Statistical Analysis Plan	
Study Code	D933YC00001
Edition Number	3.0
Date	18Mar2024

A Phase III, Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab as Consolidation Therapy in Patients with Locally Advanced, Unresectable, Non-Small Cell Lung Cancer (Stage III) Who Have Not Progressed Following Definitive, Platinum-Based, Chemoradiation Therapy (PACIFIC 5)

A Phase III, Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab as Consolidation Therapy in Patients with Locally Advanced, Unresectable, Non-Small Cell Lung Cancer (Stage III) Who Have Not Progressed Following Definitive, Platinum-Based, Chemoradiation Therapy (PACIFIC 5)

Study Statistician (IQVIA)

PPD

Date

A Phase III, Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab as Consolidation Therapy in Patients with Locally Advanced, Unresectable, Non-Small Cell Lung Cancer (Stage III) Who Have Not Progressed Following Definitive, Platinum-Based, Chemoradiation Therapy (PACIFIC 5)

Study Statistician (AstraZeneca)

PPD

Date

A Phase III, Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab as Consolidation Therapy in Patients with Locally Advanced, Unresectable, Non-Small Cell Lung Cancer (Stage III) Who Have Not Progressed Following Definitive, Platinum-Based, Chemoradiation Therapy (PACIFIC 5)

Global Product Statistician

PPD

Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
AEPI	Adverse event of possible interest
ALK	Anaplastic lymphoma kinase
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BICR	Blinded Independent Central Review
BoR	Best objective response
cCRT	Concurrent chemoradiation therapy
CI	Confidence interval
CrCL	Creatinine clearance
COVID-19	Coronavirus disease 2019
CR	Complete response
CRF	Case report form
CRT	Chemoradiation therapy
CSP	Clinical study protocol
CSR	Clinical study report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Event
DAE	Discontinuation of investigational product due to adverse events
DBL	Database lock
DBP	Diastolic blood pressure
DCO	Data cut-off
DoR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EGFR	Epidermal growth factor receptor
EORTC	European Organisation for Research and Treatment of Cancer
EQ-5D-5L	EuroQoL 5 dimension, 5-level health state utility index

Abbreviation or special term	Explanation
FSI	First subject in
GCP	Good Clinical Practice
HR	Hazard ratio
HRQoL	Health-related quality of life
IA	Interim analysis
irRECIST	Immune-related RECIST
IDMC	Independent Data Monitoring Committee
IHC	Immunohistochemistry
imAE	Immune-mediated AE
IP	Investigational product
ITT	Intent-to-treat
IV	Intravenous
IVRS	Interactive voice response system
KM	Kaplan-Meier
LD	longest diameter
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified intent-to-treat
MRI	Magnetic resonance imaging
NA	Not applicable
NE	Not evaluable
NED	No evidence of disease
NMPA	National Medical Products Administration
NSCLC	Non-small cell lung cancer
NTL(s)	Non-target lesion(s)
OAE	Other significant adverse events
ORR	Objective response rate
OS	Overall survival
OS24	Proportion of patients alive at 24 months from randomisation
PD	Progressive disease
PD-L1	Programmed cell death ligand 1
PFS	Progression-free survival
PFS12	Proportion of patients alive and progression-free at 12 months after randomisation
PFS18	Proportion of patients alive and progression-free at 12 months after randomisation

Abbreviation or special term	Explanation
PFS2	Time from randomisation to second progression
PK	Pharmacokinetic(s)
PR	Partial response
PRO	Patient-reported outcome
PT	Preferred term
q4w	Every 4 weeks
QLQ-C30	30-item core quality of life questionnaire
QLQ-LC13	13-item lung cancer quality of life questionnaire
QoL	Quality of life
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumours
SAE	Serious adverse event
SAP	Statistical analysis plan
SBP	Systolic blood pressure
sCRT	Sequential chemoradiation therapy
SD	Stable disease
SI	Standard International
TCs	Tumour cells
TL(s)	Target lesion(s)
CCI	[redacted]
TTDM	Time to death or distant metastasis
TTR	Time to response
ULN	Upper limit of normal
WHO	World Health Organisation

AMENDMENT HISTORY

Date	Brief description of change
(v1.0) 09Sep2020	N/A – first version
(v2.0) 30Jun2022	Updated to reflect CSP v4 and CSP v5, as described below: Section 1.1: Study objectives changes. Section 1.3: Sample size details updated. Section 2.1: • Full analysis set replaced with intent-to-treat set. • Modified intent-to-treat set introduced, including as analysis set for primary and key secondary efficacy analyses, for both main and China cohorts. • Analysis populations updated in Table 1 accordingly. Section 2.2: • Numbering of deviations revisited. • Inclusion criterion no.2 added for Part I and Part II. • COVID-19 related issues and important protocol deviations added. Section 3.2.1: • Condition for censoring at day 1 updated. • Definition of 2 missed visits updated Section 3.2: • Definition of subsequent anti-cancer therapies added. • Reference to analysis sets removed. • Description of summaries/analyses removed. Section 3.2.2.7: • Preconditions to have PFS2 added.

Date	Brief description of change
	• Details on subsequent anti-cancer therapies considered in PFS2 added. Section 3.2.3.1: • List of individual items updated (perceived financial impact of disease added). • Reference to Fayers et al updated. Section 3.3: • General definition of "on-treatment" assessments added. • Description of summaries/analyses removed. Section 3.3.1: • Section on Other significant AEs removed. • Section on confirmed/suspected COVID-19 infection added. Section 3.2.1: Table 6 (definition of 2 missed visits) added. Section 3.3.3: Definition of safety variable added (moved from section 4.2.5.4). Section 3.5: • Sub-section on "PFS and ORR using BICR assessments according to immune related RECIST (irRECIST)" removed • Sub-section "Collection of data due to COVID-19" added Section 4.1: • Analysis populations updated. • Repeated analysis on ITT identified. Section 4.1.1: details on baseline definition for efficacy and PRO endpoints added Section 4.1.2: • PFS IA removed.

Date	Brief description of change

	• Updated to consider mITT for PFS (primary endpoint) and OS (key secondary endpoint). Moved secondary endpoints PFS (ITT) and OS (ITT) to the 3rd and 4th layers of the MTP (figure 2 update)
	• Updated statistical calculations for PFS FA and OS analyses.
	
	Section 4.2: Table 9 updated.
	
	Section 4.2.1:
	
	• Reference to the Kaplan-Method method removed for the purpose of p-value generation
	• Sensitivity analyses
	  ○ Details on ascertainment bias analysis added
	  ○ Attrition bias sensitivity analysis added
	  ○ Analysis accounting for COVID-19 deaths added
	• Subgroup analysis: subgroups added
	
	Section 4.2.2.1: Sensitivity analysis accounting for COVID-19 deaths added
	
	Section 4.2.2.3: details on SAS® procedures used updated.
	
	Section 4.2.3: details on SAS® procedures used updated.
	
	Section 4.2.5.1:
	
	• Details on first subsequent anti-cancer therapy updated.
	• List of tables for AESI/AEPI description detailed.
	• COVID-19 related death events summarised
	
	Section 4.2.5.4: Removed definition of safety variables.
	
	Section 4.2.6.2: Analysis set updated.
	
	Section 4.2.11:
	
	• Analysis set updated
	• Disease characteristics at baseline update to include ALK status and the combine EGFR/ALK status

Date	Brief description of change
	• Details on EGRF, ALK and combined EGFR/ALK status added • Selection of summaries repeated on patients with confirmed/suspected COVID-19 infection Section 4.2.13: new section Section 4.2.14: new section Section 5 • PFS interim analysis removed • Statistical calculation updated for OS interim analysis updated Section 6: Time to deterioration in patient-reported symptoms, functioning and global health status/QoL added as secondary endpoint Section 7: References updated. Typos corrected.
(v3.0) 18Mar2024	Updated to reflect CSP v6, as described below: Section 1.12.1: update table footnote. Section 1.22.1: remove the IDMC review for efficacy interim analysis. Section 1.32.1: update the PFS and OS DCO criteria and related wording to align with CSP v6. Section 2.12.1: add flexible wording of other country/region safety analysis set for other country/region subset analysis. Section 2.22.1: add deviation g to the IPD listing. Section 3.1.12.1: add two rules of overall visit response derivation for missing TL data. Section 3.2.2.62.1: For the definition/derivation of TTDM, new lesion proven by biopsy will not be used for distant metastasis identification. Section 3.2.2.2.17 2.1: update PFS2 derivation to make it clear.

Date	Brief description of change
	Section 3.2.32.1: add 2 missing visits rule for PRO endpoints. Section 3.2.42.1: add the rule of missing TTSCAPRX form for TFST and TSST derivation. Section 4.1.22.1: update OS DCO criteria and add MTP rule according to the DCO adjustment. Section 4.2.5.12.1: • remove episode level summary. • add AEs summary with maximum CTCAE grade 3 or 4. • remove infection adverse events summary. • remove adverse events of pneumonitis and radiation pneumonitis summary. Section 5.12.1: remove the IDMC review for OS interim analysis. Section 5.1.12.1: update the OS IA DCO criteria. Section 5.1.22.1: remove the IDMC review for OS interim analysis. Section 62.1: remove biopsy test for TTDM definition.

1 STUDY DETAILS

1.1 Study objectives

1.1.1 Primary objective

Primary objective:	Endpoint/variables:
To assess the efficacy of durvalumab treatment compared with placebo in terms of PFS in randomised patients without sensitizing EGFR mutations or ALK rearrangements (i.e. mITT)	PFS using BICR assessments according to RECIST 1.1[a]

BICR Blinded Independent Central Review; PFS Progression-free survival; RECIST Response Evaluation Criteria In Solid Tumors.
[a]PFS will be based on programmatically derived PFS by BICR assessment according to RECIST 1.1. See primary efficacy outcome endpoints Section 3.2.1 and analysis methods Section 4.2 for further details.

1.1.2 Secondary objectives

Secondary objectives:	Endpoint/variables:
Key secondary objective: To further assess the efficacy of durvalumab compared with placebo in terms of OS in the mITT	OS
To further assess the efficacy of durvalumab treatment compared with placebo in terms of PFS in all randomised patients (i.e. ITT)	PFS using BICR assessments according to RECIST 1.1[a]
To further assess the efficacy of durvalumab compared with placebo in terms of OS in the ITT	OS
To further assess the efficacy of durvalumab compared with placebo in terms of: OS24, ORR, DoR, PFS12, PFS18, PFS2 and TTDM in the mITT, and separately in the ITT	OS24 ORR, DoR, PFS12, PFS18 and TTDM using BICR assessments according to RECIST 1.1[b] PFS2 as defined by local standard clinical practice
To assess the PK of durvalumab	Concentration of durvalumab in blood and non compartmental PK parameters (such as peak

	concentration and trough, as data allow) (sparse sampling)
To investigate the immunogenicity of durvalumab	ADA (confirmatory results: positive or negative; titres [ADA neutralising antibodies will also be assessed])
To assess symptoms and health-related quality of life in patients treated with durvalumab compared with placebo using EORTC QLQ-C30 v3 and QLQ-LC13 in the mITT, and separately in the ITT	Change from baseline in patient-reported symptoms, functioning and global health status/QoL. Time to deterioration in patient-reported symptoms, functioning and global health status/QoL.
To investigate the relationship between a patient's baseline tumour PD-L1 expression and efficacy outcomes with durvalumab compared with placebo in the mITT, and separately in the ITT	IHC analysis of tumoural PD-L1 expression and spatial distribution within the tumour microenvironment relative to efficacy outcomes (OS, PFS, and ORR)

ADA Anti-drug antibody; BICR Blinded Independent Central Review; DoR Duration of response; EORTC QLQ-C30 European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire; IHC Immunohistochemical; ORR Objective response rate; OS Overall survival; OS24 Proportion of participants alive at 24 months from randomisation; PD L1 Programmed death ligand 1; PFS Progression free survival; PFS2 Time from randomisation to second progression; PFS12 Proportion of participants alive and progression free at 12 months from randomisation; PFS18 Proportion of participants alive and progression free at 18 months from randomisation; PK Pharmacokinetic(s); QLQ-LC13 13-item self-administered questionnaire for lung cancer; QoL Quality of Life; RECIST Response Evaluation Criteria In Solid Tumors; TTDM Time to death or distant metastasis;

[a] PFS will be based on programmatically derived PFS by BICR assessment according to RECIST 1.1.

[b]Analysis of ORR, DoR, PFS12, PFS18 and TTDM will be based on BICR assessment according to RECIST 1.1.

See secondary efficacy outcome endpoints Section 3.2.2 and analysis methods Section 4.2 for further details.

1.1.3 Safety objective

Safety objective:	**Endpoint/variables:**
To assess the safety and tolerability profile of durvalumab compared with placebo	AEs, physical examinations, vital signs, electrocardiograms, and laboratory findings

AE Adverse event.

1.1.4 Exploratory objective

Exploratory objective:	**Endpoint/variables:**

To investigate the relationship between a patient's CCI and efficacy outcomes with durvalumab	CCI relative to efficacy outcomes (OS, PFS, and ORR)
To explore irRECIST criteria as an assessment methodology for clinical benefit of durvalumab compared with placebo by BICR	PFS and ORR using BICR assessments according to irRECIST[a]
To investigate the relationship between durvalumab PK exposure and clinical outcomes, efficacy, AEs and/or safety parameters, if deemed appropriate	A graphical and/or a data modelling approach will be used to analyse durvalumab PK exposure and the relationship with clinical outcomes, efficacy, AEs and/or safety parameters, as deemed appropriate
To describe and evaluate resource use associated with durvalumab treatment and underlying disease	Health resource utilisation measures including hospitalisation, outpatient visits, or emergency department visits
To explore the impact of treatment and disease state on health state utility using EQ-5D-5L	The EQ-5D-5L health state profile will be used to derive health state utility index representing utility based on patient reported data
CCI	
To explore the relationship(s) between a patient's biomarker status and durvalumab PK exposure and clinical outcomes before and after treatment	Biomarker status before and after treatment and durvalumab PK exposure and relationship with clinical outcomes, efficacy, AEs, and/or safety parameters, as deemed appropriate
To explore potential biomarkers in residual biological samples CCI , which may influence the progression of cancer (and associated clinical characteristics) and/or prospectively identify patients likely to respond to durvalumab treatment	Correlation of biomarkers with response to durvalumab treatment and/or the progression of cancer

AE Adverse event; BICR Blinded Independent Central Review; EQ-5D-5L EuroQoL 5-dimension; irRECIST Immune-related Response Evaluation Criteria In Solid Tumors; ORR Objective response rate; OS

Overall survival; PD-L1 Programmed death ligand 1; PFS Progression-free survival; PK Pharmacokinetic(s); CCI .

[a] Exploratory endpoints and analyses related to PFS/ORR analyses using irRECIST data objectives maybe reported outside of the main CSR.

1.2 Study design

This study is a Phase III, randomised, double-blind, placebo-controlled, multicentre study assessing the efficacy and safety of durvalumab compared with placebo as consolidation therapy in patients with locally advanced, unresectable non-small cell lung cancer (NSCLC, [Stage III]) who have not progressed following definitive, platinum-based, chemoradiation therapy (CRT). See Figure 1 for an overview of the study design. See D933YC00001 Clinical Study Protocol version 6.0 dated 28thAug2023 (CSP) Tables 1 and 2 for the schedule of activities planned, including details on the study period.

The study will maintain a balance of approximately 1:1 between the cCRT and sCRT patient populations with the majority of patients to be recruited in China.

Treatments and treatment duration:

After the Screening Period (up to 84 days), patients will receive durvalumab 1500 milligrams (mg) or placebo every 4 weeks (q4w) through the Treatment Period from Day 1 until confirmed radiological progression, initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or other reasons to discontinue study treatment occur (note: if a patient's weight falls to 30 kilograms (kg) or below then the patient should receive weight-based dosing equivalent to 20 mg/kg durvalumab q4w until the weight improves to above 30 kg, at which point the patient should receive the original 1500 mg durvalumab dosage q4w again). Up to 3 days window from randomisation to first dose will be applied.

Once a patient has had objective progression recorded and has discontinued study treatment, the patient will enter the Follow-up Period, from last dose until 12 months and then every 2 months thereafter.

Independent Data Monitoring Committee:

An independent data monitoring committee (IDMC) comprising independent experts will meet with frequency as per the IDMC charter from first subject in (FSI) the study to review safety assessments and make recommendations to continue, amend, or stop the study based on safety findings. Additional reviews of the safety data may be requested by the IDMC at additional points during the study.

Full details of the IDMC procedures and processes can be found in the IDMC Charter.

CCI

1.3 Number of subjects

The study will randomise approximately 400 patients in the ITT population and approximately 375 patients in the mITT population. Patients who received prior concurrent chemoradiation (cCRT) or sequential chemoradiation (sCRT) and have not progressed following definitive chemoradiation therapy will be randomised in a 2:1 ratio to receive durvalumab and placebo.

Randomisation will be stratified by the level of PD-L1 expression (PD-L1 <1% or PD-L1≥1%) and prior therapy (cCRT or sCRT).

The number of patients with sensitizing epidermal growth factor receptor (EGFR) mutation or anaplastic lymphoma kinase (ALK) rearrangement randomised will be capped at approximately 7% of the total number of patients randomised. In order to reflect global clinical practice, study will recruit at least 60% of patients who received prior cCRT. The majority of patients will be recruited in China.

Eligible patients will be in complete response (CR), partial response (PR), or have stable disease (SD) following definitive, platinum-based, cCRT or sCRT.

The primary objective of this study is to assess the efficacy of durvalumab compared with placebo in terms of PFS per RECIST 1.1 as assessed by BICR in the mITT population. The key secondary endpoint (i.e., included in the multiple testing procedure) is OS (mITT).

The primary PFS (mITT) analysis data cut-off (DCO) will occur when whichever of the following conditions have been met first:

- Reaching approximately CCI BICR progression-free survival events across the durvalumab and placebo treatment arms (approximately CCI maturity) in the mITT population

OR

- Approximately CCI months follow-up from last patient randomized to the study.

If the true PFS HR is CCI with an estimated CCI BICR PFS events in the mITT, the study will provide at least CCI power to demonstrate a statistically significant difference for PFS with CCI this translates to approximately a CCI-month benefit in median PFS over CCI months on placebo. The smallest treatment difference that would be statistically significant is a HR of CCI. A recruitment period of approximately CCI months and a follow-up period of CCI months are expected for the PFS (mITT) final analysis.

The overall alpha level for the statistical testing of the secondary key endpoint OS (CCI) for durvalumab versus placebo will be CCI. The final planned OS analysis DCO will occur when reaching approximately CCI death events (CCI maturity) or approximately CCI months follow-up from the last participant randomization in the CCI, whichever occurs first. If the true OS hazard ratio (HR) is CCI with an estimated CCI OS events, this study will provide CCI power to demonstrate a statistically significant difference for OS, assuming a CCI 2-sided significance level (with overall 2-sided alpha for OS as CCI); this translates to an CCI-month benefit in median OS over CCI months on placebo. The smallest treatment difference that would be statistically significant is a HR of CCI. Up to two interim analyses for OS will be conducted: 1) at the same time as primary PFS analysis and 2) at approximately CCI months after the OS first interim analysis, with approximately CCI and CCI of the target events respectively. If the expected DCO for OS final analysis is within CCI months after the OS first interim analysis, the OS second interim analysis may be removed. A recruitment period of approximately CCI months and a follow-up period of CCI months are expected for the final analysis of the OS endpoint.

2 ANALYSIS SETS

2.1 Definition of analysis sets

Summaries of the analysis sets for each outcome variable are provided in Table 1.

Table 1 Summary of outcome variables and analysis populations

Outcome variable	Populations
Efficacy data	
PFS	mITT and ITT[a]
OS, OS24, PFS12, PFS18, PFS2, PRO endpoints, TTDM, ORR* (with/without confirmed response) and DoR* (with/without confirmed response)	mITT and ITT[a]
Study population/demography	
Demography	mITT and ITT[a]
Prior anti-cancer therapy	mITT and ITT[a]
Protocol deviations	mITT and ITT[a]
Medical and surgical history	mITT and ITT[a]
Patient characteristics	mITT and ITT[a]
Concomitant medications	mITT and ITT[a]
PK/ADA data	
PK data	PK analysis set
ADA	ADA analysis set
Safety data	
Exposure	Safety analysis set
Adverse events	Safety analysis set
Laboratory measurements	Safety analysis set
Vital Signs	Safety analysis set

[a] ITT is the secondary population for efficacy analysis.
DoR Duration of response; ITT Intent-to-treat; mITT Modified Intent-to-treat; ORR Objective response rate; OS overall survival; OS24 Proportion of patients alive at 24 months from randomisation; PFS Progression-free survival; PFS2 Time from randomisation to second progression; PFS12 Proportion of patients alive and progression-free at 12 months from randomisation; PFS18 Proportion of patients alive and progression-free at 18 months from randomisation; PK Pharmacokinetic; PRO Patient-reported outcomes; TTDM Time to death or distant metastasis.
* Patients who are evaluable for the analysis of ORR are those with measurable disease at baseline. Patients who are evaluable for the analysis of DoR are those who responded in the ORR analysis.

2.1.1 Intent-to-treat (ITT) set

The ITT set will include all randomised patients. The ITT set will be used as secondary population for efficacy analyses (including PROs). Treatment arms will be compared based on randomised study treatment, regardless of the treatment actually received. Patients who were

randomised but did not subsequently go on to receive study treatment will be included in the analysis in the treatment arm to which they were randomised.

ITT set with measurable disease at baseline

A subset of the ITT set including only those patients with measurable disease at baseline will be applied to all ORR and some BoR endpoint summaries.

China ITT set

A subset of the ITT set including only those patients from sites located in mainland China.

The other country/region ITT set may be planned since the analysis based on country/region ITT set may be performed per the request from local health authority.

2.1.2 Modified intent-to-treat set

The mITT set will include all randomised patients in the ITT set who are without sensitizing EGFR mutations or ALK rearrangements. Unless otherwise specified, the mITT set will be used as primary population for all efficacy analyses (including PROs). Treatment arms will be compared based on randomised study treatment, regardless of the treatment actually received.

Information on EGFR/ALK status will be based on eCRF data. If a patient has EGFR/ALK valid results from both local and central laboratories, then results from central laboratories will be used.

China mITT set

A subset of the mITT set including only those patients without sensitizing EGFR mutations or ALK rearrangements from sites located in mainland China.

The other country/region mITT set may be planned since the analysis based on country/region mITT set may be performed per the request from local health authority.

2.1.3 Safety analysis set

The safety analysis set will consist of all patients who received at least 1 dose (full or partial) of study treatment (durvalumab or placebo). Safety data will not be formally analysed but summarised using the safety analysis set according to the treatment received; that is, erroneously treated patients (e.g., those randomised to treatment A but actually given treatment B) will be summarised according to the treatment they actually received. If a patient receives any amount of durvalumab, they will be summarised in the durvalumab arm.

China safety analysis set

A subset of the safety analysis set including all patients recruited from sites located in mainland China who received at least 1 dose (full or partial) of study treatment (durvalumab or placebo).

The other country/region safety analysis set may be planned since the analysis based on country/region safety analysis set may be performed per the request from local health authority.

2.1.4 PK analysis set

All patients who receive at least 1 dose of durvalumab per the CSP for whom any post-dose data are available and who do not violate or deviate from the CSP in ways that would significantly affect the PK analyses will be included in the PK analysis set. The population will be defined by the Study Physician, Pharmacokineticist, and Statistician prior to any analyses being performed.

China PK analysis set

A subset of the PK analysis set including all patients recruited from sites located in mainland China who receive at least 1 dose of durvalumab per the CSP for whom any post-dose data are available and who do not violate or deviate from the CSP in ways that would significantly affect the PK analyses.

2.1.5 ADA analysis set

All patients who have non-missing baseline ADA and at least 1 non-missing post-baseline ADA result will be included in the ADA analysis set.

China ADA analysis set

A subset of the ADA analysis set including all patients recruited from sites located in mainland China who have non-missing baseline ADA and at least 1 non-missing post-baseline ADA result.

2.2 Violations and deviations

The following general categories will be considered important deviations and will be listed and discussed in the clinical study report (CSR) as appropriate for the study:

- Deviation a: Patients randomised but who did not receive durvalumab/matching placebo

- Deviation b: Patients who deviate from the following key entry criteria in CSP:
  a) Part I inclusion criteria: 2, 6; Part II inclusion criteria: 2, 4, 5, 6.
  b) Exclusion criteria: 7, 11.
- Deviation c: Baseline RECIST 1.1 scan > 42 days before randomisation.
- Deviation d: No baseline RECIST 1.1 assessment on or before date of randomisation or baseline RECIST assessment is performed before the completion of Chemoradiation therapy.
- Deviation e: Patients randomised who received treatment other than that to which they were randomised.
- Deviation f: Received prohibited concomitant medication.
- Deviation g: discontinuation criteria for study product met but patients not withdrawn from study treatment.

Prior to database lock (DBL), medications will be reviewed by a physician on a regular basis, and a decision made by the study team, in order to classify medications that are 'prohibited' from permitted use during the study.

The categorisation of these as important deviations is not automatic and will depend on duration and the perceived effect on efficacy.

Patients who receive the wrong treatment at any time will be included in the safety analysis set as described in Section 2.1. During the study, decisions on how to handle errors in treatment dispensing (with regards to continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader and/or statistician.

The important protocol deviations will be listed and summarised by randomised treatment group. Deviation 'a' will lead to exclusion from the safety analysis set. None of the other deviations will lead to patients being excluded from the analysis sets described in Section 2.1 (except for the PK analysis set, if the deviation is considered to significantly impact upon PK, and except for the ADA analysis set). A per protocol analysis excluding patients with specific important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis may be performed excluding patients with deviations that may affect the efficacy of the trial therapy if > 10% of patients in either treatment group:

- Did not have the intended disease or indication or
- Did not receive any randomised therapy.

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

Protocol deviations, along with other study deviations, will be noted by the monitors (on the electronic case report from [eCRF] for inclusion/exclusion criteria) throughout the study period and these will be reviewed by the study team to determine whether they are to be classified as important deviations and to be summarised/listed in the CSR. In addition to the programmatic determination of deviations above, other study deviations captured from the electronic case report form (eCRF) module for inclusion/exclusion criteria will be tabulated and listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

In addition, the number and percentage of patients with at least one COVID-19 pandemic related important protocol deviations will be summarized. A listing for patients with reported issues due to COVID-19 pandemic will be provided, as reported by the sites.

3 PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST 1.1 visit responses

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

Part II screening/baseline radiological tumour assessments are to be performed no more than 28 days before the start of randomisation and ideally as close as possible to and prior to the start of study treatment. Tumour assessments are then performed after randomisation every 8 weeks ± 1 week for the first 48 weeks and then every 12 weeks ± 1 week thereafter until RECIST 1.1-defined radiological progression. A follow-up scan is to be collected after the initial RECIST 1.1-defined radiological PD, no less than 4 weeks after the prior assessment of PD and no later than the next regularly scheduled imaging visit and this follow-up scan is evaluated using the Confirmation of Radiological Progression criteria outlined in CSP Appendix E. If the subsequent scan confirms the immediate prior radiological PD, no additional scans are required; however, if the subsequent scan does not confirm the immediate prior radiological PD, scanning should continue until the next RECIST 1.1-defined radiological PD, which in turn will require a subsequent scan evaluated using the Confirmation of Radiological Progression criteria outlined in CSP Appendix E.

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

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

RECIST 1.1 outcomes (i.e., PFS and ORR etc.) will be calculated programmatically for the BICR and site investigator data from overall visit responses.

The above derivations will be programmed for the BICR based upon RECIST 1.1 assessments. Measurements from the reviewer selected by the adjudicator will be used when adjudication for overall visit response has occurred, but in the case where no adjudication was required, the measurements from the reviewer who reviewed the baseline scan first will be used for this analysis. Note: this window rule will be applicable to other secondary efficacy outcome endpoints that involve RECIST 1.1 scans.

3.1.1 Site investigator assessments using RECIST 1.1

3.1.1.1 Target lesions

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

Details of TL visit response categories are provided in Table 2. See CSP Appendix E for further details on how to handle special cases of TLs.

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

Note: For patients who do not have measurable disease at entry (i.e. no TLs) but have non-measurable disease, evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions (see Section 3.1.1.2 for further details). If a patient does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

Rounding of TL data

For calculation of progressive disease (PD) and PR for TLs percentage changes from baseline and previous minimum should be rounded to 1 decimal place (d.p.) before assigning a TL response. For example, 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%

Table 2 TL Visit Responses

Visit Responses	Description
Complete response (CR)	Disappearance of all TLs since baseline. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10 mm.
Partial response (PR)	At least a 30% decrease in the sum of the diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.
Progression of disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥5 mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Not evaluable (NE)	Only relevant if any of the TLs at follow-up were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions. Note: If the sum of diameters meets the PD criteria, progressive disease overrides NE as a TL response.
Not applicable (NA)	No TLs are recorded at baseline.

Missing TL data

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

- A new lesion is recorded
- An NTL visit response of PD is recorded
- The sum of available 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 an LD recorded.

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 will also be NE, unless there is a progression of non-TLs or new lesions, in which case the response will be PD.

Lymph nodes

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

TL Visit responses subsequent to CR

A CR response can only be followed by CR, PD or NE. If a CR has occurred, then the following rules at the subsequent visits must be applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node LD increases by 20% but remains < 10 mm.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD is also met.

- Step 3: If not all lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD
- Step 4: If after steps 1 – 3 a response can still not be determined, the response will be set to remain as CR

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure, then this will be indicated as such on the eCRF and a value of 5 mm will be entered into the database and used in TL calculations.

However, a smaller value may be used if the radiologist has not indicated 'too small to measure' on the eCRF and has entered a smaller value that can be reliably measured. If a TL response results in PD, then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions may be considered measurable and selected as target lesions, provided they fulfil the other criteria for measurability (i.e., reproducible measurability). This is because all patients will have received radiation therapy immediately prior to randomised study treatment as part of their standard of care.

Any TL (including lymph nodes), which has had intervention in addition to study treatment during the study (e.g. irradiation / palliative surgery / embolization), should be handled in the following way and once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours:

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

scaling results in a visit response of PD then the patient would be assigned a TL response of PD.
Note: scaling up will only be performed if ≤ 1/3 of the TL measurements have been set to missing following an intervention. If > 1/3 of TL measurements are set to missing, then TL response will be set to NE.

- Step 3: If, after both steps, PD has not been assigned then, if appropriate, the scaled sum of diameters calculated in step 2 should be used, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR on the condition that all non-intervened lesions are 0 (or < 10 mm for lymph nodes) and the lesions that have been subject to intervention also 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 will be repeated to determine the TL and overall visit response. When calculating the previous minimum (the nadir), lesions with intervention should be treated as missing and scaled up where appropriate (as per step 2 above).

Scaling (applicable only for irradiated lesions/lesion intervention)

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

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

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

Table 3 provides an example of scaling-up. Lesion 5 has received intervention prior to the follow-up visit and has been set to missing. The sum of lesions 1-4 at the follow-up is 26 cm. The sum of the corresponding lesions at the nadir visit is 26.8 cm. Scale up as follows to give an estimated TL sum of 28.4 cm:

$\frac{26}{26.8}$ x 29.3 = 28.4

Table 3 Example of scaling-up

Lesion	Longest diameter at baseline visit	Longest diameter at follow-up visit
1	7.2	7.1
2	6.7	6.4
3	4.3	4.0
4	8.6	8.5
5	2.5	Missing
Sum	29.3	26

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of TLs

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

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

3.1.1.2 Non-Target Lesions (NTLs) and new lesions

At each visit an overall assessment of the NTL response should be recorded by the investigator. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit. NTL response will be derived based on the Investigator's overall assessment of NTLs as detailed in Table 4.

Table 4 NTL Visit Responses

Visit Responses	Description
Complete response (CR)	Disappearance of all NTLs since baseline. All lymph nodes must be non-pathological in size (< 10 mm short axis).
Non CR/non PD	Persistence of one or more NTLs with no evidence of progression.
Progression (PD)	Unequivocal progression of existing NTLs, which 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.
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.
Not applicable (NA)	Only relevant if there are no NTLs at baseline.

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

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more new lesions is assessed as progression. A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

The finding of a new lesion should be unequivocal: i.e., not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumour. New lesions will be identified via a Yes/No tick box. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses. A new lesion should be unequivocal and indicates progression, so the overall visit response will be PD irrespective of the TL and NTL response.

If a new lesion is equivocal, for example because of its small size, the treatment and tumour assessments should be continued until the new lesion has been confirmed. If repeat scans confirm there is a new lesion, then the progression date should be declared using the date of the initial scan.

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

Patients with 'symptomatic deterioration' 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.

3.1.1.3 Overall visit response

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

Table 5 Overall visit responses

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
CR	CR (or NA)	No (or NE)	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	Non-PD or NE or NA	No (or NE)	PR
SD	Non-PD or NE or NA	No (or NE)	SD
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD
NE	Non-PD or NE or NA	No (or NE)	NE
NA	CR	No (or NE)	CR
NA	Non-CR/Non-PD	No (or NE)	SD
NA	NE	No (or NE)	NE
NA	NA	No (or NE)	NED

CR Complete response; PR Partial response, SD Stable disease, PD Progressive disease, NE Not evaluable, NED No evidence of disease, NA Not applicable (only relevant if there were no TL/NTL at baseline).

3.1.2 Blinded Independent Central Review (BICR)

A planned BICR of all radiological imaging data will be carried out using RECIST 1.1. All radiological scans for all patients (including those at unscheduled visits, or outside visit windows) will be collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for central analysis. The imaging scans will be reviewed by two independent radiologists using RECIST 1.1 and will be adjudicated, if

required (i.e. two reviewers' review the scans and adjudication is performed by a separate reviewer in case of a disagreement).

For each patient, the BICR will define the overall visit response (i.e. the response obtained overall at each visit by assessing TLs, NTLs and new lesions) data and no programmatic derivation of visit response is necessary. (For patients with TLs at baseline: CR, PR, SD, PD, NE; for patients with NTLs only: CR, SD, PD, NE; for patients with no disease identified at baseline: PD, no evidence of disease [NED], NE). If a patient has had a tumour assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST 1.1 assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression. Endpoints (of ORR, PFS, DoR, etc..) will be derived programmatically from this information.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision will be used). In the absence of adjudication, the records for all visits for a single reviewer will be used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to report all BICR RECIST 1.1 information including dates of progression, visit response, censoring and changes in TL dimensions.

Results of this independent review will not be communicated to investigators and the management of patients will be based solely upon the results of the RECIST 1.1 assessment conducted by the investigator.

A BICR of all patients will be performed for the final database lock for PFS, which will cover all of the scans up to the DCO.

Further details of the BICR will be documented in the Independent Review Charter (IRC).

BICR according to RECIST 1.1 will be regarded as primary in terms of the efficacy analyses.

3.2 Outcome variables

All RECIST 1.1 assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues investigational product (IP).

RECIST 1.1 outcomes (i.e. PFS, ORR etc) will be derived using the overall visit responses and relevant dates from the BICR. This will be repeated using the programmatically derived overall visit responses from investigator RECIST 1.1 assessments.

3.2.1 Progression-Free Survival (PFS)

The primary endpoint is PFS per RECIST 1.1 as assessed by BICR. PFS will be defined as the time from the date of randomisation until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from therapy or receives another anti-cancer therapy prior to progression (i.e., date of event or censoring – date of randomisation + 1).

Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the 2 missed visits (note: NE visit is not considered a missed visit). If the patient has no evaluable visits or does not have baseline data or have baseline data but no-post baseline data, they will be censored at study day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window), which then will be treated as an event with date of death as the event date.

Given the scheduled visit assessment scheme (i.e. after randomisation every 8 weeks ± 1 week for the first 48 weeks and then every 12 weeks ± 1 week thereafter until clinical progression/deterioration) the definition of 2 missed visits will change over time, as described in Table 6.

Table 6 PFS: Definition of 2 Missed Visits

Timing of Previous RECIST Assessment	Two Missed Visits	Explanation
Study day < 50 (before week 8 day 1)	17 weeks	17 weeks since the date of randomisation (if no post-baseline assessment) or 17 weeks since the previous RECIST assessment (if has post-baseline assessment) (i.e., 2 × 8 weeks + 1 week for a late assessment = 17 weeks)
Study day 50 to < 274 (i.e., week 8 to week 39)	18 weeks	18 weeks since the previous RECIST assessment (i.e., 2 × 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks)
Study day 274 to < 330 (week 40 to week 47)	22 weeks	22 weeks since the previous RECIST assessment

Timing of Previous RECIST Assessment	Two Missed Visits	Explanation
		(i.e., 8 weeks + 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 22 weeks)
Study day ≥ 330 (week 48 onwards)	26 weeks	26 weeks since the previous RECIST assessment (i.e., 2 × 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks)

Study day is calculated as (date of assessment – date of randomisation + 1)

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

RECIST 1.1 assessments/scans contributing toward a particular visit may be performed on different dates within the imaging visit window. The following rules will be applied:

- For BICR assessments, the date of progression will be determined based on the **earliest** scan dates of the component that triggered the progression on the first set of scans that indicates progression for the adjudicated reviewer selecting PD, or where both reviewers select PD as a timepoint response and there is no adjudication, then the assessments of the reviewer who completed their baseline assessments first are used for the PD timepoint.
- For Investigator assessments, the date of progression will be determined based on the **earliest** of the RECIST 1.1 assessment/scan dates of the component that indicates progression.
- When censoring a patient for PFS, the patient will be censored at the **latest** of the scan dates contributing to a particular overall visit assessment.

Note: At each imaging visit, an overall timepoint response is algorithmically derived according to assessments of target lesions, non-target lesions, and new lesions (see CSP Appendix E). For TLs, only the latest scan date, within an imaging visit window, is recorded in the RECIST 1.1 eCRF out of all scans performed at that assessment for the TLs. Similarly, for NTLs, only the latest scan date is recorded out of all scans, within an imaging visit window, is recorded in the RECIST 1.1 eCRF out of all scans performed at that assessment for the NTLs. For new lesions, only the first scan date, within an imaging visit window, is recorded in the RECIST 1.1 eCRF out of all scans performed at that assessment of new lesions.

A sensitivity analysis of PFS will be performed using Investigator assessments according to RECIST 1.1. If applicable, another sensitivity analysis will be performed by excluding sCRT

patients with 1 concurrent chemotherapy and radiation cycle using BICR tumour data (RECIST 1.1).

3.2.2 Secondary efficacy outcome endpoints

For analysis purpose and except otherwise specified, subsequent anti-cancer therapies include radiotherapies, except palliative radiotherapies.

3.2.2.1 Overall survival

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

Note: Survival calls will be made in the week following the date of DCO for the analysis (these contacts should generally occur within 7 days of the DCO). If patients are confirmed to be alive or if the death date is after the DCO date, these patients will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the final OS analysis should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly available death registries. If the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

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

a. For missing day only – using the 1st of the month
b. For missing day and month – using the 1st of January

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

Note that for any OS analysis performed prior to the final OS analysis, 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 patient was known to be alive for those patients still on-treatment (since the SURVIVE module is only completed for patients off treatment if a survival sweep is not performed). The last date for each individual patient is defined as the

latest date recorded on the CRFs. This includes, but is not limited to, the following dates recorded on the CRF:

- Adverse event (AE) start and stop dates
- Admission and discharge dates of hospitalisation
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST 1.1 CRF
- Start and stop dates of alternative anti-cancer treatment
- Date last known alive on survival status CRF
- End of study date.

3.2.2.2 **Proportion of patients alive at 24 months (OS24)**

OS24 will be defined as the Kaplan-Meier estimate of OS at 24 months after randomisation.

3.2.2.3 **Objective response rate (ORR)**

ORR (per RECIST 1.1 as assessed by BICR) is defined as the number (%) of patients with at least one visit response of CR or PR (without requirement for the first documented response to be subsequently confirmed) and will be based on all patients in the subset of the analysis population including only those patients with measurable disease at baseline per BICR.

For ORR without confirmation, data obtained up until progression, or the last assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue treatment without progression, receive a subsequent anti-cancer therapy, and then respond will not be included as responders in the ORR.

Note: For the purposes of BoR and DoR summaries, described below, 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, will be included in the BoR calculation. Patients who discontinue treatment without progression, receive a subsequent anti-cancer therapy and then respond will not be included as responders in the BoR calculation (i.e. both visits contributing to a response must be prior to subsequent therapy for the patient to be considered as a responder). In the case where a patient 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 patient will be defined as a

responder. Similarly, if a patient 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 will be assigned.

Change in tumour size

For supportive purposes percentage change from baseline in tumour size will be derived at each scheduled tumour assessment visit (i.e., 8-weekly until 48 weeks, 12-weekly until RECIST 1.1-defined radiological progression, hereafter referred to as week X for convenience) using BICR data and site investigator data. Best percentage change from baseline in tumour size will also be derived as the biggest decrease or the smallest increase in tumour size from baseline.

This is based on RECIST TL measurements taken at baseline and at the timepoint of interest. Tumour size is defined as the sum of the longest diameters of the TLs based upon RECIST assessments. TLs are measurable tumour lesions. Baseline for RECIST is defined to be the last assessment prior to randomisation. The change in TL tumour size at week X will be obtained for each patient by taking the difference between the sum of the TLs at week X and the sum of the TLs at baseline. To obtain the percentage change in TL tumour size at week X the change in TL tumour size is divided by the sum of the TLs at baseline and multiplied by

100 (i.e., (week X - baseline) / baseline * 100).

The above derivations will be programmed for the BICR based upon RECIST 1.1 assessments. Measurements from the reviewer selected by the adjudicator will be used when adjudication for overall visit response has occurred, but in the case where no adjudication was required the measurements from the reviewer who reviewed the baseline scan first will be used for this analysis.

Best objective response (BoR)

BoR is calculated based on the overall visit responses from each RECIST 1.1 BICR assessment. It is the best response a patient has had during their time in the study (following randomisation but prior to starting any subsequent anti-cancer therapy) up to and including RECIST 1.1 progression or the last evaluable assessment in the absence of RECIST 1.1 progression.

Categorisation of BoR will be based on RECIST 1.1 using the following response categories: CR, PR, SD, PD, NE and NED (only applicable when considering patients without measurable disease at baseline).

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 8 weeks minus 1 week, i.e. at least 49 days (to allow for an early assessment within the assessment window), after randomisation. For CR/PR, the initial overall visit assessment which showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

BoR will be determined programmatically based on RECIST 1.1 progression from the overall visit responses using all BICR data up until the first progression event or the last evaluable assessment in the absence of RECIST 1.1 progression. It will also be determined programmatically based on RECIST 1.1 using all site investigator data up until the first progression event. The denominators for each case will be consistent with those used in the ORR analysis.

For patients whose PFS event is death, BoR will be calculated based upon all evaluable RECIST 1.1 assessments prior to death.

For patients who die with no evaluable RECIST 1.1 assessments, if the death occurs ≤ 17 weeks (i.e., 16 weeks + 1 week for a late assessment) after randomisation, then BoR will be assigned to the progression (PD) category. For patients who die with no RECIST 1.1 assessments, if the death occurs > 17 weeks (i.e., 16 weeks + 1 week) after the date of randomisation, then BoR will be assigned to the NE category.

Progression events that were censored after two missed visits will not contribute to the BoR derivation. Given the schedule of assessment scheme, the definition of two missed visits will change depending on whether the patient is currently scheduled to have a visit every 8 weeks or every 12 weeks – please see section 3.2.1 for details of how to derive two missed visits based on the assessment schedule.

A patient will be classified as a responder if the RECIST 1.1 criteria for a CR or PR defined in the section above are satisfied at any time following randomisation, prior to RECIST 1.1 progression and prior to starting any subsequent anti-cancer therapy.

3.2.2.4 **Duration of response (DoR)**

DoR (per RECIST 1.1 as assessed by BICR) will be defined as the time from the date of first documented response until the first date of documented progression or death in the absence of disease progression (i.e., date of PFS event or censoring – date of first response+1). The end of response should coincide with the date of progression or death from any cause used for the RECIST 1.1 PFS endpoint.

The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of CR or PR. If a patient does not progress following a response, then the DoR will be censored at the PFS censoring time. DoR will be summarised for the RECIST 1.1 BICR data for patients with objective response and without the requirement for a confirmed response.

DoR will additionally be summarised, as described above, for patients with confirmed objective response. Thus, the time of the initial response will be defined as the latest of the dates contributing towards the first visit that was CR or PR that was subsequently confirmed.

Time to response (TTR)

For supportive purposes, time to response (without requirement for confirmation of response) will be reported. Time to response is defined as the time from the date of randomisation until the date of first documented response. The date of first documented response should coincide with that used for the DoR endpoint.

Analysis will be based on all patients with objective response.

3.2.2.5 Progression-free survival at 12 (PFS12) and 18 months (PFS18)

The PFS12 and PFS18 will be defined as the KM estimate of PFS (per RECIST 1.1 as assessed by the Investigator) at 12 and 18 months, respectively and both will be obtained using the algorithm described above for the RECIST 1.1 site Investigator tumour data.

3.2.2.6 Time to death or distant metastasis (TTDM)

TTDM (per RECIST 1.1 as assessed by BICR) will be defined as the time from the date of randomisation until the first date of distant metastasis or death in the absence of distant metastasis. Distant metastasis is defined as any new lesion that is outside of the radiation field according to RECIST 1.1. For this reason, the TTDM endpoint will be determined from recurrent disease that occurs outside of the structures contained within the thorax, including lymph nodes, pulmonary, pleural, and mediastinal metastatic sites and excluding the heart. The locations of distant metastases will be defined and documented prior to database lock/unblinding.

Patients who have not developed distant metastasis or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient has distant metastasis or dies after 2 or more missed visits, the patient will be censored at the time of the latest RECIST 1.1 assessment prior to the 2 missed visits. If the patient has no evaluable visits or does not have baseline data, they will be

censored at Day 1 unless they die within 2 visits of baseline, which then will be treated as an event with date of death as the event date.

3.2.2.7 Time from randomisation to second progression or death (PFS2)

Time from randomisation to second progression or death (PFS2) will be defined as the time from the date of randomisation to the earliest of the progression event subsequent to first subsequent therapy, or death.

There are two preconditions to have PFS2 second progression event: 1) observation of primary PFS event (first progression) by Investigator; 2) received subsequent anti-cancer therapy. The primary PFS event includes only radiological progression by RECIST 1.1, as per Investigator.

The date of second progression will be recorded by the Investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: objective radiological imaging, symptomatic progression, or death. Second progression status will be reviewed (q8w ± 1 week for the first 48 weeks [relative to the date of randomisation] and then q12w ± 1 week [the first progression]) following the first progression event used for the primary variable PFS and status recorded.

Subsequent anti-cancer therapy considered in PFS2 assessment include systemic treatment (chemotherapy, targeted therapy, immunotherapy, etc.) and definitive radiotherapy. Palliative radiotherapy and other local treatment such as surgery are not considered as subsequent anti-cancer therapy in PFS2 assessment.

If death occurs without first or second progression, the death will be a PFS2 death event irrespective of whether subsequent therapy has started.

Patients alive and for whom a second disease progression has not been observed should be censored at date last known alive and without a second disease progression (i.e. censored at:

- The PFS assessment date if the patient has not had a first progression or death (PFS censoring date)
- The date the patient is last known to not have received a first subsequent therapy (latest TTSCAPRX=No) or the date known to be alive for any patient without a TTSCAPRX form if a patient has had a first progression and not started a subsequent therapy (TFST censoring date)
- The latest PFS2 assessment date if the patient has started a first subsequent therapy and PFS2 event (second progression or death) has not been observed).

3.2.3 Patient-reported outcomes (PRO)

Symptoms and overall quality of life (QoL) will be assessed using the PRO questionnaires, European Organisation for Research and Treatment of Cancer (EORTC) 30-item core (QoL) questionnaire (QLQ-C30) (Aaronson et al 1993) and 13-item lung cancer QoL questionnaire (QLQ-LC13) (Bergman et al 1994) (secondary endpoints). All questionnaires will be scored according to published guidelines or the developer's guidelines, if published guidelines are not available. All PRO data will be collected until PFS2.

3.2.3.1 EORTC QLQ-C30 and QLQ-LC13

The EORTC QLQ-C30 consists of 30 questions that can be combined to produce 5 functional scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, pain, and nausea/vomiting), 6 individual items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea and perceived financial impact of disease), and a global measure of health status. The global health status/QoL will be assessed using 2 items from the QLQ-C30: "How would you rate your overall health during the past week? (Item 29) and "How would you rate your overall QoL during the past week? (Item 30).

The QLQ-LC13 is a lung cancer specific module from the EORTC for lung cancer comprising 13 questions to assess lung cancer symptoms (cough, haemoptysis, dyspnoea, and site-specific pain), treatment-related symptoms (sore mouth, dysphagia, peripheral neuropathy, and alopecia), and pain medication. Except for a multi-item scale for dyspnoea, all are single items. The dyspnoea scale will only be used if all 3 items have been scored; otherwise, the items are treated as single-item measures.

An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales/symptom items, the functional scales and the global health status/QoL scale in the QLQ-C30 according to the EORTC QLQ-C30 Scoring Manual (EORTC QLQ-C30 Scoring Manual, Third Edition) and for each of the symptom scales/items in the QLQ-LC13 according to the EORTC QLQ-LC13 instructions.

Higher scores on the global health status/QoL and functioning scales indicate better health status/function, but higher scores on symptom scales/items represent greater symptom severity. For each scale, if <50% of the scale items are missing, then the scale score will be divided by the number of non-missing items and multiplied by the total number of items on the scales (Fayers et al 2001). If at least 50% of the items are missing, then that scale will be treated as missing. Missing single items are treated as missing. The reason for any missing questionnaire will be identified and recorded. If there is evidence that the missing data are systematic, missing values will be handled to ensure that any possible bias is minimised.

The five symptoms that have been identified as primary PRO endpoints are:

- Dyspnoea: multi-item scale based on 3 questions ("Were you short of breath when you rested; walked; climbed stairs?" – QLQ-LC13)
- Cough: 1 item ("How much did you cough?" – QLQ-LC13)
- Chest pain: 1 item ("Have you had pain in your chest?" – QLQ-LC13)
- Fatigue: multi-item based on 3 questions ("Did you need rest?; Have you felt weak?; Were you tired?" – QLQ-C30)
- Appetite loss: 1 item ("Have you lacked appetite?" – QLQ-C30)

The physical functioning and overall health status domains of the EORTC QLQ-C30 are furthermore pre-specified endpoints of interest.

Change from baseline

Change from baseline in the key PRO symptom scores of dyspnoea (LC13), cough (LC13), pain in chest (LC13), fatigue (C30) and appetite loss (C30), physical functioning (C30) and overall health status (C30) will be analysed making use of all data from baseline.

Definition of clinically meaningful changes

A minimum clinically meaningful change is defined as an absolute change in the score from baseline of ≥10 for scales/items from the EORTC QLQ-C30 and the QLQ-LC13 (Osoba et al 1998). For example, a clinically meaningful improvement in physical function (as assessed by EORTC QLQ-C30) is defined as an increase in the score from baseline ≥ 10, whereas a clinically meaningful deterioration is defined as a decrease in the score from baseline of ≥ 10. At each post-baseline assessment, the change in symptoms/functioning from baseline will be categorised as improvement, no change, or deterioration, as shown in Table 7.

Table 7 Visit response in symptoms, functioning, and global health status/QoL

Score	Change from baseline	Visit response
EORTC QLQ-C30 /QLQ-LC13 symptom scales/items	≥ + 10	Worsened
	≤ - 10	Improved
	Otherwise	Stable
EORTC QLQ-C30 functional scales and global health status/QoL	≥ + 10	Improved
	≤ - 10	Worsened
	Otherwise	Stable

Time to symptom deterioration/worsening (QLQ-C30 and QLQ-LC13)

For each of the symptoms scales/items in the EORTC QLQ-C30 and QLQ-LC13, time to symptom deterioration/worsening will be defined as the time from randomisation until the date of the first clinically meaningful symptom deterioration/worsening (an increase in the score from baseline of ≥10) that is confirmed at the next available subsequent assessment at least 14 days apart, or death (by any cause) in the absence of a clinically meaningful symptom deterioration/worsening, regardless of whether the patient withdraws from IP or receives another anti-cancer therapy prior to symptom deterioration/worsening. Missed visits are allowed in between assessments confirming deterioration/worsening. This is considered a conservative approach whereby a deterioration/worsening is considered a 'negative' outcome and therefore should be assigned as such, regardless of missed visits. Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the symptom change could be evaluated. Patients with a single deterioration/worsening and no further assessments will be treated as deteriorated in the analysis.

Patients whose symptoms (as measured by EORTC QLQ-C30 and QLQ-LC13) have not shown a clinically meaningful deterioration/worsening and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if symptoms deteriorate after 2 or more missed PRO assessment visits or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated, prior to the 2 missed assessment visits. The 2 missed visit rule for ePRO will take a 'look-forward' approach i.e. if there are 2 consecutive missed visits at any time prior to the confirmed deterioration event, the event will be censored at the last available assessment prior to the 2 missed visits. Confirmation of deterioration will be first determined, then the censoring rules will be applied.

See Appendix A for further details on the derivation of the confirmation of deterioration 2 missed visit rule for ePRO and length of 2 missed visit window.

If the patient has no evaluable visits or does not have baseline data, they will be censored at day 1 unless they die within 2 visits of baseline (8 weeks plus 3 days allowing for a late assessment within the visit window).

The population for the analysis of time to symptom deterioration will include patients who have baseline scores of ≤ 90.

Time to HRQoL/function deterioration/worsening (QLQ-C30)

For HRQoL and function (as measured by EORTC QLQ-C30), time to deterioration/worsening will be defined as the time from the date of randomisation until the date of the first clinically meaningful deterioration/worsening (a decrease in the function scales or the global health status/QoL score from baseline of ≥10) that is confirmed at the next available subsequent assessment at least 14 days apart, or death (by any cause) in the absence of a clinically meaningful deterioration/worsening, regardless of whether the patient withdraws from IP or receives another anti-cancer therapy prior to HRQoL/function deterioration/worsening. Missed visits are allowed in between assessments confirming deterioration. This is considered a conservative approach whereby a deterioration is considered a 'negative' outcome and therefore should be assigned as such, regardless of missed visits. Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the HRQoL/function change could be evaluated. Patients with a single deterioration/worsening and no further assessments will be treated as deteriorated in the analysis.

Patients whose HRQoL or function have not shown a clinically meaningful deterioration/worsening and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the HRQoL/function could be evaluated, prior to the 2 missed visits. Also, if HRQoL or function deteriorates after 2 or more missed PRO assessment visits or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where HRQoL/function could be evaluated. The 2 missed visit rule for ePRO will take a 'look-forward' approach i.e. if there are 2 consecutive missed visits at any time prior to the confirmed deterioration event, the event will be censored at the last available assessment prior to the 2 missed visits. Confirmation of deterioration will first be determined, then the censoring rules will be applied. If the patient has no visits or does not have baseline data, they will be censored at day 1 unless they die within 2 visits of baseline (8 weeks plus 3 days allowing for a late assessment within the visit window).

See Appendix A for further details on the derivation of the confirmation of deterioration, 2-missed visit rule and length of 2 missed visit window.

The population for the analysis of time to QoL/function deterioration/worsening will include patients who have baseline scores of ≥10.

Symptom improvement rate (QLQ-C30 and QLQ-LC13)

The symptom improvement rate will be defined as the number (%) of patients with a minimum of 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (a decrease in score of ≥10) in that symptom from baseline. Missed visits are not allowed in between consecutive assessments of improvement. This is considered a conservative approach whereby an improvement is considered a 'positive' outcome and therefore for any avoidance of doubt in the assessment of an improvement, missed visits are not allowed.

The denominator will consist of the subset patients who have a baseline symptom score of ≥10.

HRQoL/function improvement rate (QLQ-C30)

The HRQoL/function improvement rate will be defined as the number (%) of patients with a minimum of 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (an increase in score of ≥10) in that scale from baseline. Missed visits are not allowed in between consecutive assessments of improvement. This is considered a conservative approach whereby an improvement is considered a 'positive' outcome and therefore for any avoidance of doubt in the assessment of an improvement, missed visits are not allowed.

The denominator will consist of the subset of patients who have a baseline QoL/function score of ≤90.

3.2.3.2 Compliance rate

Summary measures of overall compliance and compliance over time will be derived for each EORTC-QLQ-C30 and EORTC-QLQ-LC13, respectively. These will be based upon:

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

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

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

3.2.4 Other efficacy outcome endpoints

3.2.4.1 Time to first subsequent therapy or death (TFST)

As a supportive summary to PFS, time to first subsequent therapy or death (TFST) is defined as the time from the date of randomisation to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomised treatment, or death (i.e. date of first subsequent anti-cancer therapy/death or censoring – date of randomisation + 1). Any patient not known to have had a first subsequent anti-cancer therapy or have not died at the time of the analysis will be censored at the last date that the patient was known not to have received a first subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a patient terminated the study for reason other than death before first subsequent therapy, these patients will be censored at the earliest of their last known to be alive and termination dates. Patients not receiving randomised treatment would have TFST calculated in the same way, i.e. time from date of randomisation to the initial subsequent therapy or death. Any participant without a TTSCAPRX form (not known to have had a first subsequent anti-cancer therapy) and have not died at the time of the analysis is censored at the last date that the participant was known to be alive, where any participant recorded as alive or to have died after DCO date is censored at the date of DCO.

3.2.4.2 Time to second subsequent therapy or death (TSST)

As a supportive summary to PFS2, time to second subsequent therapy or death (TSST) is defined as the time from the date of randomisation to the earlier of start date of the second subsequent anti-cancer therapy after discontinuation of first subsequent treatment, or death (i.e. date of second subsequent anti-cancer therapy/death or censoring – date of randomisation + 1). Any patient not known to have had a second anti-cancer subsequent therapy or have not

died at the time of the analysis will be censored at the last date that the patient was known not to have received a second subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a patient terminated the study for reason other than death before second subsequent therapy, these patients will be censored at the earliest of their last known to be alive and termination dates. Patients not receiving randomised treatment would have TSST calculated in the same way, i.e. time from date of randomisation to the second subsequent therapy or death. Any participant without a TTSCAPRX form (not known to have had a second subsequent anti-cancer therapy) and have not died at the time of the analysis is censored at the last date that the participant was known to be alive, where any participant recorded as alive or to have died after DCO date is censored at the date of DCO.

3.3 Safety variables

Safety and tolerability will be assessed in terms of AEs (including serious AEs [SAEs]), deaths, laboratory data, vital signs, electrocardiograms (ECGs) and exposure. These will be collected for all patients.

Unless otherwise specified, ‘on-treatment’ will be defined as assessments between date of the first dose and 90 days following last dose of the IP (durvalumab/placebo) or initiation of the first subsequent anti-cancer therapy (excluding palliative radiotherapy), whichever occurs first.

3.3.1 Adverse events

AEs and SAEs will be collected throughout the study, from date of the patient signing the ICF of study procedure (part II screening) until 90 days after the last dose of study treatment. Events will be defined as treatment emergent if they onset, or worsen (by investigator report of a change in intensity), during the treatment period as defined in the protocol. The Medical Dictionary for Regulatory Activities (MedDRA) (using the latest or current MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 5.0).

AEs of special interest and AEs of possible interest

Some clinical concepts (including some selected individual preferred terms and higher level terms) have been considered “AEs of special interest” (AESI) and “AEs of possible interest” (AEPI) to the durvalumab program.

The AESIs reported in the AstraZeneca-sponsored durvalumab studies are defined as AEs with a likely inflammatory or immune-mediated pathophysiological basis resulting from the mechanism of action of durvaluamb and requiring more frequent monitoring and/or interventions such as corticosteroids, immunosuppressants, and/or endocrine therapy. Endocrine therapies include standard endocrine supplementation, as well as treatment of symptoms resulting from endocrine disorders (for example, therapies for hyperthyroidism include beta blockers [e.g., propranolol], calcium channel blockers [e.g., verapamil, diltiazem], methimazole, propylthiouracil, and sodium perchlorate).

The AEPIs reported in the AstraZeneca-sponsored durvalumab studies are defined as AEs that could have a potential inflammatory or immune-mediated pathophysiological basis resulting from the mechanism of action of durvaluamb but are more likely to have occurred due to other pathophysiological mechanisms, thus, the likelihood of the event being inflammatory or immune-mediated in nature is not high and/or is most often or usually explained by the other causes.

These AESIs and AEPIs have been identified as

Pneumonitis, Hepatic events, Diarrhea/Colitis, Intestinal perforations, Adrenal Insufficiency, Type 1 diabetes mellitus, Hyperthyroid events, Hypophysitis, Hypothyroid events, Thyroiditis, Renal events, Dermatitis/Rash, Pancreatic events, Myocarditis, Myasthenia gravis, Guillain-Barre syndrome, Myositis, Infusion/hypersensitivity reactions, Other rare/miscellaneous. Other categories may be added or existing terms may be merged as necessary. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician has reviewed the AEs of interest and identified which preferred terms contribute to each AESI/AEPI. A further review will take place prior to Database lock (DBL) to ensure any further terms not already included are captured within the categories.

Immune-mediated Adverse Events (imAEs)

imAEs will be identified from both AEs of special interest (AESIs) and AEs of possible interest (AEPIs) based on programmatic rules that consider interventions involving systemic steroid therapy, immunosuppressant use, and/or endocrine therapy (which, in the case of AEPIs, occurs after first considering an Investigator’s causality assessment and/or an

Investigator's designation of an event as immune-mediated). Endocrine therapies include standard endocrine supplementation, as well as treatment of symptoms resulting from endocrine disorders (for example, therapies for hyperthyroidism include beta blockers [e.g., propranolol], calcium channel blockers [e.g., verapamil, diltiazem], methimazole, propylthiouracil, and sodium perchlorate). Further details are provided in the imAE charter.

In addition, the Sponsor may perform medical review of those AESIs and classify them as imAEs or not imAEs via an independent manual adjudication process.

Confirmed/Suspected COVID-19 infection

Confirmed/suspected COVID-19 infection is defined as an AE with preferred term within the AE search criteria developed by the latest MedDRA MSSO guidance for COVID-19.

3.3.2 Laboratory data

Laboratory data will be collected throughout the study, from Part II screening to follow-up visit as described in the CSP. Blood and urine samples for determination of haematology, clinical chemistry, and urinalysis will be collected as described in section 8.2.1 of the CSP. For the definition of baseline and the derivation of post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in section 4.1.3 below will be used.

Change from baseline in haematology and clinical chemistry variables will be calculated for each post-dose visit on-treatment. CTCAE grades will be defined at each visit according to the CTCAE grade criteria using local or project ranges as required, after conversion of lab result to corresponding preferred units. The following parameters have CTCAE grades defined for both high and low values: Potassium, Sodium, Magnesium, Glucose and Corrected calcium.

Corrected calcium will be derived during creation of the reporting database using the following formula:

Corrected calcium (mmol/L) = Total Calcium (mmol/L) + ([40 – Albumin (G/L)] x 0.02)

Creatinine clearance (CrCl) (mL/min) will be derived according to the Cockroft-Gault formula (Cockcroft and Gault 1976), where the most recent weight before the laboratory assessment will be considered.

Males:

$$Creatinine\ CL = \frac{Weight(kg)\ \text{x}\ (140 - Age)}{72\ \text{x}\ serum\ creatinine\ (mg/dL)}$$

Females:

$$Creatinine\ CL = \frac{Weight(kg)\ \text{x}\ (140 - Age)}{72\ \text{x}\ serum\ creatinine\ (mg/dL)}\ \text{x}\ 0.85$$

Absolute values will be compared to the project reference range and classified as low (below range), normal (within range or on limits of range) and high (above range).

The maximum or minimum on-treatment value (depending on the direction of an adverse effect) will be defined for each laboratory parameter as the maximum (or minimum) post baseline value at any time.

Local reference ranges will be used for the primary interpretation of laboratory data at the local laboratory. Project reference ranges will be used throughout for reporting purposes.

The denominator used in laboratory summaries of CTCAE grades will only include evaluable patients, in other words those who had sufficient data to have the possibility of an abnormality.

For example:

- If a CTCAE criterion involves a change from baseline, evaluable patients would have both a baseline and at least 1 post-dose value recorded.
- If a CTCAE criterion does not consider changes from baseline, to be evaluable the patient need only have 1 post-baseline value recorded.

3.3.3 ECGs

Resting 12-lead ECGs are recorded at screening and as clinically indicated thereafter. Overall evaluation of ECG is collected in terms of normal, abnormal, or borderline, and the clinical relevance assessment for “abnormal” and “borderline” category is termed as “clinically significant” or “not clinically significant”.

In case of clinically significant ECG abnormalities, including a QT interval corrected for heart rate using Fridericia’s formula (QTcF) value >470 ms, 2 additional 12-lead ECGs should be obtained over a brief period (i.e., 30 minutes) to confirm the finding.

The QTcF will be derived during creating of the reporting database using the reported ECG values (RR and QT) using the following formula:

$QTcF = \frac{QT}{RR^{\frac{1}{3}}}$ where RR is in seconds.

ECG data obtained up until the safety follow-up (i.e. 30 days following last dose of the IP (durvalumab/placebo)) will be used for the reporting, and considered as "on-treatment".

3.3.4 Vital signs

Vital signs data obtained between the date of first dose and up until the 30 days from date of last dose of IP will be used for reporting, and considered as "on-treatment". Change from baseline in vital signs variables will be calculated for each post-dose visit on-treatment. For derivation of post baseline visit values considering visit window and to handle multiple records, derivation rules as described in section 4.1.3 below will be used. The denominator in vital signs data should include only those patients with recorded data.

3.3.5 Study treatments

Study treatment in this study refers to durvalumab, and placebo. See Table 8 for further details on the IPs. Exposure will be defined for durvalumab or placebo outlined below.

Table 8 Study treatments

	Study treatment name	Route of administration	Dosing instructions
Durvalumab	Durvalumab (MEDI4736)	IV	1500 mg IV q4w[b]
Placebo	Saline solution	IV	Saline volume matching durvalumab volume

[b] If a patient's weight falls to 30 kg or below, the patient should receive weight-based dosing equivalent to 20 mg/kg of durvalumab q4w until the weight improves to >30 kg, at which point the patient should start receiving the fixed dosing of durvalumab 1500 mg q4w.
IV Intravenous; q4w Every 4 weeks.

3.3.5.1 Treatment exposure for durvalumab or placebo

If a patient has a delay to receiving study treatment, all tumour efficacy assessments and PRO assessments should still be conducted relative to the date of randomisation.

Total (or intended) exposure of durvalumab (MEDI4736) or placebo

- Total (or intended) exposure = min (last dose date where dose>0mg+27 days, date of death, date of DCO) – first dose date +1

Actual exposure of durvalumab (MEDI4736) or placebo

- Actual exposure = intended exposure – total duration of dose delays, where intended exposure will be calculated as above.

Calculation of duration of dose delays (for actual exposure):

- Since patients will receive 1500 mg via intravenous (IV) infusion q4w, the duration of dose delays will be calculated as follows:

  For all dosing dates:

  Total duration of dose delays= Sum of (Date of the dose - Date of previous dose – 28 days)

  Thus, if no delays were encountered, the duration would sum up to 0, since infusions were done every four weeks.

Dose reductions are not permitted per the CSP for durvalumab. The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Patients who permanently discontinue during a dose delay

If a decision is made to permanently discontinue study treatment in-between cycles or during a dose delay, then the date of last administration of study medication recorded will be used in the programming.

3.3.6 Dose intensity

Dose intensity will be derived for study treatment durvalumab or placebo. Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to treatment discontinuation.

RDI will be defined as follows:

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule. When accounting for the calculation of intended cumulative dose 3 days should be added to the date of last dose to reflect the CSP allowed window for dosing.

When deriving actual dose administered the volume before and after infusion will also be considered.

Example of dose intensity for durvalumab (MEDI4736)

		Study day									
RDI	Patient	1	29	57	85	113	141	169	197	225	
100%	1	X	X	X	X	X	X	X	X	X	PD
100%	2	X	X	X	X	X	X	X	X[D]		PD
56%	3	X	X		X		O	X	X		PD

X: Dose of 1500g taken; O: Dose missed; [D] Dose discontinued; PD: Progressive disease

Patients 1-3 progressed on Day 230, so the intended dose through to progression was 9 * 1500mg of durvalumab = 13500mg (13.5g).

Patient 1 received a total of 13.5g of durvalumab, whereas other patients received less treatment due to:

- Early stopping prior to PD (Patient 2)
- Dosing delays (Patient 3)

The Patient 2 example illustrates that for RDI, the end of actual dosing period is calculated based on the smallest recovery period after the last non-zero dose.

Patient 1: RDI = (9 * 1.5g)/ 13.5g = 100%

Patient 2: RDI = (8 * 1.5g)/ 12g = 100%

Patient 3: RDI = (5 * 1.5g)/ 13.5g = 56%

3.3.7 Concomitant Medication

Any medications taken by the patient at any time between the date of the first dose (including the date of the first dose) of study treatment up to the date of last dose of study treatment in the study will be considered as concomitant medication. Any medication that started prior to the first dose of the study treatment and ended after the first dose or is ongoing will be considered as both prior and concomitant medication.

Allowed and disallowed concomitant medications will be presented by ATC classification and generic term.

3.4 Pharmacokinetic and Immunogenicity Variables

Analyses to evaluate the pharmacokinetics (PK) and immunogenicity of durvalumab will be performed by AstraZeneca/MedImmune Clinical Pharmacology group or designee. Only the

earliest available test result will be used for analysis if multiple tests are performed for one sample.

3.4.1 Pharmacokinetic analysis

Individual durvalumab concentrations will be listed by visit. Summary statistics of durvalumab concentrations will be calculated and tabulated by visit. Individual and mean blood concentration-time profiles will be generated. The following PK parameters will be determined after the first and steady-state doses: peak concentration and trough concentration (as data allow). Samples below the lower limit of quantification (LLOQ) will be treated as missing in the analyses. If the data are suitable, the relationship between PK exposure and efficacy/safety parameters may be investigated graphically or using an appropriate data modelling approach, which will not be included in the CSR of this study.

3.4.2 Immunogenicity analysis

Serum samples for anti-drug antibody (ADA) assessments will be conducted utilising a tiered approach (screen, confirm, titer), and ADA data will be collected at scheduled visits shown in the CSP. ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titer will be reported as well. In addition, the presence of neutralising antibody (nAb) may be tested for all ADA-positive samples using a ligand-binding assay. The nAb results will be reported as positive or negative.

The number of patients in the ADA analysis set who fulfil the following criteria will be determined. The percentage of ADA-positive patients in each of the category will be calculated, using the number of patients in the ADA analysis set of the treatment group as the denominator. A patient is defined as being ADA positive if a positive ADA result is available at any time, including baseline and all post-baseline measurements; otherwise ADA negative.

- ADA positive at any visit; the percentage of ADA-positive patients in the ADA analysis set is known as ADA prevalence.
- Treatment emergent ADA positive: the sum of both treatment-induced and treatment-boosted ADA; the percentage of patients fulfilling this criterion in the ADA analysis set is known as ADA incidence.
- ADA positive post-baseline and positive at baseline.
- ADA positive post-baseline and not detected at baseline (treatment-induced ADA).
- ADA not detected post-baseline and positive at baseline.
- Treatment-boosted ADA, defined as a baseline positive ADA titer that was boosted to a 4-fold or higher level (greater than the analytical variance of the assay) following drug administration.

- Persistently positive ADA, defined as having at least 2 post-baseline ADA positive measurements with at least 16 weeks (112 days) between the first and last positive measurement or an ADA positive result at the last available assessment. The category includes patients meeting these criteria who are ADA positive at baseline.
- Transiently positive ADA, defined as having at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive. The category includes patients meeting these criteria who are ADA positive at baseline.
- nAb positive at any visit.

Immunogenicity results will be analysed descriptively by summarising the number and percentage of patients who develop detectable ADAs against durvalumab. The immunogenicity titre and presence of neutralising ADAs will be reported for samples confirmed positive for the presence of ADAs. The effect of immunogenicity on PK, pharmacodynamics, efficacy, and safety will be evaluated, if the data allow.

3.4.3 Biomarker data

The relationship of programmed cell death ligand 1 (PD-L1) expression and, if applicable, of exploratory biomarkers to clinical outcomes (including but not restricted to) of PFS, OS, ORR, and DoR may be presented.

PD-L1 expression determined by immunohistochemistry (IHC) will be reported in the CSR.

Remaining exploratory analysis described in the CSP may be reported outside the main CSR, including (and not limited to) the following exploratory endpoints as specified in CSP section 3:

- Blood CCI relative to efficacy outcomes (OS, PFS, and ORR)
- Biomarker analysis of blood and tissue to assess exploratory markers
- Biomarker status before and after treatment and durvalumab PK exposure and relationship with clinical outcomes, efficacy, AEs and/or safety parameters, as deemed appropriate
- Correlation of biomarkers with response to durvalumab treatment and/or the progression of cancer

A separate SAP for the remaining exploratory analysis as above will be prepared where necessary.

3.5 Exploratory variables

3.5.1 Calculation or derivation of patient-reported health state utility (EQ-5D-5L)

The health state utility will be assessed using the EQ-5D-5L index (EuroQoL 2013). The EQ-5D is a standardised measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care.

The index comprises of 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) (EuroQoL 2013). For each dimension, respondents select which statement best describes their health on that day from a possible 5 options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems, and extreme problems). A unique EQ- 5D health state is referred to by a 5-digit code, allowing for a total of 3125 health states. For example, state 11111 indicates no problems on any of the 5 dimensions. These data will be converted into a weighted health state index by applying scores from EQ-5D value sets elicited from general population samples (the base case will be the United Kingdom valuation set, with other country value sets applied in scenario analyses). Where value sets are not available, the EQ-5D-5L to EQ-5D-3L crosswalk will be applied (Oemar and Janseen 2013). Change from baseline in utility scores will be evaluated. In addition to the descriptive system, respondents also assess their health on the day of assessment on a visual analogue scale, ranging from 0 (worst imaginable health) to 100 (best imaginable health). This score is reported separately.

3.5.2 Health care resource use

To investigate the impact of treatment and disease on health care resource, the following exploratory variables may be captured:

- Planned and unplanned hospital attendances beyond trial protocol mandated visits (including physician visits, emergency room visits, day cases and admissions).
- Primary sign or symptom for hospital/inpatient/emergency room visit
- Length of hospital stay
- Planned and unplanned hospital attendances beyond trial protocol mandated visits (including physician visits, emergency room visits, day cases and admissions)
- Primary sign or symptom for hospital/inpatient/emergency room visit
- Length of hospital stay
- Length of any time spent in an intensive care unit (ICU)

Where admitted overnight, the length of hospital stay will be calculated as the difference between the date of hospital discharge (or death date) and the start date of hospitalisation or start of study drug if the start of study drug is after start date of hospitalisation (length of hospital stay = end date of hospitalisation – start date of hospitalisation + 1). Patients with missing discharge dates will be calculated as the difference between the last day with available data and the start date of hospitalisation. The length of ICU stay will be calculated using the same method.

3.5.3 Collection of data due to COVID-19

Due to COVID-19 pandemic, two eCRF pages were added to collect details on visits impacted (eCRF page [VISITP]) and concomitant medications not started (eCRF page [CMNSP]). More details were provided in the eCRF completion guidelines (v06, 27Apr2021) for those new pages and for other pages.

4 ANALYSIS METHODS

There will be 1 treatment comparison of interest:

- Durvalumab (MEDI4736) 1500mg vs placebo

Formal statistical analyses will be performed to test the main hypothesis:

- H0: No difference between durvalumab and placebo
- H1: Difference between durvalumab and placebo

The primary endpoint is PFS as assessed by BICR. The study has been sized to characterise the PFS benefit of durvalumab relative to placebo.

Results of all statistical analyses will be presented using a 95% confidence interval (CI) and 2-sided p-value, unless otherwise stated.

4.1 General principles

Efficacy and PRO data will be summarised and analysed on the mITT set. PK data will be summarised on the PK analysis set. Safety and treatment exposure will be summarised based upon the safety analysis set, except deaths that will be analysed on the ITT set. Immunogenicity data will be summarised on the ADA analysis set. Study population and demography data will be summarised based upon the mITT set.

Selected analyses based on the mITT set will be repeated on the ITT set. For supporting listings, those will be produced on ITT and mITT patients will be identified.”

All data collected will be presented in data listings by treatment group and subject number. All summaries will be presented by treatment group, unless otherwise specified.

A year is defined to be 365.25 days and a month is defined to be 30.4375 days.

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

- Descriptive statistics will be used for all variables, as appropriate, and will be presented by treatment arm. Continuous variables will be summarised by the number of observations, mean, standard deviation, median, upper and lower quartiles, minimum, and maximum. For log transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables will be summarised by frequency counts and percentages for each category.
- Unless otherwise stated, percentages will be calculated out of the population total and for each treatment arm.
- For continuous data, the mean and median will be rounded to 1 additional decimal place compared to the original data and rounded up (i.e., if original data is provided to 1 decimal place, a mean of 2.425 will become 2.43). The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.
- For categorical data, percentages will be rounded to 1 decimal place.
- Results of all statistical analysis will be presented using a 95% CI and 2-sided p-value, unless otherwise stated.
- P-values will be rounded to 3 decimal places. P-values less than 0.0005 (e.g. 0.0002) will not be rounded to 3 decimal places (e.g. 0.000) but instead be displayed as <0.001. P-values output as <0.0001 by statistical software will not be rounded and displayed in the same way (‘<0.0001’).
- SAS® Enterprise Guide (EG) will be used for all analyses.

4.1.1 Baseline

In general, for efficacy and PRO endpoints the last observed measurement prior to or on the date of randomisation will be considered the baseline measurement. However, for PRO endpoints if an evaluable assessment is only available after randomisation but before the first dose of randomised treatment then this assessment will be used as baseline. For safety endpoints, the last observation before the first dose of IP will be considered the baseline measurement unless otherwise specified. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose.

Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to the first dose if such procedures are required by the CSP to be conducted before the first dose.

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The % change from baseline will be calculated as (post-baseline value - baseline value) / baseline value x 100.

4.1.2 Methods for multiplicity control

The multiple testing procedure, as shown in Figure 2, will define which significance levels should be applied to the interpretation of the raw p-values for the primary endpoint of PFS [redacted]

There will be up to [redacted] DCOs timepoints in the study. The DCO for the final (primary) PFS analysis (first analysis) on the mITT will be done when reaching approximately [redacted] BICR PFS events or approximately [redacted] months follow-up post last participant randomization, whichever occurs first, and the first OS interim analysis on the [redacted] will be conducted at the same time (with approximately [redacted] OS events, [redacted] maturity). The DCO for the second interim OS analysis on the mITT will be performed at approximately [redacted] months after the OS first interim analysis (approximately [redacted] OS events in the mITT with [redacted] maturity), which may be removed as specified in section 1.3. The DCO for the final OS analysis on the mITT will be performed when reaching approximately [redacted] OS events ([redacted] maturity) or approximately [redacted] months follow-up from the last participant randomization, whichever occurs first.

The overall 5% type 1 error will be allocated to the PFS analysis. The 5% alpha level allocated to the OS analyses will be controlled at the interim and final timepoints by using the Lan DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the significance level applied at the interim depends upon the proportion of information available. If [redacted] or [redacted] of the OS events required at the time of the final OS analysis is available at the time of the interim analysis (i.e. [redacted] or [redacted] OS events have occurred), the 2-sided significance level to be applied for the OS interim analysis would be [redacted] and [redacted] respectively, and the 2-sided significance level to be applied for the final OS analysis would be [redacted] Note that the actual allocation of alpha across the three analysis times will be driven by the actual information fraction associated with the analysis and alpha allocation for OS first interim analysis will assume that the OS second interim and final analysis will take place. If second interim analysis is removed after the first interim analysis is complete, the alpha has been spent for first interim analysis will not be changed, the remaining alpha will be calculated based on actual information fraction

and allocated for OS final analysis (i.e. The alpha level allocated to the OS first interim analyses and final analysis would be CCI and CCI respectively if removing the OS second interim analysis).

If the null hypothesis is rejected for either CCI interim or final analyses, 5% alpha level will be further recycled to the CCI analysis. To control the type I error rate, the DCO for the CCI analysis will be fixed at the same time as that of the CCI) analysis and approximately CCI BICR PFS events are expected in the ITT population at this time.

If the null hypothesis is rejected for the CCI analysis, CCI alpha level will be finally recycled to CCI analysis. The CCI alpha level will be controlled at the interim and final analyses by using the Lan DeMets method described as above. At the time of interim and final analyses (same time as the CCI analysis), the number of CCI events is expected to be approximately CCI maturity), and CCI maturity), respectively, which corresponds to an information fraction of approximately CCI and CCI at each analysis, the resulting alpha was calculated to be CCI and CCI at subsequent analyses. The alpha allocation for OS interim and final analysis in the ITT population will follow the same rule and consideration above for the analysis in the mITT population.

4.1.3 Visit window for safety and PRO assessments

Time windows will need defining for any presentations that summarise values by visit. The following conventions should also apply:

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the time window should be based on the actual date and not the intended date of the visit.
- All unscheduled visit data and repeated measurements should have the potential to be included in the summaries. This data should take into account the definition of windows per visits described in the point below.
- The window for the visits following baseline will be 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 will be Day 2). If an even number of days exists between two consecutive visits, then the upper limit will be taken as the midpoint value minus 1 day.

  For example, the visit windows for vital signs data are:

  − Day 29, visit window 2 – 42

  − Day 57, visit window 43 – 70

  − Day 85, visit window 71 – 98

  − Day 113, visit window 99 – 126

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on-treatment will be used (regardless of where it falls in an interval).
- Listings should display all values contributing to a time point for a patient.
- For visit based summaries:
  - If there is more than one value per patient within a time window then the closest value to the scheduled visit date should be used, or the earlier in the event the values are equidistant from the nominal visit date. If there are two values recorded on the same day and the parameter is CTCAE gradable then the record with the highest toxicity grade should be used. Alternatively, if there are two records recorded on the same day and the toxicity grade is the same (or is not calculated for the parameter) then the average of the two records should be used. The listings should highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all on-treatment values collected are used including those collected at unscheduled visits.
  - To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group visit data should only be summarised if the number of observations is greater than the minimum of 20

and > 1/3 of patients dosed (i.e., randomised and received durvalumab/placebo).

- For summaries at a patient level, all values should be included, regardless of whether they appear in a corresponding visit-based summary, when deriving a patient level statistic such as a maximum.
- Baseline will be defined as the last non-missing measurement prior to the first dose of study treatment. For laboratory data with no time collected, any assessments made on day 1 will be considered pre-dose. If there are two visits equally eligible to assess patient status at baseline (e.g., screening and baseline assessments both on the same date prior to first dose with no washout or other intervention in the screening period) with assessment time missing, the average can be taken as a baseline value. For non-numeric laboratory tests (i.e. some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline as this is the most conservative. In the scenario where there are two assessments on day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline.

Where safety data are summarised over time, study day will be calculated in relation to date of first treatment.

4.1.4 Imputation rules

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e., below the lower limit of quantification) or > x (i.e., above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings. Additionally, AEs that have missing causality (after data querying) will be assumed to be related to study drug. Missing CTCAE grades will not be imputed.

Patients with a partial date of birth (i.e. for those countries where year of birth only is given) will have an assumed date of birth of 1st Jan [given year]) for calculation of age at informed consent.

Partial dates for the following modules will be imputed: prior cancer therapy, previous radiotherapy and prior and concomitant medications and adverse events, as well as for other modules, where required.

The original incomplete or missing dates will be presented in the listings.

Adverse events and medications

Adverse events: all AEs will be considered as treatment-emergent unless the opposite can be clearly stated. Imputation will be done only in the context of identifying TEAEs.

Medications/therapies: all medications will be considered as concomitant unless the opposite can be clearly stated.

In practice, for adverse events and medications, original incomplete or missing start dates will be imputed as below:

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

When imputing a start date, ensure that the new imputed date is sensible, i.e., is prior to the end date of the AE or medication.

Original incomplete or missing stop dates for adverse events and medications/therapies will be imputed as below:

- Missing day: impute the last day of the month;
- Missing day and month: impute 31st December. Flags will be retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations would not be calculated.

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

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

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

Duration of AE will not be derived using imputed dates.

4.2 Analysis methods

Efficacy data will be summarised and analysed using the mITT set, and selected analyses repeated on the ITT set.

Table 9 below details which endpoints are to be subjected to formal statistical analysis, together with pre-planned sensitivity analyses, making it clear which analysis is regarded as primary for that endpoint.

Table 9 Study endpoints

Endpoints analysed	Notes

Progression-free survival	Stratified log-rank tests for:
	• Primary analysis using BICR tumour RECIST 1.1 assessments
	• Sensitivity analyses using BICR tumour RECIST 1.1 data
	  • Interval censored analysis – evaluation time bias
	  • Analysis using alternative censoring rules – attrition bias
	  • Analysis by excluding sCRT patients with 1 concurrent chemotherapy and radiation cycle
	  • Analysis accounting for COVID-19 deaths
	• Sensitivity analysis using study site Investigator tumour RECIST 1.1 data – ascertainment bias
	
	Analysis in the presence of non-proportional hazards assumption – Stratified Max-combo test
	
	Additional analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate
	
	Additional analysis using Cox proportional hazards model to determine the consistency of treatment effect between stratification factors via the approach of Gail and Simon 1985.
	
	Subgroup analysis using Cox proportional hazards model

Overall survival	Stratified log-rank test for: • Secondary analysis • Sensitivity analysis using alternative censoring rules – attrition bias • Analysis accounting for COVID-19 deaths Analysis in the presence of non-proportional hazards assumption – Stratified Max-combo test Additional analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate Subgroup analysis using Cox proportional hazards model

	Additional analysis using Cox proportional hazards model to determine the consistency of treatment effect between stratification factors via the approach of Gail and Simon 1985.
Proportion of patients alive at 12 and 24 months from randomisation	KM estimates of survival at 12 and 24 months
Objective response rate	Logistic regression for: • Secondary analysis using BICR tumour RECIST 1.1 data • Sensitivity analysis using study site Investigator tumour RECIST 1.1 data
Duration of response	KM estimates for DoR using BICR tumour data
Progression-free survival at 12 and 18 months from randomisation	KM estimates of progression-free survival at 12 and 18 months
Time from randomisation to second progression[a]	Stratified log-rank test
Time to death or distant metastasis	Stratified log-rank test using site BICR tumour RECIST 1.1 data
Time to first subsequent therapy or death	Stratified log-rank test
Time to second subsequent therapy or death	Stratified log-rank test
Symptom improvement rate (EORTC QLQ-C30 and QLQ-LC13 endpoints	Logistic regression for symptom improvement
HRQoL/Function improvement rate (EORTC QLQ-C30 endpoints)	Logistic regression for HRQoL/Function improvement
Time to HRQoL/Function deterioration (EORTC QLQ-C30 endpoints)	Stratified log-rank test and KM estimate plots
Time to symptom deterioration (EORTC QLQ-C30 and QLQ-LC13 endpoints)	Stratified log-rank test and KM estimate plots
Change from baseline in key symptoms (EORTC QLQ-C30 and QLQ-LC13)	Mixed model repeated measures analysis

BICR Blinded Independent Central Review; EORTC European Organisation for Research and Treatment of Cancer; HRQoL Health-related quality of life; ORR Objective response rate; PFS Progression-free survival; PFS2 Time from randomisation to second progression; QLQ-C30 v3 30-item Core Quality of Life Questionnaire, version 3; QLQ-LC13 13-item Lung Cancer Quality of Life Questionnaire; RECIST 1.1 Response Evaluation Criteria in Solid Tumours, version 1.1.

4.2.1 Primary efficacy endpoint – Progression-free survival

PFS based on the BICR data will be analysed using a stratified log-rank test adjusting for the stratification factors, i.e. level of PD-L1 expression (PD-L1<1% vs PD-L1≥1%) and prior therapy (cCRT or sCRT), for generation of the p-value. CIs for median progression-free survival will be derived based on Brookmeyer-Crowley method with a log-log transformation. The effect of durvalumab versus placebo treatment will be estimated by the HR together with its corresponding 95% CI and p-value. The HR and its CI will be estimated from a stratified Cox proportional hazards model (Cox, 1972) (with ties = Efron and the stratification variables included in the strata statement) and the CI calculated using a profile likelihood approach.

The stratification variables in the statistical modelling will be based on the values entered into Interactive voice response system (IVRS) at randomisation, even if it is subsequently discovered that these values were incorrect. Sensitivity analysis might be performed based on the values in the eCRF page if more than 10% randomised patients have discrepancies in stratification factors between IVRS and eCRF data.

These analyses, later named ‘main analysis’ below, will be repeated on the ITT set.

Supportive summaries/graphs

Kaplan-Meier plot of PFS will be presented by treatment arm, and repeated on the ITT set. Summaries of the number and percentage of patients experiencing a PFS event and the type of event (RECIST 1.1 or death) will be provided along with median PFS for each treatment and associated 95% CI, and repeated on the ITT set.

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

KM curve along with landmark analyses (i.e. 6 months, 12 months, 18 months, 24 months, 36 months PFS rate) will be presented by treatment arm

The treatment status at progression of patients at the time of analysis will be summarised. This will include the number (%) of patients who were on-treatment at the time of progression, the number (%) of patients who discontinued study treatment prior to progression, the number (%) of patients who have not progressed and were on-treatment or discontinued treatment, and the number (%) of patients who progressed prior to receiving treatment. This will also provide distribution of number of days prior to progression for the patients who have discontinued treatment.

The number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if they had not progressed (or died in the absence of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval plus 2 weeks (10 weeks if period between randomisation and DCO for that patient is 48 weeks or less; 14 weeks otherwise) prior to the DCO date.

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

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

Additionally, summary statistics for the number of weeks between the time of progression and the last RECIST 1.1 assessment prior to progression will be presented for each treatment group.

Summaries of the number and percentage of patients who miss two or more consecutive RECIST 1.1 assessments will be presented for each treatment group.

All collected RECIST 1.1 data will be listed. In addition, a summary of new lesions (i.e., sites of new lesions) will be produced.

Sensitivity Analyses

The following sensitivity analyses will be performed:

1. Evaluation-time bias

A sensitivity analysis will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the CSP-scheduled time points. The midpoint between the time of progression and the previous RECIST 1.1 assessment (using the final date of the assessment) will be analysed using a stratified log-rank test, as described for the primary analysis of PFS. Note that midpoint values resulting in non-integer values should be rounded down. For patients whose death was treated as a PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). To support this analysis, the mean of patient-level average inter-assessment times will be tabulated for each treatment. This approach will use the BICR data. This analysis will be repeated on the ITT set.

2. Attrition bias

Attrition bias will be assessed by repeating the PFS main analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following two, or more, missed tumour assessments will be included. In addition, and within the same sensitivity analysis, patients who take subsequent therapy prior to their last RECIST 1.1 assessment or progression or death will be censored at their last assessment prior to taking the subsequent therapy.

An attrition bias sensitivity analysis will also be performed. This analysis will be a repeat of the PFS primary analysis except that the actual PFS times of the patients who missed two or more tumour assessments (rather than censored times) will be used.

The above analyses will be supported by the Kaplan-Meier plot of the time to censoring using the PFS data from the primary analysis and where the censoring indicator of the PFS analysis is reversed. Except the supportive Kaplan-Meier plot, these analyses will be repeated on the ITT set.

3. Ascertainment bias

Ascertainment bias will be assessed by analysing site Investigator data. The programmatically derived PFS using the site Investigator date based upon RECIST 1.1 will be used. The stratified log-rank test, the HR and corresponding 95% CI will be provided as well as other pieces of the PFS main analysis. The number and percentage of patients experiencing a PFS event, and the type of event (RECIST 1.1 or death) along with median PFS for each treatment and associated 95% CI will also be produced. This analysis will be repeated on the ITT set.

If there is an important discrepancy between the primary analysis using the BICR and this sensitivity analysis using site Investigator data assessments, then the proportion of patients

with Investigator but no BICR progression will be summarised; such patients have the potential to introduce bias in the BICR due to informative censoring. An approach that imputes an event at the next visit in the BICR analysis may help inform the most likely HR value (Fleischer et al 2011), but only if an important discrepancy exists.

Disagreements between BICR and investigator reviews of RECIST 1.1 progression will be presented for each treatment group. The summary will include the early discrepancy rate which is the frequency of investigator declared progressions before the BICR review as a proportion of all investigator progressions, and the late discrepancy rate which is the frequency of investigator review declared progressions after the BICR as a proportion of all discrepancies.

A forest plot illustrating the HR and 95% CI will be provided to compare the primary and sensitivity analyses of progression-free survival. This analysis will be repeated on the ITT set.

4. Analysis by excluding sCRT patients with 1 concurrent chemotherapy and radiation cycle

Further sensitivity analysis will be carried out for PFS by excluding sCRT patients with 1 concurrent chemotherapy and radiation cycle using BICR tumour data based upon RECIST 1.1. The stratified log-rank test will be repeated on this data and HR and 95% CI will be presented. The following acceptance thresholds must be followed for patients to be accounted for in the excluded sCRT subpopulation:

- The window for sCRT patients receiving concurrent chemotherapy and radiation cycle is within 1 to 28 days prior to 1st dose of IP.
- Chemotherapy received must be platinum-based
- These patients must have received **at least** 3 cycles of chemotherapy (2 sequential and 1 concurrent) to be included in this excluded subgroup of patients, whereby
    - 2 sequential cycles: patients must have received at least 2 cycles of platinum-based chemotherapy before radiation therapy. The interval between administration of the last dose of the chemotherapy regimen and start of radiotherapy must be no more than 6 weeks.
    - 1 concurrent cycle: patients must have received at least 1 cycle of platinum-based chemoradiation therapy.
- The start date of last cycle of chemotherapy should be on or after the 7 days before start date of radiation therapy

5. The stratified Max-combo test will be conducted as a sensitivity analysis on the primary endpoint (PFS based on BICR data), to test for treatment differences in the case of nonproportional hazards. The analysis will be based on adaptive procedure involving selection of best test statistics with log-rank (G0,0) and the Fleming-Harrington (FH) test (G0, 1, G1, 0, and G1, 1) with alpha correction (Duke-Margolis, 2018).

6.Analysis accounting for COVID-19 deaths

This analysis is a repeat of the primary PFS analysis, but patients who have not progressed prior to death and where the primary or secondary cause of death is due to confirmed/suspected COVID-19 infection, or a confirmed/suspected COVID-19 infection is reported as a fatal AE, are censored at their last evaluable radiological assessment prior to their confirmed/suspected COVID-19 infection death date. This analysis will be performed if sufficient number of patients have such events (≥ 5 and/or ≥ 2% of the patients in the analysis set).

Subgroup analysis / Additional analysis

Subgroup analyses will be conducted comparing PFS (per RECIST 1.1 using BICR assessments) between treatments in the following subgroups (but not limited to):

- Prior therapy (cCRT or sCRT)
- Stage of disease at study entry (IIIA vs. IIIB/C)
- Sex (Female vs. Male)
- Age at randomisation (< 65 vs. ≥ 65 years)
- Type of chemotherapy (cisplatin-based vs carboplatin-based)
- Time from last radiology to randomisation (< 14 days vs. ≥ 14 days)
- PD-L1 status (tumour cells [TCs] ≥ 25% vs TCs < 25%; TCs ≥ 1% vs TCs < 1%. Both sets are high vs. low/negative;)
    - PD-L1 status will be assigned based on the following algorithm:
        - Only sample with sample date on or before first dose should be used
        - Most recent sample “sampled date” should be used, unless this sample is unevaluable, in which case most recent evaluable sample is used
        - For multiple evaluable samples with the same sampled date, the sample with the highest PD-L1 % tumour membrane staining result should be used
- EGFR status (positive vs negative) - only applicable for analyses in the ITT set
- ALK status (positive vs negative) - only applicable for analyses in the ITT set
- EGRF or ALK status (positive vs negative) - only applicable for analyses in the ITT set

- Histology (squamous vs. non-squamous)WHO performance status at baseline (normal activity [PSTAT=0] vs restricted activity [PSTAT=1])
- Smoking (smoker vs. non-smoker [never smoked])
  - Patient is categorised as smoker if there exists a record in SU_NIC with any of the following options for “What type of substance was used?” 'Cigarettes', 'Cigarillos', 'Cigars', 'Pipe Tobacco', 'Tobacco for Smoking'. Else patient is a non-smoker. Non-smoker is a patient that has never smoked.
- Race (White, Black/African-American, Asian, Other [Native Hawaiian/Pacific Islander or American Indian/Alaska Native or Others])
- Race (Asian vs. non-Asian).

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

Unless noted above, patients with unknown or missing level of baseline variables will be excluded from subgroup analyses.

Other baseline variables may also be assessed if there is clinical or biological justification or an imbalance is observed between the treatment arms. The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic factors and/or predictive factors.

No adjustment to the significance level for testing of the subgroup and sensitivity analyses will be made since all these subgroup analyses will be considered exploratory and may only be supportive of the primary analysis of PFS. For each subgroup, the HR (for the treatment comparisons of interest) and 95% CI will be calculated from a Cox proportional hazards model that contains treatment. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties, using the by statement to obtain HR and 95% CI for each subgroup level separately. These will be presented on a forest plot including the HR and 95% CI, along with the results of the overall primary analysis.

If there are too few events available for a meaningful analysis of a particular subgroup (it is not considered appropriate to present analyses where there are less than 20 events [10 events in each subgroup/treatment arm]), the relationship between that subgroup and PFS will not be formally analysed. In this case, only descriptive summaries will be provided.

Unless there is a marked difference between the results of the statistical analyses of the PFS from the BICR tumour data and that of the site Investigator, the subgroup analyses will only be performed upon the PFS endpoint using BICR data.

This analysis will be repeated on the ITT set.

Effect of covariates on the HR estimate

Cox proportional hazards modelling will be employed to assess the effect of pre-specified covariates on the HR estimate. The result from the model in the primary analysis and the model containing additional covariates will be presented.

Additional covariates for this model may include age at randomisation, sex, smoking status, stage of disease at study entry, best response to prior anti-cancer therapy (CR vs PR vs SD), histology, region and race.

The model will include the effect regardless of whether the inclusion of effect significantly improves the fit of the model provided there is enough data to make them meaningful.

Consistency of treatment effect between stratification factors

The presence of quantitative interactions will be assessed by means of an overall global interaction test for stratification factors.

This is performed by comparing the fit of a Cox proportional hazards model including treatment, all stratification variables, and all stratification variable-by treatment interaction terms, with a model that excludes the interaction terms and will be assessed at the 2-sided 10% significance level. Ties will be handled using the Efron approach.

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

Interactions between treatment and stratification factors will also be tested to rule out any qualitative interaction using the approach of Gail and Simon (Gail and Simon 1985).

4.2.2 Secondary efficacy endpoints

4.2.2.1 Overall survival

The key secondary endpoint of OS will be analysed using a stratified log-rank test, using the same methodology as described for the primary PFS endpoint. The effect of durvalumab versus placebo will be estimated by the HR together with its corresponding (1 – alpha adjusted) % CI and p-value. Kaplan-Meier plot of OS will be presented by treatment arm.

The stratification variables in the statistical modelling will be based on the values entered into Interactive voice response system (IVRS) at randomisation, even if it is subsequently discovered that these values were incorrect. Sensitivity analysis might be performed based on the values in the eCRF page if more than 10% randomised patients have discrepancies in stratification factors between IVRS and eCRF data.

Summaries of the number and percentage of patients who have died, those still in survival follow-up, those lost to follow-up, and those who have withdrawn consent will be provided along with the median OS for each treatment.

This analysis will be repeated on the ITT set.

The assumption of proportionality will be assessed in the same way as for PFS.

Overall survival status will also be listed.

Sensitivity analyses

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

Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) vs log (time) and, if these raise concerns, by fitting a time-dependent covariate to assess the extent to which this represents random variation.

The stratified Max-combo test will be conducted as a sensitivity analysis on OS, to test for treatment differences in the case of non-proportional hazards. The same algorithm as specified in section 4.2.1 for PFS will be used. This analysis will be repeated on the ITT set.

The number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if their survival status was not defined at the DCO.

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

- In censored patients who are alive at DCO only: Time from randomisation to date of censoring (date last known to be alive) for each arm.
- In all patients: Time from randomisation to the date of death or to the date of censoring for censored patients, regardless of treatment arm.

A sensitivity analysis will be performed to account for COVID-19 deaths. This analysis is a repeat of the primary OS analysis, but patients who died and the primary or secondary cause of death is due to confirmed/suspected COVID-19 infection, or a confirmed/suspected COVID-19 infection is reported as a fatal AE, are censored at their confirmed/suspected COVID-19 infection death date.

Subgroup analyses maybe performed if there is a sufficient number of OS events (≥ 20 events [≥ 10 events in each subgroup/treatment arm]). Subgroup analyses and a forest plot will be generated comparing OS between treatments in the same way as previously specified for PFS. This analysis will be repeated on the ITT set.

Cox proportional hazards modelling will be employed to assess the effect of pre-specified covariates on the HR estimate. The result from the model with and without additional covariates will be presented as specified for PFS.

Consistency of treatment effect between stratification factors will be analysed. Interactions between treatment and stratification factors will also be tested to rule out any qualitative interaction using the approach of Gail and Simon (Gail and Simon 1985), as previously specified for PFS.

4.2.2.2 Proportion of patients alive at 12 months (OS12) and at 24 months (OS24)

OS24 will be summarised (using the Kaplan-Meier curve) and presented by treatment arm along with 95% CI using the log-log transformation.

OS12 will be defined as the Kaplan-Meier estimate of OS at 12 months. The same analysis as for OS24 will be produced.

These analyses will be repeated on the ITT set.
Other landmarks may be summarised as appropriate.

4.2.2.3 Objective response rate

The ORR will be based on the programmatically derived RECIST 1.1 using the BICR tumour data, without the need for confirmation of the first documented response. The ORR will be

compared between durvalumab versus placebo using logistic regression models adjusting for the stratification factors as in the PFS analysis.

The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour the durvalumab arm) together with its associated profile likelihood 95% CI and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). CCI [redacted]

The covariates in the statistical modelling will be based on the stratification values entered into IVRS at randomisation, even if it is subsequently discovered that these values were incorrect.

Summaries will be produced that present the number and percentage of patients with a tumour response (CR, PR separately and CR/PR combined).

The above analyses, and summaries, for ORR will be presented for the subset of patients with measurable disease at baseline. This analysis will be repeated on the corresponding ITT subset.

Overall visit response data will be listed for all patients.

Sensitivity analyses/summaries

The following sensitivity analyses/summaries will be performed:

1. This logistic regression analysis of ORR, and summaries described above, will be repeated using the results of the programmatically derived PFS using the site Investigator data from all scans based upon RECIST 1.1 as a sensitivity analysis to confirm the results of the primary analysis. This analysis will be presented for the subset of patients with measurable disease at baseline.

2. For each treatment arm, BoR will be summarised by n (%) for each category (CR, PR, SD, PD, NE and NED) as follows:

- For the subset of patients with measurable disease at baseline (without the requirement for confirmed response)
- For all patients (without the requirement for confirmed response)
- For the subset of patients with measurable disease at baseline with the requirement for responses of PR and CR to be confirmed

- For all patients with the requirement for response of PR and CR to be confirmed

No formal statistical analyses are planned for BoR.

4.2.2.4 Duration of response (DoR)

Descriptive data, including Kaplan-Meier estimates, will be provided for the DoR based on BICR assessments according to RECIST 1.1 in responding patients (i.e., median DoR and 95% CIs) by treatment arm, including the associated Kaplan-Meier curves (without any formal comparison of treatment arms or p-value attached). CIs for median duration of response will be derived based on Brookmeyer-Crowley method with log-log transformation. Those patients with who do not have an objective response recorded by BICR will be censored. The summary statistics table will be performed on patients with objective response, with and without the requirement for the first documented response to be confirmed. Kaplan-Meier curve will only be repeated for DoR without the requirement for the first documented response to be subsequently confirmed. These analyses will be repeated on the ITT set.

Descriptive data will be provided for TTR without the need for confirmation of response for patients included.

The TTR will be summarised (i.e. number of patients [%] based upon the number of responders for each cohort) by the scheduled assessment timepoint that the response was first observed. Additionally, descriptive summary statistics (i.e. minimum, maximum, median, Q1 and Q3) will also be presented. This analysis will be repeated on the ITT set.

Duration of response data will also be listed.

4.2.2.5 Progression-free survival at 12 and 18 months

PFS12 and PFS18 summaries and analyses will be produced as for OS24 detailed in section 4.2.2.2.

These analyses will be repeated on the ITT set.

Other landmarks may be summarised as appropriate.

4.2.2.6 Time to death or distant metastasis (TTDM)

TTDM will be analysed using identical methods (stratified log-rank test) as outlined for the analysis of PFS and adjusting for the same set of covariates, but no subgroup analysis will be performed. Median and KM plot will be presented to support the analysis. This analysis will be repeated on the ITT set.

The sensitivity analyses outlined in section 4.2.1 will not be repeated for TTDM, except for a KM plot of the time to censoring where the censoring indicator of TTDM is reversed. TTDM will be based on BICR data.

4.2.2.7 Time from randomisation to second progression (PFS2)

PFS2 will be analysed using a stratified log-rank tests, using the same methodology as described for the primary PFS endpoint. The effect of durvalumab versus placebo will be estimated by the HR together with its corresponding 95% CI and p-value. KM plot will be presented by treatment arm. Summaries of the number and percentage of patients who have an event as well as those who were censored will be provided along with the medians for each treatment.

The number and percentage of patients experiencing a PFS2 event and the type of progression (Symptomatic progression, Objective radiological progression, Other) will also be summarised by treatment arm.

This analysis will be repeated on the ITT set.

4.2.2.8 Change in Tumour Size

Descriptive summaries of change in tumour size will also be presented for target lesion size and percentage change from baseline in target lesion size across all time points. Further, descriptive summaries of best percentage change from baseline in target lesion size will be presented, alongside a waterfall plot. Best change in target lesion size is defined as the maximum reduction of lesion size from baseline, or the minimum increase of lesion size from baseline in the absence of a reduction.

4.2.3 Patient-reported outcomes

Compliance rate

Summaries of the number of expected, received and evaluated questionnaires, as well as the compliance and evaluability rates will be presented by treatment arm at each visit (including baseline). The overall number of expected, received and evaluated questionnaires, as well as the compliance and evaluability rates over all time points will also be presented.

Symptom and HRQoL/function improvement rate

A summary of the symptom improvement rate for all symptom scales/items in EORTC QLQ-C30 and QLQ-LC13 will be produced. Similarly, a summary of function/HRQoL improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/QoL will be produced.

Symptom and HRQoL/function improvement will be analysed by comparing the treatment arms using a logistic regression model adjusting for the stratification factors. The odds ratio (an odds ratio greater than 1 will favour durvalumab) together with its associated profile likelihood 95% CI and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model) will be obtained for each scale/item using the same method as done for ORR and will be presented graphically on a forest plot. If there are very few responses in one treatment arm, a Fisher's exact test will be considered.

Time to symptom and HRQoL/function deterioration

Time to symptom and function/HRQoL deterioration will be analysed for each of the symptom scales/items, function scales, and global health status/QoL in EORTC QLQ-C30 and QLQ-LC13. This will be achieved by comparing between treatment arms using a stratified log-rank test as described for the primary analysis of PFS. The HR and 95% CI for each scale/item will be presented graphically on a forest plot.

For each of the symptom scales/items, functional scales, and global health status/QoL, time to deterioration will be presented using a Kaplan-Meier plot. Summaries of the number and percentage of patients experiencing a clinically meaningful deterioration or death and the median time to deterioration will also be provided for each treatment arm.

Time to symptom and HRQoL/function deterioration will also be listed.

Change from baseline in key symptoms, HRQoL and functioning

Change from baseline in the 5 pre-specified symptoms, the physical functioning score and the overall health status score will be analysed using a linear mixed model for repeated measures (MMRM) analysis making use of all data from baseline up to 12 months. MMRM analysis will be of change from baseline in the scores for each assessment timepoint and the Bonferroni-Holm procedure (Holm 1979) for adjusting the significance level will be used to aid interpretation. The analysis will compare the average treatment effect from the point of randomisation until 12 months unless there is excessive missing data (defined as >75% missing data). The 5 key endpoints will be tested at a 1% significance level. Physical functioning score and overall health status score will also be analysed using similar methods, except they will be tested at the 5% significance level.

The MMRM model will include the fixed categorical effects of treatment, visit, and treatment-by-visit interaction, the level of PD-L1 expression (PD-L1<1% or PD-L1≥1%) and prior therapy (cCRT or sCRT) as well as the continuous fixed covariates of baseline score and baseline score-by-visit interaction. Restricted maximum likelihood (REML) estimation will be

used. An overall adjusted mean estimate will be derived that will estimate the average treatment effect over visits giving each visit equal weight. For this overall treatment comparison, adjusted mean estimates per treatment group and corresponding 95% CIs will be presented along with an estimate of the treatment difference, 95% CI, 99% CI (for the 5 key symptoms only) and p-value. An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, autoregressive and compound symmetry.

Multiple imputation techniques for missing values may be considered to explore the robustness of any treatment effect.

It is acknowledged that patients will discontinue treatment at different timepoints during the study and that this is an important time with regards to symptoms and HRQoL data collection. To account for this, and to include the discontinuation and follow up visits, a generic visit variable will be derived for each patient in order that the average treatment effect can be analysed using the above method. Each visit will be assigned a sequential number. The time from randomisation to each of these will be derived to select only those visits occurring within the first 12 months of randomisation.

As an example, say a patient X attends the first 4 scheduled visits of a 4-weekly schedule and then discontinues treatment, whilst patient Y discontinues treatment after the first scheduled visit, the first 6 generic visits would be as in Table 10 below.

Table 10 Generic visit variable derivation example

Generic visit	**Study day (week)**	
	Patient X	**Patient Y**
Baseline	Baseline	Baseline
1	29	29
2	57	54 (discontinuation)
3	85	85
4	113	113
5	138 (discontinuation)	141
6	169	169

Summaries of original and change from baseline values of each symptom scale/item, the global HRQoL score, and each functional domain will be reported by assessment timepoint for each treatment arm. Graphical presentations may also be produced as appropriate.

Summaries of the number and percentage of patients in each response category at each assessment timepoint for each ordinal item (in terms of the proportion of patients in the categories of Improved, Stable, and Worsened as defined in Table 7) will also be produced for each treatment arm. For each of the visit level summaries of Improved/Worsened/Stable, all patients with a baseline and post-baseline score will be included, thus the denominator may differ from the time to worsening and improvement rate endpoints.

All EORTC QLQ-C30 and QLQ-LC13 raw scores and scale scores will be listed.

All PRO analyses described above below will be repeated on the ITT set.

4.2.4 Other efficacy outcome endpoints

4.2.4.1 Time to first subsequent therapy or death (TFST)

The time to the start of subsequent therapy will be analysed using the same methodology and model as that used for the primary analysis of PFS. The HR for the treatment effect together with its 95% CI will be presented. In addition, medians and a Kaplan-Meier plot of the time to the start of subsequent therapy will be presented by treatment arm and the time between progression and starting subsequent therapy will be assessed based upon the investigator data defined date of progression. This will be summarised per treatment arm but no formal comparisons will be made. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

In patients who received a subsequent anti-cancer therapy, a summary table of first subsequent anti-cancer therapies by treatment arm will be provided.

This analysis will be repeated on the ITT set.

4.2.4.2 Time to second subsequent therapy or death (TSST)

The time to second subsequent therapy will be analysed using similar methodology to that of TFST. This analysis will be repeated on the ITT set.

4.2.5 Safety data

Safety is a secondary objective. Safety and tolerability data will be presented by treatment arm using the safety population. Safety summaries will be descriptive only. No formal statistical analyses will be performed on the safety variables.

The following sections describe the planned safety summaries for AEs, vital signs, laboratory parameters, ECG and World Health Organisation (WHO)/ Eastern Cooperative Oncology Group (ECOG) performance status. However, additional safety summaries (not specified in

this statistical analysis plan [SAP]) may need to be produced to aid interpretation of the safety data.

4.2.5.1 **Adverse events**

All AEs, both in terms of current MedDRA PT and CTCAE grade, will be summarised descriptively by count (n) and percentage (%) for each treatment group. Any AE occurring before randomised treatment (i.e. before the administration of the first infusion on Study Day 1) will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'. However, any AE occurring before the administration of the first dose on Study Day 1 that increases in severity after the first dose will be regarded as treatment emergent and thus will be included in the majority of summary tables.

A treatment emergent AE (TEAE) is an AE with an onset date (and time, if applicable) following the first dose of study treatment, or a pre-existing AE which worsens following the first dose of study treatment through to 90 days after the last dose of study treatment (i.e., the last dose of durvalumab/placebo).

AEs observed up until 90 days following discontinuation of the study treatment (i.e., the last dose of randomised treatment) or until the initiation of the first subsequent anti-cancer therapy (excluding palliative radiotherapy) following discontinuation of study treatment (whichever occurs first) will be used for reporting of all the AE summary tables. For all IDMC and interim analyses (IAs) deliveries, this will take into account the earliest date from: AEs observed up until 90 days following discontinuation of the study treatment, initiation date of the first subsequent anti-cancer therapy (excluding palliative radiotherapy) or the data cut-off (DCO) date applicable to that delivery.

This will more accurately depict AEs attributable to study treatment only as some of AEs up to 90 days following discontinuation of the study treatment are likely to be attributable to subsequent therapy.

However, to assess the longer-term toxicity profile, some of the AE summaries will also be produced containing AEs observed up until 90 days following discontinuation of the study treatment (i.e. without taking subsequent therapy into account)

Any events in this period that occur after a patient has received further therapy for cancer (following discontinuation of study treatment) will be flagged in the data listings.

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved) and investigator's assessment of severity and relationship to durvalumab/placebo/

CRT. Frequencies and percentages of patients reporting each PT will be presented (i.e. multiple events per patient will not be accounted).

AEs that have missing causality (after data querying) will be assumed to be related to study drug.

A separate data listing of AEs occurring more than 90 days after discontinuation of IP will be produced. These events will not be included in AE summaries, unless otherwise stated.

Any AE occurring before treatment with IP will be included in the data listings but will not be included in the summary tables of AEs, unless otherwise stated.

Summary information (the number and percent of patients by system organ class (SOC) and PT separated by treatment group) will be tabulated for:

- All AEs (at patient level, and by event rate)
- All AEs causally related to durvalumab/placebo (as determined by the reporting investigator)
- AEs by maximum reported CTCAE grade
- AEs with CTCAE grade 3 or 4
- AEs with CTCAE grade 3 or 4, causally related to durvalumab/placebo (as determined by the reporting investigator)
- AEs with maximum CTCAE grade 3 or 4
- AEs with maximum CTCAE grade 3 or 4, causally related to durvalumab/placebo (as determined by the reporting investigator)
- Most common AEs
- AEs with outcome of death
- AEs with outcome of death causally related to durvalumab/placebo (as determined by the reporting investigator)
- All SAEs
- All SAEs causally related to durvalumab/placebo (as determined by the reporting investigator)
- AEs leading to discontinuation of durvalumab/placebo
- AEs leading to discontinuation of durvalumab/placebo, causally related to durvalumab/placebo (as determined by the reporting investigator)
- Other significant AEs
- Most common AEs with CTCAE grade 3 or higher
- AEs leading to dose delay of durvalumab/placebo

- AEs from start of first subsequent therapy until 90 days following discontinuation of study treatment
- AEs which started prior to first dose or after 90 days following date of last dose

An overall summary of the number and percentage of patients in each category will be presented. A truncated AE table of most common AEs and another table showing most common AEs with CTCAE grade 3 or 4, showing all events that occur in at least 5% of patients overall will be summarised by PT, by decreasing frequency in the total column (the total column will not be displayed in the AE tables). This cut-off may be modified after review of the data. When applying a cut-off (i.e., x %), the raw percentage should be compared to the cut-off, no rounding should be applied first (i.e., an AE with frequency of 4.9% will not appear if a cut-off is 5%). Summary statistics showing the time to onset and the duration of the first AE may also be presented as appropriate.

Each AE event rate (per 100 patient years) will also be summarised by PT within each SOC for the output summarising all AEs. For each PT, the event rate is defined as the number of patients with that AE divided by the total treatment duration (days) of randomised treatment summed over patients and then multiplied by 365.25 x 100 to present in terms of per 100 patient years.

Summaries of the number and percentage of patients with AEs will be provided by maximum reported CTCAE grade, SOC, PT and treatment group.

Fluctuations observed in CTCAE grades during study will be listed for those AEs which are CTCAE ≥ 3.

In addition, all AEs will be listed.

Deaths

Two summaries of all deaths will be provided with number and percentage of patients by treatment group, categorised as:

- Total number of deaths (regardless of date of death)
- Death related to disease under investigation only as determined by the investigator
- Death related to disease under investigation and an AE with outcome of death

- AE onset prior to subsequent therapy. Which includes AEs with an onset date (or pre-treatment AEs that increase in severity) on or after the date of first dose and up to and including 90 days following the last dose of study medication

(durvalumab/placebo), or AE start date <= the date of initiation of the first subsequent therapy (whichever occurs first).

- AE with outcome of death only
  - AE onset prior to subsequent therapy. Which includes AEs with an onset date (or pre-treatment AEs that increase in severity) on or after the date of first dose and up to and including 90 days following the last dose of study medication (durvalumab/placebo), or AE start date <= the date of initiation of the first subsequent therapy (whichever occurs first).

- Death after end of safety follow up period (last dose of study medication (durvalumab/placebo) + 90 days) and not due to disease under investigation
- COVID-19 related death
- Unknown reason for death
- Other deaths

This summary will be repeated for all deaths on-treatment or within 90 days of last dose of durvalumab/placebo.

Adverse events of special interest and possible interest

Preferred terms used to identify adverse events of special interest and possible interest (as defined in section 3.3.1) will be listed before DBL and documented in the Study Master File. Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI/AEPI grouping. Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

An overall summary will include the number and percentage of patients with:

- Any AESI/AEPI (*)
- Any AESI/AEPI, causally related to durvalumab/placebo (*)
- Any AESI/AEPI of CTCAE grade 3 or 4
- Any AESI/AEPI of CTCAE grade 3 or 4, causally related to durvalumab/placebo
- Any AESI/AEPI with outcome of death
- Any AESI/AEPI with outcome of death, causally related to durvalumab/placebo

- Any serious AESI/AEPI (*)
- Any serious AESI/AEPI, causally related to durvalumab/placebo
- Any AESI/AEPI leading to discontinuation of durvalumab/placebo (*)
- Any AESI/AEPI leading to discontinuation of durvalumab/placebo, causally related to durvalumab/placebo
- Any AESI/AEPI leading to dose delay of durvalumab/placebo
- Any AESI/AEPI leading to concomitant medication (*)
- Any AESI/AEPI leading to concomitant medication use (steroids) (*)
- Any AESI/AEPI leading to concomitant medication use (high dose steroids) (*)
- Any AESI/AEPI leading to concomitant medication use (endocrine therapy) (*)
- Any AESI/AEPI leading to concomitant medication use (other immunosuppressants) (*)

Summaries of AESI/AEPI will also be produced, presenting the number and percentage of patients by group term and preferred term, in each of the selected categories above (*) and for patients with

- Any AESI/AEPI by maximum CTCAE grade
- Any AESI/AEPI causally related to durvalumab/placebo, by maximum CTCAE grade

AESI/AEPI will also be presented by outcome.

Another summary, presented by group term and type (AESI, AEPI, AESI or AEPI), will be tabulated presenting the number and percentage of patients with:

- Any AESI/AEPI
- Any serious AESI/AEPI
- Any AESI/AEPI of CTCAE grade 3 or 4
- Any AESI/AEPI causally related to durvalumab/placebo
- Any AESI/AEPI of CTCAE grade 3 or 4 causally related to durvalumab/placebo

- Any AESPI/AEPI leading to systemic corticosteroid use
- Any AESPI/AEPI leading to high dose steroid use
- Any AESPI/AEPI leading to use of other immunosuppressants
- Any AESI/AEPI leading to discontinuation of durvalumab/placebo
- Any AESI/AEPI with outcome of death
- Any AESI/AEPI not resolved
- Any AESI/AEPI resolved

A listing of AESI/AEPI will be provided, as well as a listing for those AESI/AEPI with outcome of death.

Immune-mediated Adverse events (imAEs)

The imAEs (as classified by the Sponsor) will be summarised in the similar manner as for the summaries for AESI/AEPI described above. See further details in the imAE Charter with respect to derivation rules.

4.2.5.2 Laboratory assessments

Data obtained up until the 90 days following discontinuation of study treatment (durvalumab/placebo) or until the initiation of the first subsequent anti-cancer therapy (excluding palliative radiotherapy) following discontinuation of study treatment (whichever occurs first) will be used for reporting. This will more accurately depict laboratory toxicities attributable to study treatment only as some toxicities up to 90 days following discontinuation of the study treatment are likely to be attributable to subsequent therapy.

However, to assess the longer term toxicity profile, some summaries of laboratory data will be produced containing data collected up until 90 days following discontinuation of the study treatment (i.e., without taking subsequent therapy into account).

Any data post 90 days after the last dose of the study treatment will not be summarised.

Data summaries will be provided in preferred units.

Scatter plots (shift plots) of baseline to maximum value/minimum value (as appropriate) on-treatment may be produced for certain parameters if warranted after data review.

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

Shift tables for laboratory values by worst CTCAE grade will be produced, and for specific parameters separate shift tables indicating hyper- and hypo- directionality of change will be produced. The laboratory parameters for which CTCAE grade shift outputs will be produced are:

- Haematology: Haemoglobin; Leukocytes; Lymphocytes (count, absolute); Neutrophils (count, absolute); Platelets
- Clinical chemistry: alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), Total Bilirubin, Albumin, Magnesium – low and – high, Sodium – low and – high, Potassium – low and – high, Corrected Calcium – low and – high, Glucose – low and – high, Creatinine, Lipase and Amylase.

Additional summaries will include a shift table for urinalysis (Bilirubin, Blood, Glucose, Ketones, Protein) comparing baseline value to maximum on-treatment value if a sufficient number of urinalysis assessments are recorded.

Haematology, clinical chemistry and urinalysis laboratory data will be listed.

Liver enzyme elevations and Hy's law

A summary will include the number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study
  - ALT ≥ 3x –≤ 5x, > 5x – ≤8x, > 8x - ≤ 10x, >10x - ≤ 20x, and >20x upper limit of normal (ULN) during the study
  - AST ≥ 3x–≤ 5x, > 5x – ≤8x, > 8x - ≤ 10x, >10x - ≤ 20x, and >20x ULN during the study
  - Total bilirubin ≥2x-≤3x, >3x-≤5x, >5x ULN during the study
  - ALT or AST ≥3x-≤5x, >5x - ≤8x, >8x - ≤ 10x, >10x - ≤ 20x, >20x ULN during the study.
  - ALT or AST ≥3x ULN and Total bilirubin ≥2x ULN during the study (Potential Hy's law): The onset date of ALT or AST elevation should be prior to or on the date of Total Bilirubin elevation

This summary will be repeated based on laboratory assessments with a date ≤ 90 days after last dose.

Liver biochemistry test results over time for patients with elevated ALT or AST (i.e. ≥ 3x ULN), and elevated Total Bilirubin (i.e. ≥ 2x ULN) will be plotted (the onset date of ALT or AST elevation should be prior to or on the date of Total Bilirubin elevation). Individual patient data where ALT or AST (i.e. ≥ 3x ULN) plus Total Bilirubin (i.e. ≥ 2x ULN) are elevated will be listed also (the onset date of ALT or AST elevation should be prior to or on the date of Total Bilirubin elevation).

Plots of ALT and AST vs. Total Bilirubin by treatment group will also be produced with reference lines at 3×ULN for ALT, AST, and 2×ULN for Total Bilirubin. In each plot, Total Bilirubin will be on the vertical axis.

Liver diagnostic investigations, risk factors and lifestyle events as well as liver signs and symptoms will be listed.

Assessment of Thyroid Function Test Results

The following summaries will include the number and percentage of patients who have elevated or low TSH.

- On-treatment elevated TSH > ULN
- On-treatment elevated TSH > ULN with TSH ≤ ULN at baseline
- On-treatment elevated TSH > 3 X ULN
- On-treatment elevated TSH > 3 X ULN with TSH ≤ ULN at baseline
- On-treatment elevated TSH > 10 X ULN
- On-treatment elevated TSH > 10 X ULN with TSH ≤ ULN at baseline;
- On-treatment low TSH < LLN
- On-treatment low TSH < LLN with TSH ≥ LLN at baseline

A separate summary will present:

- Number of patients with at least one baseline and post-baseline TSH result
    - On-treatment elevated TSH > ULN and above baseline
    - On-treatment decreased TSH < LLN and below baseline

Thyroid function test laboratory data will be listed.

Assessment of Renal Function Test Abnormalities

In addition to the analysis for serum creatinine, the number and percentage of patients with CrCl rate during treatment period meeting the following categories will be presented:

- Normal: CrCl ≥ 90 mL/min
- Mild Impairment: CrCl ≥ 60 - < 90 mL/min
- Moderate Impairment: CrCl ≥ 30 - < 60 mL/min
- Severe Impairment: CrCl ≥ 15 - < 30 mL/min
- Kidney Failure: CrCl < 15 mL/min

A summary of the reversibility of CrCl will be provided as the number and percentage of patients who shifted to a worse on-treatment renal impairment category from baseline and the number and percentage of patients whose worsened renal impairment was reversible and transient, defined as subsequent CrCl value that is higher than the worst on-treatment CrCl value and in a better impairment category.

CrCl rate will be calculated using serum Creatinine and the Cockcroft-Gault formula (Cockcroft and Gault 1976).

4.2.5.3 Vital signs

Box plots for absolute values and change from baseline by week may be presented for certain vital signs parameters if warranted after data review.

Vital signs (systolic blood pressure [SBP], diastolic blood pressure [DBP], pulse rate, temperature, respiratory rate and weight) will be summarised over time in terms of absolute values and change from baseline at each scheduled measurement by actual treatment group.

Vital signs data will also be listed.

4.2.5.4 ECGs

ECG evaluations, categorised as detailed in section 3.3.3, will be summarised using a shift table of baseline evaluation to worst evaluation on-treatment if a sufficient number of ECG assessments are recorded.

ECG data and ECG abnormalities will also be listed.

4.2.5.5 Physical examination

Individual physical examination data will not be summarised or listed.

4.2.5.6 Other safety data

Data from positive pregnancy tests will not be summarised. This data will be listed only.

4.2.6 Other exploratory endpoints

4.2.6.1 EuroQol-5-Dimension 5-Level questionnaire (EQ-5D-5L)

The change from baseline in health state utility values and the visual analogue scale (VAS) will be compared between treatment arms at each visit using a MMRM analysis as detailed for the EORTC QLQ-C30 and QLQ-LC13 change from baseline in key symptoms endpoint above, which adjusts for the same factors as the primary analysis and the baseline health state utility value/VAS as appropriate. Adjusted mean differences between treatments and 95% CIs from these analyses will be presented, but, as this analysis is exploratory in nature, p-values will not be calculated.

Descriptive statistics will be reported for the health state utility index score and the VAS index score by visit, as well as the change from baseline of these scores. To support future economic evaluations, additional appropriate analyses may be undertaken, for example, mean health state utility pre- and post-treatment and pre- and post-progression.

A summary of EQ-5D compliance rates will also be presented by visit and overall using the same derivations and analysis methods as the analysis of compliance rates for EORTC QLQ-C30 and QLQ-LC13 above.

EQ-5D-5L utility and VAS scores will be listed.

4.2.6.2 WHO/ECOG performance status

All WHO/ECOG performance status data will be summarised over time for the mITT set, and will also be listed.

4.2.6.3 Health care resource use

An exploratory health economic analysis of hospital episodes including type of contact (hospitalisation, outpatient, or day case), reason, length of stay by ward type (including intensive care unit), procedures, and tests may be undertaken to examine the impact of disease and treatment on health care resource use to primarily support the economic evaluation of durvalumab for the purposes of submissions to payers. To support submissions to payers, additional analyses (including appropriate descriptive statistics such as means, median and ranges) may be undertaken and these will be outlined in a separate Payer Analysis Plan.

Health care resource use data will also be listed.

4.2.7 Pharmacokinetic analysis

Durvalumab concentration data will be listed, and a summary will be provided for all evaluable patients, also by ADA category along with a figure for individual concentration

data. Data collected in this study may be utilised to perform population PK analysis; if this is produced, it will be presented in a separate population PK report.

4.2.8 Immunogenicity analysis

Immunogenicity results of all patients will be listed. The number and percentage of patients who develop detectable ADA to MEDI4736 within each ADA response category listed in section 3.4.2 will be summarised based on the ADA analysis set. The immunogenicity titre and neutralising ADA data will be listed for samples confirmed positive for the presence of anti-MEDI4736 antibodies. Details for the presentation and derivation of ADA data is provided in section 3.4.2.

The effect of immunogenicity on PK, efficacy and safety will be evaluated if data allow.

4.2.9 Biomarkers analysis

The relationship between PD-L1 and clinical outcomes PFS, OS, ORR, DoR and PFS2 will be presented; plans for this analysis are detailed in the clinical outcomes sections 4.2.1 (PFS), 4.2.2.1 (OS), 4.2.2.3 (ORR), 4.2.2.4 (DoR) and 4.2.2.7 (PFS2).

4.2.10 PK/PD relationships

If the data are suitable, the relationship between durvalumab PK exposure and efficacy/safety parameters may be investigated graphically or using appropriate data modelling approach, which will be presented in a separate exposure-response report.

4.2.11 Demographic and baseline characteristics data

Unless otherwise specified, the following will be summarised for all patients in the mITT set by treatment group, and repeated for ITT set:

- Patient disposition (All patients only)
- Important protocol deviations
- Inclusion in analysis populations
- Demographics (age, age group [< 50, ≥ 50 - < 65, ≥ 65 - < 75 years and additionally ≥ 75 years], sex, race and ethnicity). Age will be derived from the date of informed consent
- Patient characteristics at baseline (height, weight, weight group [< 70, 70 to 90 and > 90 kg], body mass index (BMI) and BMI group)
- Patient recruitment by region, country and centre
- Prior radiotherapy and chemotherapy

- Time from last dose of radiation to randomisation, and time from last dose of radiation to randomisation group (< 14 days vs. ≥ 14 days)
- Disease characteristics at baseline (WHO performance status, primary tumour location, histology type, AJCC staging and overall disease classification, best response to previous therapy, EGFR status, ALK status and combined EGFR/ALK status)
- Extent of disease at study entry
- TNM classification at baseline
- Medical history (past and current)
- Relevant surgical history (as appropriate)
- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer therapy
- Nicotine use, categorised (never, current, former)
- Stratification factors (level of PD-L1 expression, prior therapy) as per IVRS and eCRF data. In addition, cross tabulation will be provided for prior therapy per IVRS with eCRF data.

Information on EGRF/ALK status will be based on eCRF data. The combined EGRF/ALK status (i.e. "EGRF mutation or ALK rearrangement") will be classified as 'positive' if positive for either EGFR mutation or ALK rearrangement, 'negative' if both are negative, 'unknown'/'missing' if both are unknown/missing. In case of multiple sources, i.e. records from both local and central laboratories, data from the central laboratory will be used. These descriptions of EGRF and/or ALK status are only applicable on the ITT set.

The demographic characteristics at baseline and ECOG performance along with past medical history and ongoing conditions will be summarised additionally in patients with confirmed or suspected COVID-19 infection.

The patient disposition table will be repeated on the mITT population without counts of patients enrolled and screen failure description.

Demographic and baseline characteristics will also be listed.

4.2.12 Treatment exposure and intensity

The following summaries related to study treatment will be produced for the safety analysis set by treatment group:

- Total exposure
- Actual exposure (durvalumab or matching placebo only)

- Reasons for dose delays/interruptions. Dose interruptions will be based on investigator initiated dosing decisions
- Number of infusions received
- Relative dose intensity (RDI) (durvalumab or matching placebo only)
- Cumulative exposure over time

For patients on study treatment at the time of the IA, the DCO date will be used to calculate exposure.

4.2.13 Impact of COVID-19

Depending on the extent of any impact, summaries of data relating to impact of COVID-19 on study conduct may be generated, including:

- Disposition (discontinued IP due to COVID-19 and withdrew study due to COVID-19)
- Deviations (overall deviations plus if due to COVID-19 and not due to COVID-19)
- Summary of COVID-19 disruption (visit impacted, drug impacted)
- Summary of confirmed/suspected COVID-19 infections for all patients and for patients who died.
- PFS sensitivity analysis (see more details in section 4.2.1)
- OS sensitivity analysis (see more details in section 4.2.2.1)
- COVID-19 related death (see more details in section 4.2.5.1)
- a repeat of the main AE summary table, if more than 20 patients

These confirmed/suspected AEs will also be listed.

4.2.14 China Cohort Analyses

Per China National Medical Products Administration's (NMPA) guidance, in addition to the evaluation of the global cohort data for primary, secondary and safety objectives, evaluation of consistency in efficacy and safety between the China subgroup and the overall global cohort is required to facilitate the benefit-risk assessment for China patients.

The efficacy and safety data for the China cohort are summarised and analysed using the same methodology as described above for the overall global population.

All statistical analyses in the China cohort will be considered as exploratory, and therefore no alpha is assigned to the efficacy assessment for this cohort.

5 INTERIM ANALYSES

5.1 Analysis methods

Interim safety monitoring will be conducted by an IDMC. The recommendations from the IDMC will not reveal the results of the analysis but will take the form of 'Continue/Modify/Stop'. Interim analyses will be performed for efficacy as described below. The OS interim analysis will be conducted by study team and will not be reviewed by IDMC as the study will be unblinded at PFS final (primary) analysis.

5.1.1 OS interim analyses

Up to two OS IAs will be performed for superiority, in the mITT set. The DCO for the first IA of OS will occur at the same time as final (primary) PFS analysis (estimated approximately CCI OS events in the mITT set at the time of the analysis). The second IA of OS will be conducted at approximately CCI months after the first OS interim analysis (approximately CCI OS events in the mITT with CCI maturity), which may be removed as specified in section 1.3.

The alpha level allocated to OS will depend on the results of the PFS analysis (see section 4.1.2). It will be controlled at the interims and final analysis by using the Lan DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the significance level applied at the IA depends upon the proportion of information available.

The significance level for the OS analyses will be calculated using the statistical software package EAST by specifying the information fraction for each analysis. The information fraction is calculated as the number of OS events at the analysis time-point divided by the total number of events at the final analysis time-point. For example, if the alpha level available for OS is 5%, and if CCI or CCI of OS events required at the time of the final OS analysis are available at the time of the interim analysis (i.e., CCI or CCI OS events have occurred), the 2-sided significance level to be applied for the OS IA would be CCI and CCI respectively, and the 2-sided significance level to be applied for the final OS analysis would be CCI.

5.1.2 Independent Data Monitoring Committee

An IDMC composed of independent experts will be convened to confirm the safety and tolerability of the proposed dose and schedule of durvalumab. The first safety review will take place approximately 6 months after FSI and then periodically in accordance with the IDMC charter to continue safety monitoring until final PFS analysis. Additional reviews of the safety data may be requested by the IDMC at additional points during the study.

Safety reviews will be carried out by the IDMC in an unblinded manner. After review of the unblinded data, the IDMC will make a recommendation on whether the study should continue recruitment as planned or hold recruitment.

Full details of the IDMC procedures, processes can be found in the IDMC Charter.

6 CHANGES OF ANALYSIS FROM PROTOCOL

Time to deterioration in patient-reported symptoms, functioning and global health status/QoL was added as secondary endpoint considering relevant analysis for this endpoint was described in the CSP.

For the definition/derivation of TTDM, distant metastasis is defined as any new lesion that is outside of the radiation field according to RECIST 1.1 only. New lesion proven by biopsy will not be used for distant metastasis identification.

7 REFERENCES

Aaronson et al 1993
Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85(5):365-76.

Breslow, 1974
Breslow, NE. Covariance Analysis of Censored Survival Data. Biometrics 1974; 30:89–99.

Bergman et al 1994
Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. Eur J Cancer 1994;30A(5):635-42.

Burman et al 2009
Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med 2009 Feb 28;28(5):739-61.

Cockcroft and Gault 1976
Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron 1976; 16(1):31-41.

Duke-Margolis 2018
Public Workshop: Oncology Clinical Trials in the Presence of Non-Proportional Hazards. The Duke Margolis Center for Health Policy Feb.2018.

EORTC QLQ-C30 Scoring Manual, Third Edition
EORTC QLQ-C30 Scoring Manual, Third Edition. 2001, EORTC, Brussels.

EuroQoL 2013
EuroQol Group. EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument, Version 2.0, October 2013. Available from: https://euroqol.org/wp-content/uploads/2021/01/EQ-5D-5LUserguide-08-0421.pdf

Fayers et al 2001
Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A; EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer; 2001.

Fleischer et al 2011
Fleischer F, Gaschler-Markefski B, Bluhmki E. How is retrospective independent review influenced by investigator-introduced informative censoring: a quantitative approach. Stat Med 2011 Dec 20;30(29):3373-86.

Gail and Simon 1985
Gail M, Simon R. Tests for qualitative interactions between treatment effects and patient subsets. Biometrics 1985 Jun;41(2):361–72.

Holm 1979
Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979; 6:65-70.

Klein et al 2007
Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. Stat Med 2007 Oct 30; 26(24): 4505–19.

Lan and DeMets 1983
Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659-63.

Oemar and Janseen 2013
Oemar M, Oppe M. EQ-5D-3L User Guide: Basic information on how to use the EQ-5D-3L instrument V5.0 (October 2013). Available from URL: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-3L_UserGuide_2013_v5.0_October_2013.pdf. Accessed 07 January 2014.

Osoba et al 1998
Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998 Jan;16(1):139-44.

Sun and Chen 2010
Sun X, Chen C. Comparison of Finkelstein's Method with the Conventional Approach for Interval-Censored Data Analysis. Stat Biopharm Res 2010;2(1):97-108.

Whitehead and Whitehead 1991
Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomised clinical trials. Statistics In Medicine 1991;10:1665-77.

8 APPENDIX

APPENDIX A. 2 MISSED VISITS

ePRO Assessments

As per Section 3.2.3.1, confirmation of deterioration/worsening will first be determined, then the censoring rules will be applied. For ePRO assessments, we will apply a 'look-forward' approach to the 2 missed visit rule.

For example:

Baseline	V1	V2	V3	V4	V5	V6	V7	V8	Observed confirmed deterioration visit*	Time to Deterioration derivation
	NC	DT	DT	NC	MV	NC	IMP	NC	Event at V2	Event at V2
	MV	MV	NC	DT	MV	DT	DT	NC	Event at V4	No event, censor at baseline
	MV	DT	NC	MV	MV	DT	DT	NC	Event at V6	No event, censor at V3
	NC	DT	NC	MV	IMP	DT	NC	DT	Event not confirmed, but if last DT then confirmed	Event at V8

	NC	DT	NC	MV	IMP	DT	MV	DT	Event at V6	Event at V6
	NC	IMP	NC	IMP	MV	MV	IMP	DT	Event at V8, last DT	No event, censor at V4
	MV	MV	NC	MV	MV	DT	DT	NC	Event at V6	No event, censor at baseline
	NC	DT	MV	MV	DT	NC	IMP	DT	Event at V2	Event at V2

DT Deterioration; IMP Improvement; MV Missed visit; NC No change (/stable).

*Missed visits are allowed in between assessments confirming deterioration/worsening, and confirmation of deterioration/worsening requires at least 14 days between assessments.

Length of 2 missed visit window for EORTC QLQ-C30 and QLQ-LC13

The scheduled visit assessments and visit windows for the QLQ-C30 and QLQ-LC13 are every 4 weeks (±1 week) for the first 8 weeks from randomization, and then every 8 weeks (±1 week) up to 48 weeks relative to date of randomization, and then every 12 weeks (±1 week) until confirmed PD by RECIST 1.1 per Investigator assessment, and then every 12 weeks (±1 week) thereafter until PFS2.

Missing baseline or within 2 visits of baseline:

- A patient without baseline will be censored at day 1 unless they die ≤ day 63.
- Any assessment ≤ day 63 is within 2 visits (2 × 4 weeks + 1 week) of baseline and therefore cannot be censored for missing visits.

Schedule: q4w ± 1 week:

For any assessment > day 1 then visit schedule is q4w ± 1 week and thus the maximum time between 2 consecutive assessments is 7+28+28+7=70 days.

Schedule: q8w ±1 week:

- For any assessents where the visit schedule changes from q4w ± 1 week to q8w ±1 week, the maximum time between 2 consecutive assessments is 7+28+56+7 = 98 days.
- For any assessments where the visit scheulde is q8w ±1 week, the maximum time between two consecutive assessments is 7+56+56+7 = 126 days.

Schedule: q12w ±1 week:

- For any assessents where the visit schedule changes from q8w ±1 week to q12w ±1 week, the maximum time between 2 consecutive assessments is 7+56+84+7 = 154 days.
- For any assessments where the visit scheulde is q12w ±1 week, the maximum time between two consecutive assessments is 7+84+84+7 = 182 days.

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: d933yc00001-sap-ed-3		
Document Title:	Statistical Analysis Plan Edition 3	
Document ID:	CCI [redacted]	
Version Label:	3.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
20-Mar-2024 14:09 UTC	PPD	Content Approval

CCI [redacted]