeneca.

Adverse events of special interests for durvalumab include but are not limited to events with a potential inflammatory or immune-mediated mechanism and which may require more frequent monitoring and/or interventions such as steroids, immunosuppressants and/or hormone replacement therapy. An immune-mediated AE is defined as an AESI that is associated with drug exposure and is consistent with an immune-mediated mechanism of action and where there is no clear alternate aetiology.

If the investigator has any questions in regards to an event being an immune-mediated AE, the investigator should promptly contact the AstraZeneca study physician or clinical lead.

Adverse events of special interests/ immune-mediated AE observed with anti-PD-L/PD-1 agents such as durvalumab include pneumonitis, hepatitis, diarrhoea/colitis, intestinal perforation, endocrinopathies (hypo- and hyperthyroidism, adrenal insufficiency,

hypophysitis/hypopituitarism and Type 1 diabetes mellitus), nephritis, rash/dermatitis, myocarditis, myositis/polymyositis, pancreatitis and rare/less frequent events (including, but not limited to haematological events, neuromuscular toxicities [such as myasthenia gravis and Guillain-Barré syndrome], non-infectious encephalitis, non-infectious meningitis, pericarditis, rheumatological events, sarcoidosis, skin events, uveitis [and other events involving the eye] and vasculitis).

In addition, infusion-related reactions and hypersensitivity/anaphylactic reactions with a different underlying pharmacological aetiology are also considered AESIs.

More detailed guidelines for their evaluation and treatment are described in detail in the Dose Modification and Toxicity Management Guidelines (version 28-Oct-2021). These guidelines have been prepared by the sponsor to assist the investigator in the exercise of his/her clinical judgement in treating these types of toxicities. These guidelines apply to AEs considered causally related to the study treatment/study regimen by the reporting investigator.

There are no currently identified AESI for Ceralasertib but if any are identified, these will be reported in similar/appropriate outputs to the durvalumab AESI.

Other significant adverse events (OAE)

During the evaluation of the AE data as part of the monthly offline listing review process, an AstraZeneca medically qualified expert reviews the list of AEs that were not reported as SAEs and AEs leading to discontinuation of study treatment after Parexel's review took place. Based on the expert's judgement, adverse events of particular clinical importance may, after consultation with the Global Safety Physician, be considered OAEs and reported as such in the CSR. A similar review of laboratory values, vital signs, ECGs and other safety assessments are performed for identification of other significant adverse events. This review takes place prior to database lock, and any AEs identified are fully documented in meeting minutes. Further review following database lock may result in ad-hoc OAEs being identified, in this case, the OAEs and resulting summaries are fully documented in the CSR.

Examples of these are 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.

4.3.12.2.2 Presentation

All TEAEs will summarised and listed. AEs which are not treatment emergent will be listed for the Safety analysis set and will be listed individually by subject and treatment arm.

TEAEs will be counted once for each subject for calculating percentages of subjects experiencing TEAE. In addition, if the same TEAE occurs multiple times within a particular subject, the highest severity and level of relationship observed will be reported. For tables by MedDRA system organ class (SOC) and MedDRA preferred term (PT), subjects with multiple TEAEs will be counted once for each SOC/PT.

An overall summary table of the number of subjects experiencing each category of adverse event will be produced. The number of subjects experiencing treatment emergent adverse events by MedDRA SOC and PT will be summarised and incidence rates will be provided. Severity and relationship to IP will be summarised. Further splits by CTCAE grade, causal relationship to IP and adverse events with Grade 3-4 will be also summarised.

Separate tables will present adverse events leading to discontinuation, serious adverse events, IP-related adverse events, and other significant adverse events.

Details of any deaths will be summarised and listed for all subjects. AEs leading to death will also summarised.

SAEs

SAEs are summarised as described above for the TEAEs.

AEs of special interest

Grouped summary tables of certain MedDRA preferred terms are produced and may also show the individual preferred terms which constitute each AESI grouping. These groupings may be defined using MedDRA terms, SMQs (standardized MedDRA queries) and PTs (preferred terms). Groupings are provided by the coding team prior to DBL, and a listing of the preferred terms in each grouping is provided. Summaries of the above-mentioned grouped AE categories include number and percentage (%) of subjects who have:

- At least one AESI presented by event outcome
- At least one AESI causally related to IP
- At least one AESI leading to discontinuation of study treatment.

A summary of total duration (days) of AESI are provided for events which have an end date and this may be supported by summaries of ongoing AESIs at death and, separately, at data cut-off.

4.3.12.3 Clinical Laboratory, Blood Samples

4.3.12.3.1 Definitions and Derivations

Laboratory tests are grouped according to chemistry and haematology. Laboratory parameters are assessed at baseline as well as throughout the study.

For chemistry and haematology parameters, laboratory abnormalities with toxicity grades according to the NCI CTCAE version 5.0 are derived.

Change from baseline in haematology and clinical chemistry endpoints is calculated for each post-dose visit. CTC grades are calculated at each visit. Maximum post-baseline CTC are also calculated. Absolute values are compared to the local laboratory reference range and classified as low (below range), normal (within range or on limits of range) and high (above range). All values classified as high or low are flagged on the listings.

Liver Function Parameters

Subjects with elevated post-baseline alanine aminotransferase (ALT), aspartate aminotransferase (AST) or Total Bilirubin that fall into these categories are identified.

Table 7 Liver Function Parameters

Liver Function Parameters	Category
ALT	• ≥3 × – ≤5 × ULN
AST	• ≥3 × – ≤5 × ULN
Total bilirubin	• >3 × – ≤5 × ULN
ALT or AST	• ≥3 × – ≤5 × ULN
Potential Hy's law	• (AST ≥ 3 × ULN or ALT ≥ 3 × ULN) and (Total Bilirubin ≥ 2×ULN)[a]

ULN: upper limit of normal range.

[a] It includes all subjects who have ALT or AST≥3xULN and total bilirubin (BILI) ≥2xULN, and in which the elevation in transaminases precede or coincide with (that is, on the same day as) the elevation in BILI.

4.3.12.3.2 Presentation

The change in each laboratory parameter from baseline to each post-baseline visit are summarised graphically.

Laboratory abnormalities occurring from the start of IP administration to the last assessment on study are presented. Worst toxicity grade, rates of grade 3-4 toxicity, and grade shifts of 2 or more from baseline to the maximum grade are presented. Summaries indicating hyper- and hypo- directionality of change are produced, where appropriate. Laboratory parameters that cannot be graded via NCI CTCAE are summarised with frequencies of post-baseline laboratory values categorized as low (L), normal (N), or high (H) using laboratory normal ranges comparing to baseline.

Listings are provided for all laboratory results.

Liver Function Parameters

Number and percentage of subjects with elevated post-baseline ALT, AST or Total Bilirubin (BILI) are tabulated. Individual subject data with elevated ALT or AST plus BILI falling into the "Potential Hy's law" are summarised and listed.

4.3.12.4 Clinical Laboratory, Urinalysis

4.3.12.4.1 Definitions and Derivations

Change from baseline in urinalysis endpoints are calculated for each post-dose visit. CTC grades are calculated at each visit. Maximum post-baseline CTC is also calculated. Absolute values are compared to the local laboratory reference range and classified as low (below range), normal (within range or on limits of range) and high (above range). All values classified as high or low are flagged in the listings.

4.3.12.4.2 Presentation

Listings are provided for urinalysis. Urinalysis endpoints are summarised by study visit which will include descriptive statistics for the value of the parameters and the changes from baseline. Absolute values and change from baseline at each timepoint are presented. Maximum post-baseline CTC grades which are not normal will be summarized.

For urinalysis, shift from baseline to worst on treatment results are presented.

Listings are provided for urinalysis.

4.3.12.5 Vital Signs

4.3.12.5.1 Definitions and Derivations

Vital signs (weight, body temperature, systolic and diastolic blood pressure, respiratory rate and heart [pulse] rate) are performed at timepoints as specified in the schedule of assessments. Any changes in vital signs should be recorded as an AE if applicable.

4.3.12.5.2 Presentation

Vital signs are summarised by study visit which includes descriptive statistics for the value of the parameters and the changes from baseline. Absolute values and change from baseline for vital signs data at each timepoint are presented using box plots.

4.3.12.6 Electrocardiogram

4.3.12.6.1 Definitions and Derivations

Electrocardiogram (ECG) parameters are assessed in triplicate at baseline and subsequent triplicate measurements and should only be taken if clinically indicated. Single ECG readings are collected at visits other than baseline. ECG parameters include: PR, RR, QRS, QT, and QTcF. The QTcF is considered as the primary correction method to assess subject cardiac safety.

From these resting 12-lead ECGs values of the QT and RR intervals and the QT interval corrected for heart rate using Fridericia's correction (QTcF) is derived using the following formula:

- QTcF = QT/RR^ (1/3) where RR is in seconds

The values of QTcF (milliseconds) are re-derived from the values of RR and QT during the creation of the reporting database.

4.3.12.6.2 Presentation

ECG parameters are listed.

4.3.12.7 Echocardiogram/Multigated Acquisition

4.3.12.7.1 Definitions and Derivations

An echocardiogram or Multigated Acquisition (MUGA) scan to assess left ventricular ejection fraction (LVEF) will be performed at the visits screening and treatment discontinuation and as clinically indicated.

4.3.12.7.2 Presentation

LVEF will be listed.

4.3.12.8 Bleeding events

Information obtained by the bleeding event questionnaires is depicted in a listing.

4.4 Exploratory Analyses

4.4.1 CCI

CCI

CCI

CCI

CCI

- CCI :
  - CCI
  - CCI
  - CCI
  - CCI
  - CCI
- CCI :
  - CCI
  - CCI
  - CCI
- CCI
  - CCI
  - CCI
  - CCI
  - CCI
  - CCI
  - CCI
- CCI

CCI

CCI

CCI

CCI :

- CCI
- CCI

CCI

CCI

CCI .

CCI

CCI

CCI

CCI

CCI

- CCI
- CCI ,
- CCI ,
- CCI
- CCI
- CCI
- CCI .

CCI

- CCI

CCI

- CCI
- CCI
- CCI
- CCI
- CCI

CCI

CCI

CCI

CCI

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted].

CCI [redacted]

CCI [redacted]

CCI [redacted]

.

CCI

CCI

- CCI
- CCI
- CCI
- CCI
- CCI

CCI

CCI

4.4.2 CCI

CCI

CCI

CCI

4.4.3 Biomarkers

Predictive biomarkers of ceralasertib including CCI will be measured from samples collected for testing at Screening (Days -21 to -1) [baseline assessment] and on-treatment, Cycle 0 (on Day 7) of ceralasertib monotherapy [on-treatment assessment]. All analysis of this data will be performed on the PD Analysis Set.

CCI results related analyses are beyond the scope of this SAP but may be analysed at a later timepoint according to the below descriptions which might be extended to explore biomarkers even further.

CCI [redacted] results are summarised descriptively at baseline and on-treatment assessments, including change from baseline. The summary statistics provided include number of observations (n), geometric mean (calculated as exp[μ], where μ is the mean of the data on a logarithmic scale), coefficient of variation (CV, calculated as 100 √ [exp(s2)-1], where s is the standard deviation of the data on a log scale), arithmetic mean, arithmetic SD, median, minimum and maximum.

Boxplots of change from baseline will be produced on a log scale for on-treatment timepoint. The geometric means are presented using a diamond symbol in these plots.

Plots will be produced showing each patient's baseline and on-treatment results over time on a log scale with separate lines indicating the results for each patient.

Analysis of additional biomarkers, if required, will be carried out by AZ Biomarker team and covered in a separate document.

4.5 Pharmacokinetics

Plasma concentrations-time data are not analysed to determine the PK parameters. Pre-dose and post-dose plasma concentrations for each scheduled time-point are summarised by treatment, PK Day and Cycle using appropriate descriptive statistics. PK concentration data are listed for each subject in the safety analysis set.

Individual concentrations may be excluded from summary tables, graphical or statistical analysis for legitimate scientific reasons. Any exclusions, together with justification for the exclusion, is clearly documented in the CSR. These individual data are still presented in the listings, but are flagged to identify that they are excluded from summary outputs.

The following descriptive statistics are presented for plasma concentrations:

- n
- n below lower limit of quantification (LLOQ)
- geometric mean (gmean)
- geometric standard deviation (gSD)
- gmean±gSD
- geometric coefficient of variance (%) (gCV)
- arithmetic mean of non log-transformed data (mean)
- standard deviation of non log-transformed data (Std Dev)
- coefficient of variance (%) (CV)
- median
- minimum (min)
- maximum (max)

The gmean is calculated as exp(μ), where μ is the mean of the data on the natural log scale.

The gSD is calculated as exp(σ), where σ is the standard deviation of the data on the natural log scale.

The gCV is calculated as 100 x sqrt[exp(s^2)-1], where s is the Std Dev of the data on the natural log scale.

The gmean ± gSD (gmean-gSD and gmean+gSD) is calculated as exp[μ ± s].

Protocol scheduled times are used to present the PK concentration summary tables and corresponding gmean concentration-time figures.

Boxplots of plasma concentration at pre-dose and 1h post dose is presented for cycles until cycle 3 for both Ceralasertib and Durvalumab separately. The geometric means are presented using a diamond symbol in these plots.

Handling of Non-Quantifiable Concentrations

Individual concentrations below the LLOQ of the bioanalytical assay are reported as NQ in the listings with the LLOQ defined in the footnotes of the relevant TFLs. Individual plasma concentrations that are Not Reportable are reported as NR and those that are missing are reported as NS (No Sample) in the listings. Plasma concentrations that are NQ, NR or NS are handled as follows for the provision of descriptive statistics:

- Any values reported as NR or NS are excluded from the summary tables and corresponding figures.
- At a time point where less than or equal to 50% of the concentration values are NQ, all NQ values are set to the LLOQ, and all descriptive statistics are calculated accordingly.
- At a time point where more than 50% (but not all) of the values are NQ, the gmean, gmean ± gSD and gCV% are set to NC. The maximum value is reported from the individual data, and the minimum and median are set to NQ.
- If all concentrations are NQ at a time point, no descriptive statistics are calculated for that time point. The gmean, minimum, median and maximum are reported as NQ and the gCV% and gmean ± gSD as NC.
- The number of values below LLOQ (n < LLOQ) are reported for each time point together with the total number of collected values (n).
- If data are available for less than three subjects, no summary statistics other than minimum, maximum and n are presented.

Graphical presentation of PK data

- Box plots of pre-dose and post-dose concentration mean (arithmetic mean and/or gmean) by PK visit are displayed on the same plot. Preferably, cycles are differentiated by colours on the same figure.

Precision and Rounding Rules for Pharmacokinetic Data /PK concentration data

PK concentration data listings are presented to the same number of significant figures as the data received from the bioanalytical laboratory (usually but not always to 3 significant figures) and against the same units as received.

PK concentration descriptive statistics present 4 significant figures with the exception of the min and max which present 3 significant figures and n and n<LLOQ which present as integers.

4.6 Immunogenicity

Immunogenicity results are listed for each subject and summarised for the Safety analysis set. Number and percentage of subjects in the following categories are provided.

- Anti-drug antibody (ADA) positive at baseline and/or post-baseline visits.
- Persistent positive, defined as positive at ≥2 post-baseline assessments (with ≥16 weeks between first and last positive) or positive at last post-baseline assessment.
- Transient positive, defined as negative at last post-baseline assessment and positive at only one post-baseline assessment or at ≥2 post-baseline assessments (with <16 weeks between first and last positive).
- Treatment-boosted, defined as baseline positive ADA titre that was boosted to a 4-fold or higher level following drug administration.

All valid assay results from subjects who receive any study drug are included in immunogenicity summaries. Blood samples collected outside of the protocolled window are summarised at the closest nominal time point that does not already have a value. All subjects with titre information are shown in the data listing.

5 INTERIM ANALYSIS

The IDMC for the MONETTE study is constituted to provide an independent, external, and unbiased assessment of safety, tolerability and efficacy during the conduct of the trial. The IDMC is an independent external group of experts and consists of two independent

therapeutic area experts and a biostatistician with experience in the monitoring of randomised clinical studies. IDMC meetings consist of different sessions. In general, no information deemed to unblind attendees is presented or discussed during open session. Closed sessions are attended by the IDMC members and a very limited number of unblinded people including an independent statistician and data is presented and discussed in a manner that reveals treatment group details. Further information about the IDMC and the format of the IDMC meetings can be found in the IDMC charter.

Interim Analysis[CCI]

The [CCI] interim analysis is conducted once approximately [CCI] subjects have been enrolled in total and had the opportunity for approximately [CCI] follow up. The [CCI] follow up enables [CCI] RECIST scans for confirmation of response. The rationale for this approach is to ensure the objective response rate (ORR) is not artificially underestimated at an interim analysis e.g., if there are subjects ongoing treatment with unconfirmed response who are yet to have their confirmatory scan.

The IDMC may recommend to stop recruitment to the ceralasertib plus durvalumab treatment arm if there are [CCI] or [CCI] responses out of [CCI] subjects, i.e., ORR ≤ [CCI]%. This stop rule corresponds to < [CCI]% predictive probability of success should recruitment continue to the end of the study. There is [CCI]% risk to incorrectly stop if the true ORR = [CCI]%.

The IDMC may recommend to stop recruitment to the ceralasertib monotherapy arm if there are [CCI] responses out of [CCI] subjects i.e., ORR= [CCI]%. Should the combination treatment arm be terminated at the interim analysis, the monotherapy arm may also be terminated if there are [CCI] or [CCI] responses out of [CCI] subjects, i.e., ORR ≤ [CCI]% since in this situation it is unlikely to achieve the target ORR to warrant future development should it continue to the end of the study.

Interim Analysis[CCI]

The [CCI] interim analysis is carried out once approximately [CCI] subjects have been enrolled in total and had the opportunity for approximately [CCI] weeks follow up. The [CCI] weeks follow up enables [CCI] RECIST scans for confirmation of response and also preliminary characterisation of the durability of response, which is a key secondary endpoint.

The IDMC may recommend to stop recruitment to combination treatment arm at the [CCI] interim analysis if there are [CCI] or [CCI] responses out of [CCI] subjects, i.e., ORR ≤ [CCI]%. The IDMC may recommend to stop recruitment to monotherapy treatment arm if there are [CCI] or [CCI] responses out of [CCI] subjects i.e., ORR ≤ [CCI]% and the exact [CCI] [CCI]% CI for the proportion of subjects with non-progressive disease at week [CCI] (Disease Control Rate) is < [CCI]%. Should the combination treatment arm be terminated at the interim analysis, the

monotherapy arm may also be terminated if there are [redacted] or [redacted] responses out of [redacted] subjects, i.e., ORR ≤ [redacted]% since in this situation it is unlikely to achieve the target ORR to warrant future development should it continue to the end of the study.

In addition to the ORR, the BOR is presented as well.

6 REFERENCES

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika 1934; 26:404–416.

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

Frewer P, Mitchell P, Watkins C, Matcham J. Decision-making in early clinical drug development. Pharma Statist. 2016; 15:255-263.

Guo W, Wang SJ, Yang S, Lynn H, Ji Y. A Bayesian interval dose-finding design addressing Ockham's razor: mTPI-2. Contemporary Clinical Trials 2017; 58:23-33.

Hatcher RA, Trussell J, Nelson AL, Cates W Jr, Stewart F, Kowal D, eds. Contraceptive technology. 19th edition. New York: Ardent Media 2007(a): 24. Table 3-2.

Jennison C, Turnbull BW. "Group Sequential Methods with Applications to Clinical Trials," Chapman & Hall/CRC, Boca Raton 2000.

Ji Y, Liu P, Li Y, Bekele BN. A modified toxicity probability interval method for dose-finding trials. Clinical Trials 2010; 7:653-63.

Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clinical Trials 2008; 5:93-106.

Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al on behalf of the RECIST working group. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol. 2017; 18: e143–52.

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: d533ac00001-sap-ed-4		
Document Title:	Statistical Analysis Plan Edition 4	
Document ID:	Doc ID-004936813	
Version Label:	4.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
19-Mar-2024 11:31 UTC	PPD	Author Approval
19-Mar-2024 12:21 UTC	PPD	Content Approval
19-Mar-2024 10:50 UTC	PPD	Content Approval

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

STATISTICAL ANALYSIS PLAN	
Study Code	D533AC00001
Edition Number	1.0
Date	15-Mar-2024

MONETTE (Sub-study): A Non-randomised, Open-Label, Biopsy Sub-study In Patients Suitable for 3 Mandatory Biopsies.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or Specialized Term	Definition
ADA	Anti-Drug Antibody
AE	Adverse event
AESI	Adverse Event of Special Interest
ALT	Alanine Aminotransferase
ARID1A	AT-rich interactive domain-containing Protein 1A
AST	Aspartate Aminotransferase
ATM	Ataxia Telangiectasia Mutated
ATR	Ataxia Telangiectasia and Rad3-Related Protein
BILI	Total Bilirubin
BMI	Body mass index
BOR	Best Objective Response
BRAF	B-Rapidly Accelerated Fibrosarcoma Gene
BSR	Baseline Scaled Ratio
CI	Confidence Interval
COVID-19	Coronavirus Disease 2019
CR	Complete Response
CRF	Case Report Form
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for AEs
CV	Coefficient of Variation
DCO	Data Cut-Off
DoR	Duration of Response
DRM	Data Review Meeting
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
GCP	Good Clinical Practice
gCV	Geometric Coefficient of Variance
Gmean	Geometric Mean
gSD	Geometric Standard Deviation
HR	Hazard Ratio
IO	Immune-Oncology

ICF	Informed Consent Form
IL	Item Library
IP	Investigational Product
IPD	Important Protocol Deviation
KIT	Proto-oncogene c-KIT
LDH	Lactic Acid Dehydrogenase
LLOQ	Lower Limit of Quantification
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary for Regulatory Activities
max	Maximum
min	Minimum
MMRM	Mixed Model for Repeated Measures
MSI	Activating Mutations and Microsatellite Instability
MUGA	Multigated Acquisition
NA	Not Applicable
NE	Not Evaluable
NF1	Neurofibromatosis type 1
NTL	Non-Target Lesion
NR	Not Reportable
NRAS	NRAS proto-oncogene, GTPase
NS	No Sample
OAE	Other significant adverse event
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease
PD-(L)1	Programmed Cell Death 1 Ligand
PFS	Progression Free Survival
PK	Pharmacokinetics
PR	Partial Response
PT	Preferred Term
RDI	Relative Dose Intensity
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable Disease
SE	Standard Error
SMQ	Standardized MedDRA Queries

SOC	System Organ Class
Std Dev	Standard Deviation
TL	Target Lesion
TLF	Tables Listings Figures
TACE	Transarterial Chemoembolization
TARE	Transarterial Radioembolization
TEAE	Treatment Emergent Adverse Events
CCI	CCI
TOC	Table of Contents
TTR	Time to Objective Response
ULN	Upper Limit of Normal

AMENDMENT HISTORY

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
N/A	Click or tap to enter a date.	Initial approved SAP for the biopsy sub-study	N/A	N/A

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of the biopsy sub-study of D533AC00001 supporting the clinical study report (CSR). The reader is referred to the clinical study protocol (CSP) and the case report form (CRF) for details of study conduct and data collection. This statistical analysis plan (SAP) is based on Amendment *2* of the CSP dated 14 March 2022. In the event of future amendments to the protocol, this SAP may be modified to account for changes relevant to the statistical analysis.

A separate SAP was created for the main study. There are no plans to pull clinical data for efficacy from the main study and the sub-study.

2 CHANGES TO PROTOCOL PLANNED ANALYSES

Not applicable.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

The details of the analyses planned during the conduct of this study are outlined below and summarised in the following table.

Table 1: Summary of planned analyses

Analysis	Trigger	Scope
Primary Analysis	Once all enrolled subjects have had a minimum of CCI months of follow-up after start of study intervention or have discontinued from the study, whichever occurs first.	All available data
CCI Analysis	Once CCI % of the subjects have died or CCI months after last subject is enrolled, whichever is the earliest event.	All available data

DCO: Data cut-off

The primary analysis for the biopsy sub-study is conducted once all enrolled subjects have had a minimum of CCI months of follow-up after start of study intervention or have discontinued from the study, whichever occurs first. At this primary analysis the CSR for

the biopsy sub-study is written, which may be finalised after the CSR is finalised for the main study. A CCI analysis is performed either CCI months after the last subject is enrolled, or CCI % of subjects have died, whichever is the earlier event.

3.2 General Considerations

The below-mentioned general principles will be followed throughout the study:

- Descriptive statistics will be used for all variables, as appropriate.
  - Continuous variables will be summarised by the number of observations, mean, standard deviation, median, upper, and lower quartiles where indicated, minimum and maximum. For log-transformed data (ie, pharmacokinetic [PK] concentration data) it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum, and maximum.
  - If data are available for less than 3 subjects, no summary statistics other than minimum, maximum and number of observations are presented.
  - Categorical variables will be summarised by frequency counts and percentages for each category. Unless otherwise stated, percentages will be calculated based on the analysis set population and by timepoint as appropriate.
- For continuous data, the mean, standard deviation, and median are rounded to one additional decimal place compared to the original data. Minimum and maximum are displayed with the same accuracy as the original data.
- Time-to-event variables will be presented using the Kaplan-Meier (KM) methodology, including median time calculated from the KM curves.
- For categorical data, percentages are rounded to one decimal place with the exception of 100% which is presented as a whole number.
- SAS® version 9.4 (or higher) and/or other validated software (as appropriate) will be used for the analyses. In case R is used, the R build version and the version of each used R package will be documented.

3.2.1 Sample Size Determination

Approximately CCI patients are planned to be enrolled such that approximately 30 patients have at least 2 evaluable paired biopsies (at the baseline and off ceralasertib treatment) including at least 15 patients with at least 3 evaluable biopsies (at baseline, on ceralasertib treatment and off ceralasertib treatment).

Sample sizing is based on expected CCI fold increase in CD8+ T-cells infiltration. The CCI fold assumption represents the observed increase in proliferating CD8+ T-cells infiltration in

association with Pembrolizumab and Nivolumab or Nivolumab/Ipilimumab efficacy inmelanoma (Tumeh et al, 2014; Grasso et al, 2020).

The sample size of 30 evaluable patients is expected to give adequate sample size to assess changes of CD8+ T-cells from baseline.

3.2.2 Study Treatment

During Cycle 0, patients will be treated with ceralasertib 240 mg BD, as oral tablets, from Days 1 to 7, Q28D, followed by an off-treatment period between Days 8 to 28 (i.e. no-dose period). From Cycle 1, patients will be treated with the combination of ceralasertib 240 mg BD, oral tablets, from Days 1 to 7, plus durvalumab 1500 mg Day 8, as an IV infusion, Q28D, until progressive disease (confirmed RECIST 1.1-defined radiological progression), unacceptable toxicity, withdrawal of consent, or if a study treatment discontinuation criterion is met.

3.2.3 Baseline

In general, the last observed measurement prior to first dose of study intervention will be considered the baseline measurement.

For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, serves as sufficient evidence that the assessment occurred prior to first dose. Assessments on the day of the first dose where neither time nor a nominal predose indicator are captured is considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose. If no value exists before the first dose/administration, then the baseline value is treated as missing.

For clinical data, if two measurements are equally eligible to assess patient status at baseline (e.g., two assessments both on the same date with no time recorded), the average is used as the baseline value. For non-numeric laboratory tests (ie, some of the urinalysis parameters) where taking the average is not possible, the best value (value closest to none/normal/negative) is used as baseline as this is most conservative. In the scenario where there are two eligible baseline assessments recorded on the same day, one with time recorded and the other without time recorded, the one with the time recorded is selected as baseline. Where safety data will be summarised over time, time on study is calculated in relation to date of first treatment.

In all summaries change from baseline variables are calculated as the post-treatment value minus the value at baseline. The percentage change from baseline is calculated as (post-baseline value - baseline value)/(baseline value) x 100. For any endpoint subjected to log transformation, the change from baseline calculated and summarised on the log scale are back-transformed and presented as a ‘baseline scaled ratio’ (BSR). Percentage change is then calculated as (BSR - 1) × 100.

3.2.4 On-treatment

For the purposes of summarizing safety data assessed at visits, in addition to baseline data, only on-treatment data are included in the summary tables. On-treatment data is defined as data on/after the first dose of study treatment (including Cycle 0) and with assessment date up to and including the date of last study treatment + 90 days (safety follow-up), and prior to the start of any subsequent anti-cancer therapy.

For biomarker data the on-treatment assessment is strictly defined as Cycle 0 Day 7. The Cycle 0 Days 15 – Day 28 period will be defined as off-treatment.

3.2.5 Visit Window

For safety, time windows are defined for any presentations that summarise values by visit. The following conventions apply:

- For safety data study day reference is date of first dose of study treatment as Day 1, for PK the reference is the time of study treatment administration on the day PK blood samples are taken, for efficacy data study day references date of first dose as Day 1.
- The time windows are exhaustive so that data recorded at any timepoint has the potential to be summarised. Inclusion within the time window are based on the actual date and not the intended date of the visit.
- All unscheduled visit data have the potential to be included in the summaries.
- The window for the visits following baseline are constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit is Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as scheduled visit day + (duration between scheduled visits)/2. (i.e .don't apply the minus 1 day).
- Subjects may delay durvalumab dosing under certain circumstances.
    - Dosing may be delayed per the Dosing Modification and Toxicity Management Guidelines, due to either an immune or a non-immune-related AE.
    - If dosing must be delayed for reasons other than treatment-related toxicity, dosing will resume as soon as feasible.
    - Dosing intervals of subsequent cycles may be shortened as clinically feasible in order to gradually align treatment cycles with the schedule of tumour efficacy (RECIST) assessments.

- Visit windowing is done separately for each assessment based on the schedule of events specific to that assessment.

- Should Study Day be missing (due to partial dates), then visit is assigned to the nominal visit at which that assessment was recorded, and no windowing will be performed for that specific assessment.
- Visit windowing is conducted up to and including the "Last dose of study treatment" visit. That is, the "Last dose of study treatment" visit is reassigned to a scheduled visit based on the study day that visit occurred at.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment are used (regardless of where it falls in an interval).
- Listings display all values contributing to a time point for a subject.
- For visit based summaries, if there is more than one value per subject within a time window then the closest value to the scheduled visit date are summarised, or the earlier, in the event the values are equidistant from the nominal visit date. The listings highlight the value for the subject 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 or not the value is close to the scheduled visit date.
- For summaries at a subject level, all values are included, regardless of whether they appear in a corresponding visit based summary, when deriving a subject level statistic such as maximum.

3.2.6 Handling of Unscheduled Visits

Unscheduled visits are included in the method of assigning data to scheduled visits described in the rules in Section 3.2.5 above. Unscheduled visits are not included as a separate visit in the summary tables.

For summaries at patient level, such as of extreme values, all post-baseline values collected are used to derive a patient level statistic including those collected at unscheduled visits and regardless of whether they appear in the corresponding visit-based summary.

3.2.7 Multiplicity/Multiple Comparisons

No adjustments for multiplicity are planned.

3.2.8 Missing Dates

Generally, the imputation of dates is used to decide if an observation is treatment emergent for adverse events (AEs) or concomitant medications. Flags are retained in the database indicating where any programmatic imputation has been applied, and in such cases, any imputed dates should not be used to calculate durations, where the results would be less accurate.

The following are the guidelines used when partial dates are detected in the study:

- For missing diagnostic dates (e.g. disease diagnosis), if day and/or month are missing use 01 and/or Jan. If year is missing, put the complete date to missing.

- For missing AE and concomitant medication start dates, the following is applied:
    a. Missing day–- impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date.

    b. Missing day and month–- impute 1st January unless year is the same as first dose date then impute first dose date.

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

    d. Imputed start date should be no later than the end date.

- For missing AE and concomitant medication end dates, the following is applied:
    a. Missing day–- impute the last day of the month unless month is the same as month of study discontinuation, then impute as study discontinuation.

    b. Missing day and month–- impute 31st December unless year is the same as year of study discontinuation then impute study discontinuation date.

    c. Completely Missing – If an AE/medication has a completely missing end date then it is treated as ongoing, unless this is for a prior anti-cancer medication then impute the date of informed consent.

- If a subject is known to have died where only a partial death date is available then the date of death is 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:
    a. For Missing day only – using the 1st of the month.

    b. For Missing day and Month – using the 1st of January.

For patients with a treatment start date for ceralasertib, the following rules are applied for missing treatment end dates for ceralasertib:

• Impute the missing ceralasertib end date as min(start date of ceralasertib + 6 days, end of study date, DCO, date of investigator decision to stop ceralasertib, start date of durvalumab – 1 day)..

For partial subsequent anti-cancer therapy dates, the following rules will be applied for missing start dates:

• Missing day: If the month is the same as treatment end date then impute to the day after treatment, otherwise first day of the month.

• Missing day and month: If year is the same as treatment end date then impute to the day after treatment, otherwise 1st January of the same year as anti-cancer therapy date.

For time to event endpoints, dropouts and missing data are handled according to the censoring rules detailed within the relevant sections for the endpoint .

Other rules for handling missing data are described under the derivation rules for that particular variable

3.2.9 Global/Country Situation

Impact of global or country situation is captured in the eCRF. Impact in terms of

- Disposition (discontinued IMP or withdrew study due to global/country situation)
- Disruption (visit impact, drug impacted)
- Important Protocol Deviations (IPDs) (Subjects with at least one IPD related to global/country situation)

will be summarized descriptively based on the safety Analysis Set.

3.2.10 Derivations of RECIST 1.1 Visit Responses

For all subjects, the RECIST tumour response data is used to determine each subject's visit response according to RECIST version 1.1. It is also used to determine if and when a subject has progressed in accordance with RECIST and their best objective response to study treatment.

Baseline radiological tumour assessments should be performed no more than 21 days before the start of study treatment and ideally as close as possible to the start of study treatment. Tumour assessment is conducted every 8 weeks (±1 week) after the first dose (Cycle 1 Day 1) up to 18 months, then every 12 weeks (±1 week) until objective disease progression as per RECIST 1.1, irrespective of treatment decisions. Treatment start date will only be used to determine tumour assessment schedule, date of first dose will be used for analysis.

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

From the investigator's review of the imaging scans, the RECIST tumour response data is used to determine each subject's visit response according to RECIST version 1.1. At each visit, subjects are programmatically assigned a RECIST 1.1 visit response of CR, PR, SD or PD, using the information from TLs, NTLs and new lesions and depending on the status of

their disease compared with baseline and previous assessments. If a subject has had a tumour assessment that cannot be evaluated then the subject is assigned a visit response of NE, (unless there is evidence of progression in which case the response is assigned as PD).

Please refer to Table 5 for the definitions of CR, PR, SD and PD overall visit response.

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

If there are patients with non-measurable disease or no evidence of disease assessed at baseline RECIST has been modified to allow the assessment of progression due to new lesions in subjects with no evidence of disease at baseline.

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 subject 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 target lesions (TLs). If more than one baseline scan is recorded then measurements from the one that is closest and prior to first dose is used to define the baseline sum of TLs. It is 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, is selected.

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

Note: For subjects who do not have measurable disease at entry (i.e. no TLs) but have non-measurable disease, evaluation of overall visit responses are based on the overall NTL assessment and the absence/presence of new lesions (see section below for further details). If a subject does not have measurable disease at baseline then the TL visit response is Not Applicable (NA).

For subjects with no disease at baseline (i.e. no TLs and no NTLs), evaluation of overall visit responses is based on absence/presence of new lesions. If no TLs and no NTLs are recorded at a visit, both the TL and NTL visit response is recorded as NA and the overall visit response is no evidence of disease (NED). If a new lesion is observed then the overall visit response is PD.

Table 3 TL 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 <10mm.
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 ≥ 5mm, 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.
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). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
Not Applicable (NA)	No TLs are recorded at baseline.

Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum are rounded to one d.p. before assigning a TL response. For example 19.95% is rounded to 20.0% but 19.94% is rounded to 19.9%

Missing TL data

For a visit to be evaluable then all TL measurements are recorded. However, a visit response of PD is still assigned if any of the following occurred

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

Note: the nadir can only be taken from assessments where all the TLs had a LD recorded.

If the TL visit response is not recorded as PD, then the TL visit response is NE.

If there is at least one TL measurement missing and a visit response of PD cannot be assigned, the visit response is NE.

If all TL measurements are missing then the TL visit response is NE. Overall visit response is also NE, unless there is a progression of non-TLs or new lesions, in which case the response is PD.

Lymph nodes

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

- If all lesions meet the CR criteria (i.e. 0 mm or < 10mm for lymph nodes) then response is set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node short axis increases by 20% but remains < 10mm.

- If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response is set to NE irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node short axis increases by 20% but remains < 10mm.

- If not all lesions are missing, and those that are non-missing do not meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis ≥ 10mm or the reappearance of previously disappeared lesion), then response is set to PD.

- If all lesions are missing the response is set to NE.

TL too big to measure

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

TL too small to measure

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

Irradiated lesions/lesion intervention

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

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolization but note this does not include protocol specified biopsies), are handled in the following way. Once a lesion has had intervention then it is treated as having intervention for the remainder of the study noting that an intervention most likely shrinks the size of tumours:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) are summed and the calculation is performed in the usual manner. If the visit response is PD, this remains as a valid response category.

- Step 2: If there is no evidence of progression after step 1, the lesion diameter (for those lesions with intervention) will be treated 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 patients will be assigned a TL response of PD.

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

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

Lesions that split into two

If a TL splits in two, then the LDs of the split lesions are 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 are recorded for one of the TL sizes and the other TL size is recorded as 0mm.

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 is 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 are treated as missing.

Non-target lesions (NTLs) and new lesions – site investigator data.

At each visit, the investigator is to 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 is derived based on the investigator's overall assessment of NTLs as follows:

Table 4 NTL Visit Responses

Visit Responses	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.
Not evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For subjects 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.

To achieve 'unequivocal progression' on the basis of 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 are also recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and indicates 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 are identified via a Yes/No tick box. The absence and presence of new lesions at each visit are listed alongside the TL and NTL visit responses.

A new lesion indicates progression so the overall visit response is 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 study therapy and is not included in any assessment of NTLs.

Subjects 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.

Overall visit response – site investigator data

Table 5 defines how the previously defined TL and NTL visit responses are combined with new lesion information to give an overall visit response. Confirmation of progression is not required.

Table 5 Overall visit responses

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
CR	CR or NA	No (or NE)	**CR**

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
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)	**Non-CR/Non-PD**
NA	NE	No (or NE)	**NE**
NA	NA	No (or NE)	**NED**

4 STATISTICAL ANALYSIS

This section provides information on definitions, derivation and analysis/data presentation per domain.

4.1 Analysis Sets

The analysis sets for this study are described in Table 6.

Table 6 Analysis Sets

Analysis set	Description	Endpoint/Output
Enrolled analysis set	All subjects who sign the informed consent form (ICF).	Disposition
Safety analysis set	All subjects receiving at least 1 dose of study treatment. Subjects are summarised according to the actual treatment received.	Baseline and demography Exposure Safety ORR BOR TTR DoR PFS OS Percentage Change in Target Lesion Tumour size

Analysis set	Description	Endpoint/Output
		Best Percentage Change in target lesion tumour size PK concentration listings
Pharmacokinetics (PK) analysis set	All dosed subjects with reportable ceralasertib or durvalumab plasma concentrations. Subjects are summarised according to the treatment received.*	PK concentrations
Pharmacodynamic analysis set	All subjects who receive at least 1 dose of study treatment with at least 1 reportable post-baseline pharmacodynamic measurement. Subjects are summarised according to the treatment received.*	Pharmacodynamic endpoints

ORR: Objective Response Rate, TTR: Time to Response, DoR: Duration of Response, PFS: Progression Free Survival, OS: Overall Survival, BOR: Best Objective Response; TTR and DoR are reported for the subset of subjects with confirmed objective response.

*Individual PK concentration data for any subjects who are excluded from the descriptive summary tables and/or figures are included in the listings and are flagged with an appropriate footnote.

Subject's data are flagged as being unevaluable in case no signed ICF is available. Subjects who are not assigned to treatment due to violation of one or more protocol defined inclusion or exclusion criteria are handled as screening failures.

4.1.1.1 Presentation

The analysis sets are summarised. Any exclusions from analysis sets are listed.

4.2 Study Population

The domain study population covers subject disposition, analysis sets, protocol deviations, demographics, baseline characteristics, medical history, disease characteristics, prior and concomitant medication, and study drug compliance.

4.2.1 Subject Disposition and Completion Status

4.2.1.1 Definitions and Derivations

Subjects screened is defined as informed consent received.

A subject is considered to have completed the study if they have completed all phases of the study including the last visit or the last scheduled procedure. Details can be found in section 4.4 of the CSP.

4.2.1.2 Presentation

Subject disposition including screen failures and reason for screen failure is summarised and listed based on all subjects screened as defined by the current relevant tables, figures,

listings (TFL) standards. The number and percentage of subjects for the following are summarised if applicable:

- Subjects screened;
- Screening failures.
- Subjects assigned to treatment
- Subjects assigned, did not receive study treatment
- Subjects started treatment
- Subjects who received ceralasertib and subjects who did not receive ceralasertib;
- Subjects who received durvalumab and subjects who did not receive durvalumab;
- Subjects ongoing study treatment at data cut-off;
- Subjects who discontinued treatment (including reason),
- Subjects ongoing study at data cut-off;
- Subjects who terminated study.

Summaries on disposition due to global/country situation due to a pandemic are added to the disposition table if applicable. The number and percentage of subjects for the following summaries are added on if applicable:

- Subjects who discontinued treatment due to global/country situation;
- Subjects who withdrew from study due to global/country situation.

The study disruptions due to the global/country situation is also summarised as a separate table.

The number and percentage of subjects with confirmed or suspected Coronavirus Disease 2019 (COVID-19) infection during the course of the study are presented separately, including details on COVID-19 related interruptions impacting on visits, and study drug administration. Discontinuation of study intervention and/or withdrawal from study due to COVID-19 are presented separately.

Listings of subjects affected by the COVID-19 pandemic are presented detailing any affect and impact on the study. Issues reported in the Clinical Trial Management System are considered for presentation in listings as well.

4.2.2 Protocol Deviations

4.2.2.1 Definitions and Derivations

Important protocol deviations (IPDs) are defined as protocol deviations that may significantly affect the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being. IPDs are defined in detail

in the Protocol Deviation Specification Document., but will generally include, but are not limited to:

- Key Inclusion criteria
- Key Exclusion criteria
- Discontinuation criteria for study product met but subject not withdrawn from study treatment
- Investigational Product (IP) Deviation
    - Subjects deviating from the prescribed dosing regimen.
    - Dose modifications (not related due to either an immune or a non-immune-related AE according to Dosing Modification and Toxicity Management Guidelines)
- Excluded medication taken
    - Received prohibited concomitant medications or therapies (including other anticancer agents).
- Deviations related to study procedure
    - Baseline Response Evaluation Criteria in Solid Tumours (RECIST) version 1.1 scan > 28 days before enrolment
    - No baseline RECIST 1.1 assessment on or before date of enrolment.
    - Fresh tumour biopsy not done at visit
- Other important protocol deviations
    - Deviation from Good Clinical Practice (GCP) as determined by medical review

If a deviation is serious enough to have a potential impact on the primary analysis, sensitivity analyses may be performed. A list of IPDs according to the Protocol Deviation Specification Document is compiled and new potential IPDs are classified by the global lead physician and an unblinded biostatistician trailed by AZ study physician and AZ clinical operations review approximately every 4-5 weeks. The Protocol Deviation Specification Document and the final IPD list will be finalised prior to database lock.

4.2.2.2 Presentation

The incidence of important protocol deviations (IPDs) are summarised for the safety analysis set.

- Number of subjects with at least 1 important protocol deviation further split by reason related to;
- Number of subjects with at least 1 pandemic related important protocol deviation further split by reason related to;

- Number of subjects with at least 1 important protocol deviation, excluding pandemic related IPDs further split by reason related to.

A listing is provided with the important protocol deviation details.

4.2.3 Demographics

4.2.3.1 Definitions and Derivations

Age group is defined as a categorical variable with levels <18, ≥ 18,<50, ≥ 50, < 65, ≥ 65, < 75 and ≥ 75 years. Each race category counts subjects who selected only that category.

4.2.3.2 Presentation

Demographics are summarised and listed based on the Safety analysis set as defined by the current relevant TFL standards. The following are summarised: age (years) at baseline, age group, sex, race and ethnicity. Additionally, the number and percentage of subjects recruited in each country and each centre are presented.

4.2.4 Baseline Characteristics

4.2.4.1 Definitions and Derivations

Body Mass Index is calculated as weight/height^2 (kg/m^2).

4.2.4.2 Presentation

Baseline characteristics are listed and summarised according to Section 3.2 for the Safety analysis set as defined by the current relevant TFL standards. The following are summarised: weight (kg), height (cm),body mass index (kg/m^2).

4.2.5 Disease Characteristics

4.2.5.1 Definitions and Derivations

Not applicable.

4.2.5.2 Presentation

Disease characteristics at baseline are listed and summarised in terms of absolute counts and percentages affected of subjects for the Safety analysis set as defined by the current relevant TFL standards.

The following are summarised:

- Extent of disease at study entry (metastatic, locally advanced disease, both metastatic and locally advanced, location)
- Disease characteristics at baseline (Eastern Cooperative Oncology Group [ECOG] performance status, primary tumour location, primary tumour type, number of metastases, histology type, primary tumour stage, regional lymph nodes stage, distant metastases stage, tumour grade, Stage/AJCC stage, resistance to prior IO treatment [IVRS], and baseline LDH level [IVRS])

- Mutation status at study entry (Proto-oncogene c-KIT [KIT], neurofibromatosis type 1 [NF1], ataxia telangiectasia and Rad3-related protein [ATR], ataxia telangiectasia mutated [ATM], AT-rich interactive domain-containing protein 1A [ARID1A], B-Rapidly Accelerated Fibrosarcoma gene [BRAF] V600 Mutation and subtype, NRAS proto-oncogene, GTPase [NRAS] Activating Mutations and microsatellite instability [MSI]-High Genetic or Immunohistochemistry)

Recurrence of earlier cancer is depicted in a listing.

4.2.6 Medical History and Concomitant Disease

4.2.6.1 Definitions and Derivations

Medical and surgical history is coded using Medical Dictionary for Regulatory Activities (MedDRA).

4.2.6.2 Presentation

Medical history and concomitant disease are listed and summarised for the Safety analysis set as defined by the current relevant TFL standards.

A separate summary of medical history for subjects who had confirmed or suspected COVID-19 infection during the study is presented. If less than five events appear and less than 2% of subjects assigned to treatment are affected the summary will not be visualised.

Additionally, the following is summarised in terms of numbers and percentage of subjects who had:

- Previous disease related treatment modalities (cytotoxic chemotherapy, targeted therapy, antiangiogenic therapy, taxane chemotherapy, radiopharmaceuticals, platinum chemotherapy, experimental therapy, transarterial chemoembolization [TACE], transarterial radioembolization [TARE], immunotherapy, other, number of previous chemo regimens, number of prior lines of systemic therapy, reason for therapy failure of previous cancer therapy, and treatment status of previous cancer therapy)

- Previous Anti-PDL1 therapies (Therapy class, treatment status, Reason for Therapy discontinuation, type of resistance of prior Anti-PDL1, name of prior Anti-PDL1 agent)

Summaries for the number and percentage of subjects wo had a certain number of prior regimens and response assessment results are produced.

The number and percentage of subjects who had prior cancer therapies are summarised by ATC classification and generic drug name.

Results from the Hepatitis B and C and HIV tests at screening are depicted in a listing.

4.2.7 Prior and Concomitant Medications

4.2.7.1 Definitions and Derivations

For the purpose of inclusion in prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates are imputed as detailed in Section 3.2.8.

Prior medications, concomitant 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 the first dose of study treatment.
- Concomitant medications are those with a stop date on or after the first dose date of study treatment, and must have started prior to or during treatment so there is at least one day in common with the study treatment.

4.2.7.2 Presentation

The following summaries are provided in terms of absolute counts and percentages affected of subjects for the Safety analysis set as defined by the current relevant TFL standards using ATC classification code and the generic term coded by standard drug dictionary WHODrug Global B3 Mar 2022 or later:

- summary of prior medications or therapies
- summary of concomitant medications or therapies

Additionally, the following is summarised in terms of numbers and percentage of subjects who had:

- Previous Anti-PDL1 therapies (Therapy class, treatment status, Reason for Therapy discontinuation, type of resistance of prior Anti-PDL1, name of prior Anti-PDL1 agent)

Also, all important recorded information for previous Anti-PDL1 therapies will be visualised in a listing.

4.3 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints.

Efficacy analyses are based on investigator assessments. For each post-baseline visit scan, the investigator defines the overall visit response data (complete response [CR], partial response [PR], stable disease [SD], progressive disease [PD], or not evaluable [NE] depending on the status of their disease compared with baseline and previous scans by

assessing TLs, non-target lesions (NTLs) and new lesions) and the relevant scan dates for each time point (i.e., for visits where response or progression is/is not identified). Subjects who have disease progression on study treatment based on progression of non-target disease, may also require submission of additional scans/images/photographs e.g., brain scan or photographs of cutaneous lesions. If a subject has had a tumour assessment that cannot be evaluated, then the subject is assigned a visit response of not evaluable (NE) (unless there is evidence of progression, in which case the response is assigned as PD). The date of progression is provided based on the earliest of the scan dates of the component that triggered the progression. Endpoints ORR, DoR, BOR, TTR, and PFS are then derived from the scan dates and overall visit responses. Confirmation of progression is not required.

Efficacy analysis, except for OS, are based on programmatic application of RECIST 1.1. (Eisenhauer et al, 2009) to investigator assessed tumour measurements. All RECIST 1.1 assessments, whether scheduled or unscheduled, are included in the calculations. This is also regardless of whether a subject discontinues study treatment or receives another anti cancer therapy. At each visit, subjects are programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, PD, or NE depending on the status of their disease compared with baseline and previous assessments.

Baseline should be assessed within the 21 days prior to first dose. The tumour response endpoints (ORR, DoR, BOR, TTR, PFS, and change in TL tumour size) are then derived from the scan dates and overall visit responses. Programmatic derivation guidance used for the application of RECIST 1.1 are provided in Section3.2.10, which is to determine disease response.

RECIST data, overall visit response and best objective response are listed.

4.3.1 Primary Endpoint: CD8+ T cells tumour infiltration

The primary efficacy endpoint is CD8+ T cells tumour infiltration. This data will be collected for four different tests, CD8+ T Cells density (Per mm2, unit) and area (Percent, unit) in the center tumor region; CD8+ T Cells density (Per mm2, unit) and area (Percent, unit) in the invasive margin region;).

4.3.1.1 Primary Analysis of Primary Endpoint

All study patients in the biopsy sub-study are required to provide 3 mandatory fresh tumour biopsy samples for testing at Screening (Days -21 to -1) [baseline assessment], on-ceralasertib treatment, Cycle 0 (on Day 7) of ceralasertib monotherapy [on-treatment assessment], and off-ceralasertib treatment, Cycle 0 (between Days 15 to 28) [off-treatment assessment]. All analysis of this data will be performed on the PD Analysis Set.

CD8+ T cells tumour infiltration are summarised descriptively in baseline, ceralasertib on-treatment and ceralasertib off-treatment samples for each of the tests described in Section 4.3.1

The summary statistics provided include number of observations (n), geometric mean (calculated as exp[μ], where μ is the mean of the data on a logarithmic scale), coefficient of variation (CV, calculated as 100 √ [exp(s2)-1], where s is the standard deviation of the data on a log scale), arithmetic mean, arithmetic SD, median, minimum and maximum.

Boxplots of change from baseline will be produced on a log scale for on-treatment and off-treatment timepoints. The geometric means are presented using a diamond symbol in these plots. These will be produced for each of the tests described in Section 4.3.1.

Plots will be produced showing each patient's baseline, on-treatment and off-treatment results over time on a log scale with separate lines indicating the results for each patient. These will be produced for each of the tests described in Section 4.3.1.

Data are transformed as appropriate e.g., log2-transformation, and analysed by MMRM with fixed effect for timepoint. An unstructured variance-covariance matrix is used to accommodate repeated measures within a subject, assuming model convergence otherwise alternative structures including autoregressive (TYPE=AR(1)), compound symmetry (TYPE=CS), and Toeplitz (TYPE=TOEP) are applied and selected based on best fit by examining best fit statistics AICC.

Association of biomarkers to clinical outcomes is presented visually in terms of a figure of individual data and geometric mean of CD8+ T cells tumour infiltration counts at baseline, on-treatment and off-treatment. Additionally, a figure of individual data and geometric mean of CD8+ T cells tumour infiltration fold changes is presented on a log scale. In both figures, different line/symbol colours are used to present the subjects' response status in BOR. These will be produced for each of the tests described in Section 4.3.1.

4.3.2 Secondary Endpoint: PD-L1

4.3.2.1 Analysis of Secondary Endpoint

PD-L1 expression will be evaluated applied to sections stained using Ventana SP263 immunohistochemistry assay. Other exploratory cut-offs may also be assessed as required. PD-L1 data will be collected using the SP263 assay (PD-L1 expression will be reported using the Tumor Area Positivity score and Tumor Positive Membrane score). Summaries of the number and percentage of patients with PD-L1 expression <1% and >= 1% will be provided at each sample timepoint and each test. No formal statistical analysis will be performed. Unknown: sample where PD-L1 expression was not available either due to a test fail (unevaluable sample or assay failure) or sample slide out of cut-slide stability.

4.3.3 Secondary Endpoint: pRAD50

4.3.3.1 Analysis of Secondary Endpoint

Pre-treatment presence and/or on-treatment and/or off-treatment changes in pRAD50 (by means of percent positive cells and H-scores, separately) are summarised in the same way as the primary endpoint. No formal statistical analysis will be performed.

4.3.4 Secondary Endpoint: Proliferation (using Ki67+ marker) of carcinoma and/or immune cells (including CD8+ T cells)

4.3.4.1 Analysis of Secondary Endpoint

For secondary efficacy endpoint analysis, Ki67+ marker data. This data will be collected for four different tests, Ki67+ cell density (Per mm2, unit) and area (Percent, unit) in the center tumor region; Ki67+ Cells density (Per mm2, unit) and area (Percent, unit) in the invasive margin region. Pre-treatment presence and/or on-treatment and/or off-treatment changes in proliferation are summarised in the same way as the primary endpoint for each test. No formal statistical analysis will be performed.

4.3.5 Secondary Endpoint: Objective Response Rate

The secondary efficacy endpoint is ORR by investigator.

4.3.5.1 Definition

ORR is defined as the proportion of subjects who have a complete response (CR), or partial response (PR), as determined by investigator per RECIST 1.1, prior to any evidence of progression (as defined by RECIST 1.1), that is confirmed at least 4 weeks later. ORR based on investigator is defined as the percentage of subjects with at least two post-baseline investigator-assessed visit response of CR or PR, with the denominator defined as the number of subjects in the Safety analysis set.

4.3.5.2 Derivations

A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit. Data obtained up until progression, or last evaluable assessment in the absence of progression, are included in the assessment of ORR. Also, only data obtained before the start of subsequent anticancer treatment (excluding radiotherapy) are included. Therefore, both visits contributing to a confirmed response must be prior to progression and prior to subsequent anticancer treatment.

In the case where a subject has two non-consecutive visit responses of PR, then, as long as the time between the 2 visits of PR is greater than 4 weeks and there is no PD between the PR visits, the subject is defined as a responder. Similarly, if a subject has visit responses of

CR, NE, CR, then, as long as the time between the 2 visits of CR is greater than 4 weeks, then a best response of CR is assigned.

4.3.5.3 Analysis of Secondary Endpoint

Summaries are produced that present the number and percentage of subjects with a confirmed tumour response (CR/PR). The ORR is presented with a two-sided 95% CI using the Clopper-Pearson (exact probability) method. Subjects that have missing overall response assessments at all visits are considered as non-responders and are therefore counted in the denominator of ORR. The main analysis of ORR is based on the Safety analysis set.

4.3.6 Best Objective Response

4.3.6.1 Definition

Best objective response (BoR) will be calculated based on the overall visit responses from each RECIST assessment, described in section 3.2.10. It is the best response a patient has had following first dose, 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.

4.3.6.2 Derivations

BoR is based on RECIST and will be determined using investigator data until the earliest of the first progression event (including death)/last evaluable assessment in the absence of RECIST progression or start of any subsequent cancer therapy. The denominators are consistent with those used in the ORR analysis.

BoR will be based on RECIST using the following response categories and in the following order:

- Response with categories: CR (confirmed), PR (confirmed)
- Non-response with categories:
    - SD with categories unconfirmed CR or PR, SD (>=16 weeks) and Non-CR/Non-PD (>=16 weeks)
    - NE with categories SD (<16 weeks), Non-CR/Non-PD (<16 weeks) and no baseline RECIST or post baseline RECIST assessments in the absence of death.
    - Progression with categories: RECIST progressions and death in the absence of RECIST progression

For determination of a best response of SD and Non-CR/Non-PD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD/ Non-CR/Non-PD should be recorded at least 16 weeks from Cycle 1 Day 1 minus 1 week, i.e. at least 15

weeks from Cycle 1 Day 1 (to allow for an early assessment within the assessment window), after first dose. For CR/PR, the initial overall visit assessment that showed a response should use the latest of the dates contributing towards a particular overall visit assessment.

4.3.6.3 Analysis of best objective response

The number of patients with a response and non-response along with each of the response/non-response categories will be summarised.

4.3.7 Secondary Endpoint: Duration of Response

DoR is a secondary efficacy endpoint.

4.3.7.1 Definition

DoR is defined as the time from the date of first documented confirmed objective response until the date of documented progression per RECIST 1.1 as assessed by investigator or death due to any cause.

4.3.7.2 Derivations

DoR (months) = (date of PFS event [progression/death] or censoring – date of first documented confirmed objective response + 1) / (365.25/12)

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 is defined as the latest of the dates contributing towards the first visit response of PR or CR. If a subject does not progress following a response, then their DoR is censored on the PFS censoring date. Only subjects who have achieved objective response (CR or PR) are evaluated for DoR.

4.3.7.3 Handling of Dropouts and Missing Data

DoR is calculated for the subset of subjects with confirmed objective response (CR or PR) only. Subjects without confirmed objective response are excluded from the analysis of DoR. Dropouts and missing data will be handled according to the censoring rules detailed in Section 4.3.11.2,

4.3.7.4 Primary Analysis of Secondary Endpoint

Only subjects who have achieved confirmed objective response (CR or PR) are included in the summaries of DoR. The median DoR as well as landmark estimates at 6, 9, 12, 15, 18 months of DoR and two-sided 95% CI are estimated using the Kaplan-Meier method. The analysis of DoR is based on the safety analysis set.

Swimmer plots that clearly show the profile of each subject who responds are produced.

4.3.8 Secondary Endpoint: Time to Response

Time to Response (TTR) is a secondary efficacy endpoint. The analysis of TTR is based on the safety analysis set.

4.3.8.1 Definition

TTR is defined as the time from first dose until the date of first documented objective response, which is subsequently confirmed per RECIST 1.1 as assessed by investigator.

4.3.8.2 Derivations

TTR (months) = (date of first confirmed objective response – date of first dose + 1) / *(*365.25/12*)*.

4.3.8.3 Handling of Dropouts and Missing Data

Subjects without confirmed objective response, are treated as missing for the calculation of TTR. Only subjects who have achieved confirmed objective response (CR or PR) are evaluated for TTR.

4.3.8.4 Primary Analysis of Secondary Endpoint

The TTR is summarised (i.e., number of subjects [%] based on the number of responders) by the scheduled assessment timepoint that the response was first observed. The number of subjects with response at different disease assessment timepoints is provided. Additionally, descriptive summary statistics are presented.

The median TTR and two-sided 95% CI is assessed using the Kaplan-Meier method (without censoring because all subjects have events).

4.3.8.5 Supplementary Analyses of the Secondary Endpoint

Not applicable

4.3.9 Secondary Endpoint: Percentage Change in target lesion tumour size

Percentage Change in target lesion tumour size is a secondary efficacy endpoint.

4.3.9.1 Definition

An outcome endpoint for this study is percentage change from baseline in TL tumour size. This is based on RECIST TL measurements taken at baseline and each post-baseline assessment. Tumour size is the sum of the longest diameters (or short axis measurements for lymph nodes) of the TLs. Baseline for RECIST is defined to be the last evaluable assessment prior to first dose.

The TL tumour size and percentage change from baseline in the sum of TL tumour size at each assessment are calculated.

4.3.9.2 Derivations

Whenever TL tumour size data for the week 20 visit (note: or visit at which progression was documented if before week 20) is available then this is used in the analysis. A windowing rule is applied and follows the protocol allowed visit window, therefore RECIST scan performed within ±1 week of the protocol scheduled visit is used for that visit. Subjects who progress before week 20 (plus 1 week allowing for a late assessment within the visit window) should have had a tumour assessment performed at the time of progression prior to study treatment discontinuation. The TL tumour size from their latest progression assessment is used instead of the week 20 assessment for these subjects.

The percentage change from baseline in TL tumour size is obtained for each subject by taking the difference between the sum of the TLs at week 20 (or earlier if the patient progressed) and the sum of the target lesions at baseline divided by the sum of the TLs at baseline times 100 (i.e. [week 20 – baseline] / baseline × 100).

4.3.9.3 Primary Analysis of Secondary Endpoint

The TL tumour size and percentage change in TL tumour size from baseline is summarised using descriptive statistics and presented at each timepoint for the Safety set. Descriptive statistics for the percentage change from baseline at week 20 are presented, together with the number of subjects with a baseline and at least 1 post-baseline assessment.

The percentage change from baseline in TL tumour size is presented graphically using waterfall plots, with the bars ordered from the largest increase to the largest decrease. Separate waterfall plots are used to present each subject's week 20 percentage change in TL tumour size as a separate bar. A reference line at the –30% change in TL tumour size level is added to the plots, which corresponds with the definition of 'partial response'. All progressions are marked with a '●' or designated with patterns or colours for ORR categories. Flagged progressions on the percentage change in TL tumour size at week 20 are based upon NTL or new lesion progression at that timepoint and flagged progressions on the best percentage change are based upon NTL or new lesion progression at the same timepoint as the best percentage change. The scale in these plots is 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 are marked with '#'. Values are ordered in descending order with the imputations due to death appearing first followed by a gap followed by all other subjects. These plots will be produced for the Safety Set.

Additionally, 'spider' plots of percentage change from baseline in target lesion size by subject are presented. This depicts each subject's percentage change from baseline in TL tumour size as a line over time and progression due to non- target and/or new lesions are indicated. These plots will be produced for the Safety Set.

4.3.9.4 Supplementary Analyses of the Secondary Endpoint

Not applicable.

4.3.10 Secondary Endpoint: Best Percentage Change in target lesion tumour size

Best Percentage Change in target lesion tumour size is a secondary efficacy endpoint.

4.3.10.1 Definition

The TL tumour size and percentage change from baseline in the sum of TL tumour size at each assessment are calculated. The best absolute change in TL tumour size from baseline (i.e. depth of response) is the largest decrease from baseline or the smallest increase from baseline in the absence of reduction and includes all assessments:

- up to and including the first visit at which the overall visit response is PD,
- prior to death in the absence of progression,
- prior to the start of subsequent anti-cancer therapy (excluding radiotherapy);
- or up to and including the last evaluable RECIST assessment if the subject has not died, progressed or started subsequent anti-cancer therapy.

Each post-baseline disease assessment for a subject that meets the following conditions are included: all target lesions identified at baseline have measurements recorded at the current visit (i.e. they cannot be not done or not evaluable). If a lesion is recorded as ≤ 5 mm then this will be used in the calculations.

4.3.10.2 Derivations

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 the best percentage change from baseline in the following situations (otherwise best percentage change is 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

4.3.10.3 Handling of Dropouts and Missing Data

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

- If a subject has no post-baseline assessment and has died

- If a subject has new lesions or progression of non-target lesions (NTLs) or TLs
- If a subject has withdrawn due to PD and has no evaluable TL data before or at PD

4.3.10.4 Primary Analysis of Secondary Endpoint

Descriptive statistics for the best percentage change from baseline are presented, together with the number of subjects with a baseline and at least 1 post-baseline assessment.

The best percentage change from baseline in TL tumour size is presented graphically using waterfall plots, with the bars ordered from the largest increase to the largest decrease. Separate waterfall plots are used to present each subject's best percentage change in TL tumour size as a separate bar. A reference line at the –30% change in TL tumour size level is added to the plots, which corresponds with the definition of 'partial response'. All progressions are marked with a '●' or designated with patterns or colours for ORR categories. Flagged progressions on the best percentage change in TL tumour size at week 20 are based upon NTL or new lesion progression at that timepoint and flagged progressions on the best percentage change are based upon NTL or new lesion progression at the same timepoint as the best percentage change. The scale in these plots is 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 are marked with '#'. Values are ordered in descending order with the imputations due to death appearing first followed by a gap followed by all other subjects. These plots will be produced for the Safety set.

4.3.10.5 Supplementary Analyses of the Secondary Endpoint

Not applicable.

4.3.11 Secondary Endpoint: Progression Free Survival

PFS is a secondary efficacy endpoint.

4.3.11.1 Definition

PFS is defined as the time from first dose of Cycle 0 until the date of first documented disease progression as assessed by investigator or death (by any cause in the absence of disease progression), regardless of whether the subject withdraws from study therapy or receives another anti-cancer therapy prior to progression.

4.3.11.2 Derivations

PFS (months) = (date of PFS event (progression/death) or censoring – date of first dose + 1) / (365.25/12)

Subjects who have not progressed or died at the time of analysis are censored at the time of the latest date of assessment from their last evaluable disease assessment. However, if the

subject progresses or dies immediately after two or more consecutive missed visits, the subject is censored at the time of the latest evaluable disease assessment prior to the two missed visits. (Note: a not evaluable visit is not considered as a missed visit).

For the first time period starting at cycle 0, two or more consecutive missing visits result in ≥ 13 weeks (4 weeks + 8 weeks+ 1 week for late assessment = 13 weeks). For the time period where the 8-weekly schedule applies, two or more consecutive missing visits resulting in ≥ 18 weeks after the last evaluable post-baseline disease assessment, 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). The first interval in the 8-weekly schedule cycle 1 day 1 allows for 3 days early assessment despite 7 days for the interval that follow resulting in 122 days compared to 126 days for the later intervals.

If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from 8-weekly to 12-weekly this equates to 22 weeks (i.e. the average of 8 and 12 weeks which gives 10 weeks and then apply the same rationale, hence 2 x 10 weeks + 1 week for early assessment + 1 week for late assessment = 22 weeks).

When the scheduling changes to 12-weekly assessments, two missing visits equate to 26 weeks (2 x 12 weeks + 1 week for early assessment + 1 week for late assessment = 26 weeks).

In summary the two or more consecutive missed visits result in the following:

1. If previous assessment is baseline: 4 weeks + 8 weeks+ 1 week (last assessment) ≥ 13 weeks (91 days),
2. If cycle 1 day 1 scan was the previous evaluable visit: 8 weeks + 8 weeks + 3 days (visit window for cycle 1 day 1) + 1 week (3 days for early and 1 week for late assessment) ≥ 122 days,
3. If 8 week schedule applies but cycle 1 day 1 was not previous visit: 8 weeks + 8 weeks + 2 weeks (1 week for early and 1 week for late assessment) ≥ 18 weeks
4. If schedule changes from 8 weeks to 12 weeks: 8 weeks + 12 weeks +2 weeks ≥ 22 weeks
5. During 12-weekly assessments: 12 weeks + 12 weeks + 2 weeks ≥ 26 weeks.

If subjects have no evaluable disease assessments post-baseline or do not have baseline tumour assessment data they will be censored at Day 1 unless they die within two visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window) when the death date qualifies as a PFS event.

A summary of censoring rules and the date of PD/death or censoring are given in Table 8. Note that censoring overrides event in certain specified cases.

Table 8 Summary of Censoring Rules for PFS

Situation	Date of PD/Death or Censoring	PFS Outcome
Documented PD or death in the absence of progression	Date of earliest documentation of PD or date of death in the absence of progression	Event
Either no tumour assessment at baseline or no evaluable assessments post-baseline AND death prior to second scheduled post-baseline disease assessment	Date of death	Event
Either no tumour assessment at baseline or no evaluable assessments post-baseline AND no death prior to second scheduled post-baseline disease assessment	Date of first dose	Censored
PD or death (in the absence of progression) immediately after ≥ 2 consecutive missed disease assessments as per the protocol specified assessment schedule	Last evaluable progression-free disease assessment prior to missed assessments	Censored
On-going with neither PD nor death at the time of analysis or lost to follow-up or withdrawn consent	Date of last evaluable disease assessment	Censored

PD = progressive disease; PFS = progression-free survival

The PFS time is always derived based on scan/assessment dates not visit dates.

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

- For investigator assessments, the date of progression is determined based on the earliest RECIST assessment/scan dates of the component that indicates progression.

- When censoring a subject for PFS the subject is censored at the latest of the dates contributing to a particular overall visit assessment.

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

The proportion of subjects alive and progression-free at 3, 6, 9, 12 and 18 months from first dose is defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1) at these specific months.

Potential duration of follow-up for PFS, applicable only for PFS censored subjects, is defined as follows:

Potential duration of follow-up for PFS in censored subjects (months) = (date of PFS censoring – date of first dose + 1) / (365.25/12).

4.3.11.3 Handling of Dropouts and Missing Data

See Section 4.3.11.2.

4.3.11.4 Primary Analysis of Secondary Endpoint

The analysis of PFS is based on the Full analysis set. The number and percentage of subjects experiencing a PFS event (broken down by type of event/censoring) are provided along with the median PFS and its two-sided 95% CI estimated using the Kaplan-Meier method. Kaplan-Meier plots of PFS are presented. Moreover, the landmark estimates of PFS at 3, 6, 9, and 12 months from first dose are summarised.

The treatment status at progression of subjects at the time of analysis are summarised. This includes the number (%) of subjects who were on treatment at the time of progression, the number (%) of subjects who discontinued study treatment prior to progression, the number (%) of subjects who have not progressed and were on treatment or discontinued treatment.

A summary of 'potential' duration of follow-up for PFS is included using median (range). This is presented for censored subjects (including all types of PFS censoring).

4.3.12 Secondary Endpoint: Overall Survival

OS is a secondary efficacy endpoint.

4.3.12.1 Definition

OS is defined as the time from the date of first dose of Cycle 0 until death due to any cause regardless of whether the subject withdraws from study therapy or receives another anti-cancer therapy.

4.3.12.2 Derivations

OS (months) = (date of death or censoring – date of first dose+ 1) / (365.25/12)

Any subject not known to have died at the time of analysis is censored based on the last recorded date on which the subject was known to be alive.

Note: Survival calls are made in the week following the date of DCO for the analysis, and if subjects are confirmed to be alive or if the death date is post the DCO date these subjects are censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" subjects at the time of any formal OS analysis should be obtained by the

site personnel by checking the subject's notes, hospital records, contacting the subject's general practitioner and checking publicly-available death registries. In the event that the subject has actively withdrawn consent to the processing of their personal data, the vital status of the subject can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note: For any other OS analysis performed, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the subject was known to be alive for those subjects still on treatment (since the *SURVIVE* module is only completed for subjects off treatment if a survival sweep is not performed). The last date for each individual subject is defined as the latest among the following dates recorded on the CRFs:

- AE start, stop and change in severity dates
- Admission and discharge dates of hospitalisation
- Study treatment date
- End of treatment date
- Concomitant medication start and stop dates
- Laboratory test dates
- Date of vital signs
- Date of ECGs
- Physical examination dates
- Cardiovascular data
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anticancer treatment
- Date last known alive on survival status CRF
- End of study date (excluding patients lost to follow-up)
- Patient Reported Outcomes (PRO) questionnaire dates

The proportion of subjects alive at 6, 9, 12, and 18 months is defined as the Kaplan-Meier estimate of OS at that particular month.

"Potential" duration of follow-up for OS, applicable only for OS censored subjects, is defined as follows:
Potential duration of follow-up for OS (months) = (date of OS censoring [date last known to be alive] – date of first dose + 1) / (365.25/12).

4.3.12.3 Handling of Dropouts and Missing Data

If a subject 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:

a. For Missing day only – using the 1st of the month.

b. For Missing day and Month – using the 1st of January.

If death has been recorded but the date is entirely missing, then date of death will be imputed as the date the patient was last known to be alive +1..

4.3.12.4 Primary Analysis of Secondary Endpoint

The analysis of OS is based on the Full analysis set. The number and percentage of subjects experiencing an OS event and Kaplan-Meier plots of OS are presented. Additionally, summaries of the numbers and percentages of subjects who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent are provided. The median OS and two-sided *95*% CI are estimated using the Kaplan-Meier method.

Moreover, the landmark estimates of OS at 6, 9, 12, and 18 months from first dose are summarised with corresponding 95% CI.

A summary of the potential duration of follow-up for OS is included using median (range). This is presented for censored subjects.

4.3.12.5 Sensitivity Analyses of the Secondary Endpoint

A sensitivity analysis is conducted to assess the potential impact of COVID-19 related deaths on OS. That is, subjects who had a death event where the primary or secondary cause was COVID-19 infection or COVID-19 infection was reported as a fatal AE, are censored at the date of their COVID-19 infection related death. If less than five events appear and less than 2% of enrolled subjects are affected the sensitivity analysis will not be carried out.

4.3.12.6 Supplementary Analyses of the Secondary Endpoint

Not applicable

4.3.13 Secondary Endpoint: PK of ceralasertib

PK of ceralasertib alone and when in combination with durvalumab is a secondary endpoint.

See section 4.5

4.3.14 Safety Analysis

The domain safety covers exposure, adverse events, clinical laboratory, vital signs, and electrocardiogram (ECG).

Tables are provided for the safety set, listings are provided for all subjects or the safety set depending on the availability of data.

4.3.14.1 Exposure

4.3.14.1.1 Definitions and Derivations

For the below derivations exposure will be calculated from Cycle 1 onwards i.e. Cycle 0 will not be counted.

Duration of exposure

- Duration of exposure to ceralasertib = min(date of last dose where dose > 0 mg +◦21, date of death, date of DCO) – first dose* date +◦1
- Duration of exposure to durvalumab = min(date of last dose where dose > 0 mg +◦20, date of death, date of DCO) – first dose* date +◦1

*First dose refers to earliest dose of either treatment

Actual duration of exposure (weeks)

- Actual exposure to ceralasertib [months] = total exposure to ceralasertib – total duration of dose interruptions and dose delays
- Actual exposure to durvalumab [months] = total exposure to durvalumab – total duration of dose delays

The total duration of dose delays is defined as any length of time in addition to the planned cycle length where the patient has not taken/received any of the planned doses.

- Definition of dose delay for ceralasertib: if the number of days between Day 1 of a cycle and Day 1 of the previous cycle is > 28 days, then the number of additional days above 28 days is the duration of delay.
- Definition of dose delay for durvalumab: for Cycle 1, if the number of days between the first administration of durvalumab on Day 8 and the first administration of ceralasertib on Day 1 is > 7 days, then the number of additional days above 7 days is the duration of the delay. For all subsequent cycles, if the number of days between Day◦8 of a cycle and Day 8 of the previous cycle is > 28 days, then the number of additional days above 28 days is the duration of delay.

The total duration of dose interruption is defined as any length of time during the cycle where the patient has not taken/received any of the planned doses.

- Definition of dose interruptions for ceralasertib: the total number of full-day treatment interruptions, including missed and forgotten doses. Half-day interruptions are not considered in the actual treatment exposure derivation

Note: In case the last days of cycle doses of ceralasertib are not taken (eg, Day 7) and no further medication will be taken, this is an interruption but will not be included in the total duration of dose interruptions as the total treatment exposure already includes this reduction via "date of last dose where dose > 0 mg".

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

Missed and forgotten doses should be recorded as a dose interruption with the reason recorded as "Patient forgot to take dose". These missed or forgotten doses are not included as dose interruptions in the summary tables, but the information appears in the listing for dosing. However, these missed and forgotten doses are considered in the derivation of actual exposure.

Dose intensity of study treatment(s) is addressed by considering relative dose intensity (RDI), where RDI is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. More specifically, RDI is defined as follows:

- RDI = 100% × d/D, where d is the actual cumulative dose delivered up to min(date of treatment discontinuation, date of death, data of DCO) and D is the intended cumulative dose up to the day of treatment discontinuation. D is the total dose that would be delivered if there were no modification to dose or schedule.

Intended cumulative dose is calculated by summing the individual doses that should have been received up to and including the last day of study participation according to the planned dose and schedule.

This is given by the intended dose multiplied with the number of administrations that were planned to occur. These are different per treatment component.

- Intended cumulative dose of ceralasertib [mg] = daily dose × number of dosing days =◦daily dose × [7 dosing days per number of complete cycles + dosing days of incomplete cycle]:
  - If mod28(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib + 1) ≤ 7 then:
    - Intended cumulative dose of ceralasertib =◦240◦mg × 2 × [7 × floor((min(date of treatment discontinuation, date of DCO) – date of first dose of ceralaserib +◦1)/28) + mod_{28}(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib + 1)]
  - Else if mod_{28}(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib + 1) > 7 then:

- Intended cumulative dose of ceralasertib = 240 mg × 2 × [7 × floor(min(date of treatment discontinuation, date of DCO) – date of first dose of ceralasertib +◦1)/28) + 7]

- Intended cumulative dose of durvalumab [mg] = dose × number of complete cycles =◦1500◦mg × ceil(min(date of treatment discontinuation, date of DCO) – date of first dose of durvalumab + 1)/28)

Note: ceil() refers to rounding up and floor() refers to rounding down to the next integer. $mod_{28}(n)$ is the number n modulo 28.

Actual cumulative dose is calculated by summing all administered doses:

- ceralasertib dose = Actual dose recorded x actual frequency x duration
- durvalumab dose = Actual dose recorded

4.3.14.1.2 Presentation

Duration of exposure to Investigational Products (IP)(s) in weeks and cycles are summarised by descriptive statistics and by frequency. Dose intensity is summarised by descriptive statistics. Exposure to IP(s) i.e. total amount of study drug received is listed for all subjects. Exposure swimmer plot(s) are produced, with a line presented for each subject to display relevant exposure and disposition details.

Dosing deviations for IP(s) are summarised with reasons for deviations for the following categories: delays, reductions, and interruptions. Dosing delays are derived based on the scheduled dosing dates using the previous dose given as reference. The number and percentage of subjects with dosing delays and total dose delays per subject are summarised.

4.3.14.2 Adverse Events

4.3.14.2.1 Definitions and Derivations

The Medical Dictionary for Regulatory Activities (MedDRA) (version 25 or higher) is used to code the AEs. AEs are graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE) (version 5.0).

Treatment emergent adverse events (TEAEs) are all AEs which onset or worsen in severity following the first administration of IP in Cycle 0 within the duration of the treatment period, up to and including 30 days after the last dose of study treatment for ceralasertib monotherapy or 90 days after last dose of study treatment for ceralasertib and durvalumab combination. Worsening in severity is determined by comparison with the pre-treatment CTCAE grade of the AE recorded closest to the start of dosing.

AEs with a missing start time or where time is not collected, which occur on the same day as first IP administration, are reported as treatment emergent.

For rules on missing or partial dates, see Section 3.4.6.

SAEs

A serious adverse event (SAE) is any AE that:

- Results in death
- Is immediately life-threatening
- Requires inpatient hospitalisation or prolongation of existing hospitalisation
- Results in persistent or significant disability/incapacity.
- Is a congenital anomaly/birth defect in offspring of the subject
- Is an important medical event that may jeopardize the subject.

Adverse Events of Special Interest (AESI)

Adverse events of special interests are events of scientific and medical interest specific to the further understanding of ceralasertib and durvalumab safety profile and require close monitoring and rapid communication by the investigators to AstraZeneca. An AESI can be serious or non-serious. All AESIs will be recorded in the eCRF. Additionally, AESI will be identified programmatically using a definition file maintained by Astra Zeneca.

Adverse events of special interests for durvalumab include but are not limited to events with a potential inflammatory or immune-mediated mechanism and which may require more frequent monitoring and/or interventions such as steroids, immunosuppressants and/or hormone replacement therapy. An immune-mediated AE is defined as an AESI that is associated with drug exposure and is consistent with an immune-mediated mechanism of action and where there is no clear alternate aetiology.

If the investigator has any questions in regards to an event being an immune-mediated AE, the investigator should promptly contact the AstraZeneca study physician or clinical lead.

Adverse events of special interests/ immune-mediated AE observed with anti-PD-L/PD-1 agents such as durvalumab include pneumonitis, hepatitis, diarrhoea/colitis, intestinal perforation, endocrinopathies (hypo- and hyperthyroidism, adrenal insufficiency, hypophysitis/hypopituitarism and Type 1 diabetes mellitus), nephritis, rash/dermatitis, myocarditis, myositis/polymyositis, pancreatitis and rare/less frequent events (including, but not limited to haematological events, neuromuscular toxicities [such as myasthenia gravis and Guillain-Barré syndrome], non-infectious encephalitis, non-infectious meningitis, pericarditis, rheumatological events, sarcoidosis, skin events, uveitis [and other events involving the eye] and vasculitis).

In addition, infusion-related reactions and hypersensitivity/anaphylactic reactions with a different underlying pharmacological aetiology are also considered AESIs.

More detailed guidelines for their evaluation and treatment are described in detail in the Dose Modification and Toxicity Management Guidelines (version 28-Oct-2021). These guidelines have been prepared by the sponsor to assist the investigator in the exercise of his/her clinical judgement in treating these types of toxicities. These guidelines apply to AEs considered causally related to the study treatment/study regimen by the reporting investigator.

There are no currently identified AESI for Ceralasertib but if any are identified, these will be reported in similar/appropriate outputs to the durvalumab AESI.

Other significant adverse events (OAE)

During the evaluation of the AE data as part of the monthly offline listing review process, an AstraZeneca medically qualified expert reviews the list of AEs that were not reported as SAEs and AEs leading to discontinuation of study treatment after Parexel's review took place. Based on the expert's judgement, adverse events of particular clinical importance may, after consultation with the Global Safety Physician, be considered OAEs and reported as such in the CSR. A similar review of laboratory values, vital signs, ECGs and other safety assessments are performed for identification of other significant adverse events. This review takes place prior to database lock, and any AEs identified are fully documented in meeting minutes. Further review following database lock may result in ad-hoc OAEs being identified, in this case, the OAEs and resulting summaries are fully documented in the CSR.

Examples of these are 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.

4.3.14.2.2 Presentation

All TEAEs will be summarised and listed. AEs which are not treatment emergent will be listed for the Safety analysis set and will be listed individually by subject.

TEAEs will be counted once for each subject for calculating percentages of subjects experiencing TEAE. In addition, if the same TEAE occurs multiple times within a particular subject, the highest severity and level of relationship observed will be reported. For tables by MedDRA system organ class (SOC) and MedDRA preferred term (PT), subjects with multiple TEAEs will be counted once for each SOC/PT.

An overall summary table of the number of subjects experiencing each category of adverse event will be produced. The number of subjects experiencing treatment emergent adverse events by MedDRA SOC and PT will be summarised and incidence rates will be provided.

Severity and relationship to IP will be summarised. Further splits by CTCAE grade, causal relationship to IP and adverse events with Grade 3-4 will be also summarised.

Separate tables will present adverse events leading to discontinuation, serious adverse events, IP-related adverse events, and other significant adverse events.

Details of any deaths will be summarised and listed for all subjects. AEs leading to death will also summarised.

SAEs

SAEs are summarised as described above for the TEAEs.

AEs of special interest

Grouped summary tables of certain MedDRA preferred terms are produced and may also show the individual preferred terms which constitute each AESI grouping. These groupings may be defined using MedDRA terms, SMQs (standardized MedDRA queries) and PTs (preferred terms). Groupings are provided by the coding team prior to DBL, and a listing of the preferred terms in each grouping is provided. Summaries of the above-mentioned grouped AE categories include number and percentage (%) of subjects who have:

- At least one AESI presented by event outcome
- At least one AESI causally related to IP
- At least one AESI leading to discontinuation of study treatment.

A summary of total duration (days) of AESI are provided for events which have an end date and this may be supported by summaries of ongoing AESIs at death and, separately, at data cut-off.

4.3.14.3 Clinical Laboratory, Blood Samples

4.3.14.3.1 Definitions and Derivations

Laboratory tests are grouped according to chemistry and haematology. Laboratory parameters are assessed at baseline as well as throughout the study.

For chemistry and haematology parameters, laboratory abnormalities with toxicity grades according to the NCI CTCAE version 5.0 are derived.

Change from baseline in haematology and clinical chemistry endpoints is calculated for each post-dose visit. CTC grades are calculated at each visit. Maximum post-baseline CTC are also calculated. Absolute values are compared to the local laboratory reference range and classified as low (below range), normal (within range or on limits of range) and high (above range). All values classified as high or low are flagged on the listings.

Liver Function Parameters

Subjects with elevated post-baseline alanine aminotransferase (ALT), aspartate aminotransferase (AST) or Total Bilirubin that fall into these categories are identified.

Table 7 Liver Function Parameters

Liver Function Parameters	Category
ALT	• ≥3 × – ≤5 × ULN
AST	• ≥3 × – ≤5 × ULN
Total bilirubin	• >3 × – ≤5 × ULN
ALT or AST	• ≥3 × – ≤5 × ULN
Potential Hy's law	• (AST ≥ 3 × ULN or ALT ≥ 3 × ULN) and (Total Bilirubin ≥ 2×ULN)[a]

ULN: upper limit of normal range.

[a] It includes all subjects who have ALT or AST≥3xULN and total bilirubin (BILI) ≥2xULN, and in which the elevation in transaminases precede or coincide with (that is, on the same day as) the elevation in BILI.

4.3.14.3.2 Presentation

The change in each laboratory parameter from baseline to each post-baseline visit are summarised graphically.

Laboratory abnormalities occurring from the start of IP administration to the last assessment on study are presented. Worst toxicity grade, rates of grade 3-4 toxicity, and grade shifts of 2 or more from baseline to the maximum grade are presented. Summaries indicating hyper- and hypo- directionality of change are produced, where appropriate. Laboratory parameters that cannot be graded via NCI CTCAE are summarised with frequencies of post-baseline laboratory values categorized as low (L), normal (N), or high (H) using laboratory normal ranges comparing to baseline.

Listings are provided for all laboratory results.

Liver Function Parameters

Number and percentage of subjects with elevated post-baseline ALT, AST or Total Bilirubin (BILI) are tabulated. Individual subject data with elevated ALT or AST plus BILI falling into the "Potential Hy's law" are summarised and listed.

4.3.14.4 Clinical Laboratory, Urinalysis

4.3.14.4.1 Definitions and Derivations

Change from baseline in urinalysis endpoints are calculated for each post-dose visit. CTC grades are calculated at each visit. Maximum post-baseline CTC is also calculated. Absolute values are compared to the local laboratory reference range and classified as low (below range), normal (within range or on limits of range) and high (above range). All values classified as high or low are flagged in the listings.

4.3.14.4.2 Presentation

Listings are provided for urinalysis. Urinalysis endpoints are summarised by study visit which will include descriptive statistics for the value of the parameters and the changes from baseline. Absolute values and change from baseline at each timepoint are presented. Maximum post-baseline CTC grades which are not normal will be summarized.

For urinalysis, shift from baseline to worst on treatment results are presented.

4.3.14.5 Vital Signs

4.3.14.5.1 Definitions and Derivations

Vital signs (weight, body temperature, systolic and diastolic blood pressure, respiratory rate and heart [pulse] rate) are performed at timepoints as specified in the schedule of assessments. Any changes in vital signs should be recorded as an AE if applicable.

4.3.14.5.2 Presentation

Vital signs are summarised by study visit which includes descriptive statistics for the value of the parameters and the changes from baseline. Absolute values and change from baseline for vital signs data at each timepoint are presented using box plots.

4.3.14.6 Electrocardiogram

4.3.14.6.1 Definitions and Derivations

Electrocardiogram (ECG) parameters are assessed in triplicate at baseline and subsequent triplicate measurements and should only be taken if clinically indicated. Single ECG readings are collected at visits other than baseline. ECG parameters include: PR, RR, QRS, QT, and QTcF. The QTcF is considered as the primary correction method to assess subject cardiac safety.

From these resting 12-lead ECGs values of the QT and RR intervals and the QT interval corrected for heart rate using Fridericia's correction (QTcF) is derived using the following formula:

- QTcF = QT/RR^ (1/3) where RR is in seconds

The values of QTcF (milliseconds) are re-derived from the values of RR and QT during the creation of the reporting database.

4.3.14.6.2 Presentation

ECG parameters are listed.

4.3.14.7 Echocardiogram/Multigated Acquisition

4.3.14.7.1 Definitions and Derivations

An echocardiogram or Multigated Acquisition (MUGA) scan to assess left ventricular ejection fraction (LVEF) will be performed at the visits screening and treatment discontinuation and as clinically indicated.

4.3.14.7.2 Presentation

LVEF will be listed.

4.3.14.8 Bleeding events

Information obtained by the bleeding event questionnaires is depicted in a listing.

4.4 Exploratory Analyses

4.4.1 CCI

CCI

CCI

CCI

4.5 Pharmacokinetics

Plasma concentrations-time data are not analysed to determine the PK parameters. Pre-dose and post-dose plasma concentrations for each scheduled time-point are summarised, PK Day and Cycle using appropriate descriptive statistics. PK concentration data are listed for each subject in the safety analysis set.

Individual concentrations may be excluded from summary tables, graphical or statistical analysis for legitimate scientific reasons. Any exclusions, together with justification for the exclusion, is clearly documented in the CSR. These individual data are still presented in the listings, but are flagged to identify that they are excluded from summary outputs.

The following descriptive statistics are presented for plasma concentrations:

- n

- n below lower limit of quantification (LLOQ)
- geometric mean (gmean)
- geometric standard deviation (gSD)
- gmean±gSD
- geometric coefficient of variance (%) (gCV)
- arithmetic mean of non log-transformed data (mean)
- standard deviation of non log-transformed data (Std Dev)
- coefficient of variance (%) (CV)
- median
- minimum (min)
- maximum (max)

The gmean is calculated as exp(μ), where μ is the mean of the data on the natural log scale.

The gSD is calculated as exp(σ), where σ is the standard deviation of the data on the natural log scale.

The gCV is calculated as 100 x sqrt[exp(s^2)-1], where s is the Std Dev of the data on the natural log scale.

The gmean ± gSD (gmean-gSD and gmean+gSD) is calculated as exp[μ ± s].

Protocol scheduled times are used to present the PK concentration summary tables and corresponding gmean concentration-time figures.

Box-plots of plasma concentration at pre-dose and 1h post dose is presented for cycles until cycle 3 for both Ceralasertib and Durvalumab separately. The geometric means are presented using a diamond symbol in these plots.

Handling of Non-Quantifiable Concentrations

Individual concentrations below the LLOQ of the bioanalytical assay are reported as NQ in the listings with the LLOQ defined in the footnotes of the relevant TFLs. Individual plasma concentrations that are Not Reportable are reported as NR and those that are missing are reported as NS (No Sample) in the listings. Plasma concentrations that are NQ, NR or NS are handled as follows for the provision of descriptive statistics:

- Any values reported as NR or NS are excluded from the summary tables and corresponding figures.

- At a time point where less than or equal to 50% of the concentration values are NQ, all NQ values are set to the LLOQ, and all descriptive statistics are calculated accordingly.
- At a time point where more than 50% (but not all) of the values are NQ, the gmean, gmean ± gSD and gCV% are set to NC. The maximum value is reported from the individual data, and the minimum and median are set to NQ.
- If all concentrations are NQ at a time point, no descriptive statistics are calculated for that time point. The gmean, minimum, median and maximum are reported as NQ and the gCV% and gmean ± gSD as NC.
- The number of values below LLOQ (n < LLOQ) are reported for each time point together with the total number of collected values (n).
- If data are available for less than three subjects, no summary statistics other than minimum, maximum and n are presented.

Graphical presentation of PK data

- Box plots of pre-dose and post-dose concentration mean (arithmetic mean and/or gmean) by PK visit are displayed on the same plot. Preferably, cycles are differentiated by colours on the same figure.

Precision and Rounding Rules for Pharmacokinetic Data /PK concentration data

PK concentration data listings are presented to the same number of significant figures as the data received from the bioanalytical laboratory (usually but not always to 3 significant figures) and against the same units as received.

PK concentration descriptive statistics present 4 significant figures with the exception of the min and max which present 3 significant figures and n and n<LLOQ which present as integers.

4.6 Immunogenicity

Immunogenicity results are listed for each subject and summarised for the Safety analysis set. Number and percentage of subjects in the following categories are provided.

- Anti-drug antibody (ADA) positive at baseline and/or post-baseline visits.
- Persistent positive, defined as positive at ≥2 post-baseline assessments (with ≥16 weeks between first and last positive) or positive at last post-baseline assessment.

- Transient positive, defined as negative at last post-baseline assessment and positive at only one post-baseline assessment or at ≥2 post-baseline assessments (with <16 weeks between first and last positive).
- Treatment-boosted, defined as baseline positive ADA titre that was boosted to a 4-fold or higher level following drug administration.

All valid assay results from subjects who receive any study drug are included in immunogenicity summaries. Blood samples collected outside of the protocolled window are summarised at the closest nominal time point that does not already have a value. All subjects with titre information are shown in the data listing.

5 REFERENCES

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika 1934; 26:404–416.

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

Frewer P, Mitchell P, Watkins C, Matcham J. Decision-making in early clinical drug development. Pharma Statist. 2016; 15:255-263.

Guo W, Wang SJ, Yang S, Lynn H, Ji Y. A Bayesian interval dose-finding design addressing Ockham's razor: mTPI-2. Contemporary Clinical Trials 2017; 58:23-33.

Hatcher RA, Trussell J, Nelson AL, Cates W Jr, Stewart F, Kowal D, eds. Contraceptive technology. 19th edition. New York: Ardent Media 2007(a): 24. Table 3-2.

Jennison C, Turnbull BW. "Group Sequential Methods with Applications to Clinical Trials," Chapman & Hall/CRC, Boca Raton 2000.

Ji Y, Liu P, Li Y, Bekele BN. A modified toxicity probability interval method for dose-finding trials. Clinical Trials 2010; 7:653-63.

Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clinical Trials 2008; 5:93-106.

Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al on behalf of the RECIST working group. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol. 2017; 18: e143–52.

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: d533ac00001-sap-ed-1		
Document Title:	Statistical Analysis Plan Edition 1	
Document ID:	Doc ID-005361799	
Version Label:	1.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
19-Mar-2024 11:29 UTC	PPD	Author Approval
19-Mar-2024 12:35 UTC	PPD	Content Approval
19-Mar-2024 10:51 UTC	PPD	Content Approval

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