AstraZeneca

Statistical Analysis Plan

Study Code	D933QC00001
Edition Number	5.0
Date	20 December 2023

A Phase III, Randomized, Double-blind, Placebo-controlled, Multi-center, International Study of Durvalumab or Durvalumab and Tremelimumab as Consolidation Treatment for Patients with Limited Stage Small-Cell Lung Cancer Who Have Not Progressed Following Concurrent Chemoradiation Therapy (ADRIATIC)

TABLE OF CONTENTS **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse events of special interest
ALT	Alanine aminotransferase
ALP	Alkaline phosphatase
AST	Aspartate aminotransferase
BICR	Blinded independent central review
BoR	Best objective response
CI	Confidence interval
CPAS	Combination PD-L1 analysis set
CR	Complete response
CrCl	Calculated creatinine clearance
CRF / eCRF	Case Report Form (electronic)
CRO	Contract Research Organisation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
cCRT	Concurrent chemoradiation therapy
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Event
ctDNA	circulating tumor DNA
DAE	Discontinuation of investigational product due to adverse event
DBL	Database lock
DCO	Data cut-off
DoR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organization for Research and Treatment of Cancer
EQ-5D-5L	EuroQoL 5-dimension, 5-levels
FAS	Full analysis set
FPAS	Full PD-L1 analysis set
GHS	Global Health Status

Abbreviation or special term	Explanation
HR	Hazard ratio
HRQoL	Health related quality of life
IA	Interim analysis
IC	Immune cells
ICU	Intensive care unit
IDMC	Independent data monitoring committee
IP	Investigational product
irRECIST 1.1	Immune-related response criteria modified
IV	Intravenous
IVRS	Interactive voice response system
LD	Longest diameter
LS-SCLC	Limited stage small cell lung cancer
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milli-gram
MRI	Magnetic resonance imaging
NA	Not applicable
nAb	neutralizing antibody
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NTL	Non-target lesions
OAE	Other significant adverse event
ORR	Objective response rate
OS	Overall survival
OS24	Proportion of patients alive at 24 months from randomization
OS36	Proportion of patients alive at 36 months from randomization
PCI	Prophylactic cranial irradiation
PD	Progressive disease
PD-L1	Programmed cell death ligand 1
PFS	Progression-free survival
PFS2	Time from randomization to second progression or death
PFS18	Progression-free survival at18 months following randomization

Abbreviation or special term	Explanation
PFS24	Progression-free survival at 24 months following randomization
PGIS	Patient's Global Impression of Severity
PK	Pharmacokinetic(s)
PR	Partial response
PRO	Patient reported outcomes
qxw	Every x weeks
QLQ-C30	EORTC 30-item core quality of life self-administered questionnaire
QLQ-LC13	EORTC 13-item lung cancer module self-administered questionnaire
QoL	Quality of life
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria In Solid Tumors
REML	Restricted maximum likelihood
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
T3	Triiodothyronine
T4	Thyroxine
TC	Tumor cells
TEAE	Treatment emergent adverse event
TFST	Time to first subsequent therapy or death
TL	Target lesions
TMB	Tumor mutational burden
TNM	Tumor, node, and metastatic classification
TSH	Thyroid-stimulating hormone
TTDM	Time to death or distant metastasis
ULN	Upper limit of normal
WHO	World Health Organization
WHODD	WHO drug dictionary

AstraZeneca

AMENDMENT HISTORY

Category*: Change refers to	Date and version	Description of change	In line with the CSP?	Rationale
Primary or secondary endpoints	23 July 2021, V2.0	The primary endpoint has been revised to look at dual primary endpoints of PFS and OS for the comparison of durvalumab monotherapy versus placebo.	Y (v4.0)	In line with CSP v4.0 update.
Primary or secondary endpoints	23 July 2021, V2.0	The comparison of durvalumb and tremelimumab combination versus placebo for PFS and OS have been changed to secondary endpoints. Section 1.1	Y (v4.0)	In line with CSP v4.0 update.
Primary or secondary endpoints	23 July 2021, V2.0	The exploratory endpoint, PD-L1 expression in tumor and/or immune cells relative to response/efficacy outcomes, has been promoted to a secondary endpoint. Section 1.1	Y (v4.0)	In line with CSP v4.0 update.
Primary or secondary endpoints	23 July 2021, V2.0	The secondary endpoint, TMB relative to response/efficacy outcomes, has been changed to a secondary endpoint. Section 1.1	Y (v4.0)	In line with CSP v4.0 update.
Study design	23 July 2021, V2.0	The planned number of patients to be recruited and screened and sample size	Y (v4.0)	In line with CSP v4.0 update.

		calculation has been updated. Section 1.2, 1.3 and Figure 1.		
Study design	23 July 2021, V2.0	The planned number of events required for the primary PFS and OS analyses, and OS interim analyses, and other associated parameters (e.g. % maturity, HR) have been updated. Section 1.3, 4.2.1 and 5.1.	Y (v4.0)	In line with update to dual primary endpoints and increase in sample size and change to recruitment and follow-up period.
Study Design	20 March 2023, V4.0	The study design has been updated for the following: inclusion of an interim analysis for PFS (per BICR); target number of events for the PFS primary analysis, associated statistical parameters and timing of the analysis; multiplicity control for PFS; assumed number of events at OS-IA1 and associated statistical parameters. Section 1.3, 4.2.1, 5.1, Appendix A.	Y (v5.0)	In line with CSP v5.0 update. Review of blinded event predictions show a potential shift in study timelines which has afforded an opportunity to introduce an IA for the primary endpoint PFS analysis.
Analysis sets	23 July 2021, V2.0	Inclusion of the combination analysis set and combination safety analysis set. Section 2.1, 2.1.2, 2.1.4.	Y (v4.0)	In line with CSP v4.0 update.
Analysis sets	23 July 2021, V2.0	Update to safety analysis set to provide further detail on how patients who receive incorrect therapy will be summarized. Section 2.1.3.	N/A	Further clarification provided.

Analysis sets	20 March 2023, V4.0	Update to the ADA analysis set to include subsets for the durvalumab and tremelimumab ADA analysis sets. Section 2.1 and Appendix D.	No	Certain ADA tables and listings are only based on either durvalumab or tremelimumab ADA samples.
Analysis sets	20 March 2023, V4.0	Update to the combination analysis set to clarify this includes the comparison of D+T vs placebo, and D vs D+T. Section 2.1 and 4.2.	No	Clarification
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Update to derivation of RECIST visit responses to remove text stating that baseline tumor assessments are to be performed 'ideally as close as possible prior to the start of study treatment'.	Y (v4.0)	In line with CSP v4.0 update.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Update to 2 missed visit rule for RECIST. Section 3.2.1	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Update to OS section to refer to Appendix C which details the eCRFs which will be used to determine the last data known to be alive, for patients that do not have a record In the SURVIVE module.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further detail provided for the EORTC QLQ-C30 functional and symptom scale derivations and scoring. Section 3.3.1.	N/A	Further clarification provided.

Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further detail provided of the PRO measures identified as 'primary' for the MMRM analysis. Section 3.3.1	Y (v4.0)	Further clarification provided
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Clarification added that RECIST 1.1 progression will not be considered HRQoL/function deterioration in the time to EORTC QLQ-C30 and QLQ-LC13 symptom/function deterioration. Section 3.3.1.1, 3.3.1.2.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further detail added on the response options for the PGIS questionnaire. Section 3.3.3.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Update to definition of an expected questionnaire to state that PFS2 or death as well as the date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms. Section 3.3.5.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Update to derivation of creatine clearance. Section 3.4.2.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Valid dose rule added to clarify what counts as a valid dose of treatment for exposure summaries. Section 3.4.5.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further clarification provided for the RDI calculation. Section 3.4.6.	N/A	Further clarification provided.

Derivation of primary or secondary endpoints	23 July 2021, V2.0	Inclusion of text related to AEPIs and Immune-mediated AEs. Section 3.4.1.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Inclusion of text related to the definition of concomitant medications. Section 3.4.7.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further clarification provided on definition of ADA positive patients. Section 3.5.2.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further clarification provided on flagging PD-L1 results outside of the recommended cut slide stability. Section 3.6.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further clarification provided that subsequent anti-cancer therapy does not include radiotherapy. Section 3.2, 3.4.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Further clarification provided on the definition of a treatment-emergent AE, and AEs with missing causality. Section 3.4.1.	N/A	Further clarification provided.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Additional text added to clarify the rule for handling multiple records within the same analysis visit window for ePRO. Section 4.1.3.	N/A	Further detail provided.

Derivation of primary or secondary endpoints	23 July 2021, V2.0	Update to which partial dates will be imputed, including rule for partial death dates. Section 4.1.3.	N/A	Correction.
Derivation of primary or secondary endpoints	23 July 2021, V2.0	Clarification added on reporting period for thyroid function tests. Section 4.2.13.3.	N/A	Clarification.
Derivation of primary or secondary endpoints	08 August 2022, V3.0	Section 3.2.1 (PFS) and Appendix B. Clarification that the 2 missed visit rule for PFS will follow a 'look-back' approach. Corrections to the study days when calculating the 2 missed visit rule.	N/A	Clarification and corrections provided.
Derivation of primary or secondary endpoints	08 August 2022, V3.0	Section 3.2.3 (ORR). Detail added on the definition of a 'confirmed' response.	N/A	Further detail provided.
Derivation of primary or secondary endpoints	08 August 2022, V3.0	Section 3.3.1.1 (Time to QLQ-C30 and QLQ-LC13 symptom deterioration) and Appendix B. Clarification that the 2 missed visit rule for ePRO will follow a 'look-forward' approach, what constitutes as 2 missed visits and that deterioration should be confirmed at the next subsequent assessment at least 14 days apart.	N/A	Clarification and further detail provided after input from ePRO SME
Derivation of primary or secondary endpoints	08 August 2022, V3.0	Section 3.3.1.3 (QLQ-C30 and QLQ-LC13 symptom improvement rate). Further detail provided on derivation of	N/A	Clarification and further detail provided.

		improvement rate for QLQ-LC13 during weekly assessments.		
Derivation of primary or secondary endpoints	20 March 2023, V4.0	Update to clarify that in regards to the 2 missed visit rule for PFS and PROs, that it is 2 consecutive missed visits, as opposed to any 2 missed visits. Section 3.2.1, 3.2.6, 3.3.1.1 and Appendix B,	N/A	Clarification.
Derivation of safety or exploratory endpoints	08 August 2022, V3.0	Section 3.3.3 (PGIS). Clarification on the derivation of 'overall' score for PGIS.	N/A	Clarification.
Derivation of safety or exploratory endpoints	08 August 2022, V3.0	Section 3.4.3 (ECGs). Further detail added on handling of triplicate values.	N/A	Further detail provided.
Derivation of safety or exploratory endpoints	08 August 2022, V3.0	Section 3.4.5 (treatment exposure). Details provided on dose unit conversions for durvalumab and tremeliumumab.	N/A	Further detail provided.
Derivation of safety or exploratory endpoints	08 August 2022, V3.0	Section 4.1.2 (Visit window for safety and PRO assessments). Clarification on how to handle missing laboratory assessment dates.	N/A	Clarification.
Derivation of safety or exploratory endpoints	08 August 2022, V3.0	Section 4.1.2 (Visit window for safety and PRO assessments). Correction provided on how to handle multiple ePRO records within the same analysis window.	N/A	Correction.

Derivation of safety or exploratory endpoints	08 August 2022, V3.0	Section 3.3.1 (EORTC QLQ-C30 and QLQ_LC13). Further detail provided on components of QLQ-C30 and QLQ-LC13 questionnaires including updates to Table 5 to detail scoring of LC13 scales/items. Update to terminology to use 'deterioration/deteriorated' rather than 'worsening/worsened'.	N/A	Further detail provided on review of ePRO SME. Consistency with other studies.
Derivation of safety or exploratory endpoints	20 March 2023, V4.0	Update to the definition of total (or intended) exposure. Section 3.4.5.	N/A	To correspond with Oncology TA SAP.
Derivation of safety or exploratory endpoints	20 March 2023, V4.0	Inclusion of rules for handling split and differing frequencies of radiotherapy dose. Section 3.4.8.	N/A	Previously not included.
Derivation of safety or exploratory endpoints	20 March 2023, V4.0	Inclusion of specific rules for what constitutes as a violation/deviation from the protocol in relation to the PK analysis set. Section 2.1.5 and Appendix D.	N/A	Previously not included.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Further detail added on which analyses will be presenting using the combination analysis set and combination safety analysis set. Section 4.1, 4.2, Table 7.	Y (v4.0)	In line with CSP v4.0 update.

Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Update to statistical hypotheses for new dual primary objectives. Section 4.1.1.	Y (v4.0)	In line with CSP v4.0 update.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Update to states that HRQoL/function improvement rate analysis will be performed using logistic regression analysis as opposed to summary statistics.	N/A	Correction.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Update to methods for multiplicity control due to update primary endpoints. Section 4.2.1.	Y (v4.0)	In line with CSP v4.0 update.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Further clarification provided on analysis of PFS including methods for handling non-proportional hazards. Section 4.2.2.	N/A	Further clarification provided.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	PD-L1 status added as a subgroup analysis for PFS and subgroup analyses for the stratification factors to also be repeated for values recorded on the eCRF. Section 4.2.2.	N/A	Added as analyses of interest.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Clarification provided on how the adjusted alpha levels for the interim and primary analysis of OS will be derived. Section 4.2.3, 5.1	N/A	Further clarification provided.

Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Additional sensitivity analysis added for OS to summarize duration of OS follow-up. Section 4.2.3.	N/A	Added as analysis of interest.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Correction added to state that the ORR analysis will be performed in a subset of the FAS, including all patients with measurable disease. Also to specify for the method for CI calculation. Section 4.2.4	N/A	Correction and clarification.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Additional PRO items added to the MMRM analysis, and removal of the Bonferroni-Holm procedure for adjusting the significance level. Section 4.2.11.1.	N/A	Correction.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Clarification on which laboratory assessments will be included in the summaries, and approach for calculated creatinine clearance values. Section 4.2.13.2.	N/A	Clarification.
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Inclusion of approach for handling NQ values in the PK analysis has been added. Section 4.2.15.	N/A	Clarification and in line with AZ SOP (AZ SOP LDMS_001_00201968)
Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	Updated to the expected timelines for conducting the IAs, including the expected number of events. Section 5.1.	Y (v4.0)	In line with CSP v4.0 update.

Statistical analysis method for the primary or secondary endpoints	23 July 2021, V2.0	The Haybittle-Peto method will be used to adjust the alpha level applied at the final analysis. Section 1.3, 4.2.1 and 5.1.	N/A	To account for actual alpha spent at the 2 interim analyses, based on the final total number of events, to maintain control of the Type I error.
Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Section 4.2.2 (PFS) and 4.2.3 (OS). Further detail provided on derivation of CIs for median PFS and OS.	N/A	Further detail provided.
Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Section 4.2.2 (PFS). Further detail provided on the PFS analysis methods for the consistency of treatment effect between subgroups.	N/A	Further detail provided.
Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Section 4.2.4 (ORR). Additional analysis added for ORR using the Cochran-Mantel-Haenszel test. Clarification provided that the analysis of ORR should be repeated for both confirmed and unconfirmed responses.	N/A	In line with FDA request for other Oncology studies. Clarification.
Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Section 4.2.7 (TTDM). Removal of time to censoring sensitivity analysis for TTDM.	N/A	Not required for secondary endpoints.
Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Addition of Appendix D which outlines pre-specified analyses to be conducted for China and Asia subsets.	N/A	Further detail provided.

Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Section 4.1 Analysis Methods - General Principals. Update to specify that p-values presented for objectives included in the MTP, will be rounded to 5 dps (rather than 4).	N/A	In order to provide more accuracy when comparing against the significance boundaries (Appendix A – Table 10)
Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Appendix A. Corrections provided to Table 10 and clarification provided that the significance boundaries will be calculated based on the actual number of events at the time of analysis. And for statistical significance to be declared, the p-value for the HR for the treatment effect must be < 2-sided significance boundary (rounded to 5 decimal places). Significance boundaries will also be calculated using the 'round' function rather than 'floor' function in SAS.	N/A	In order to provide more accuracy when comparing against the significance boundaries (Appendix A – Table 10)
Statistical analysis method for the primary or secondary endpoints	08 August 2022, V3.0	Section 4.2.2 Progression-free survival – Subgroup analyses. Update to the derivation and categorization of the subgroup analyses: Time from end date of cCRT to randomization, and Time from last dose of radiotherapy to ranandomization.	N/A	Clarification to definitions, and new cut-points agreed with medical team based on blinded review of the data, and clinical relevance.
Statistical analysis method for the primary or secondary endpoints	20 March 2023, V4.0	To update the analysis methods of ORR from logistic regression analysis to the Cochran-Mantel-Haenszel test.	Y (v5.0)	To correspond with Oncology TA SAP.

		Table 7 and Section 4.2.4.		
Statistical analysis method for the primary or secondary endpoints	20 March 2023, V4.0	To clarify that for the subgroup analyses, for subgroups with less than 20 events, this refers to across both treatment groups. Section 4.2.2.	N/A	Clarification.
Statistical analysis method for the primary or secondary endpoints	20 March 2023, V4.0	To clarify directionality of change for certain lab parameters. Section 4.2.13.2.	N/A	Clarification.
Data presentations	23 July 2021, V2.0	Inclusion of text related to COVID-19 related summaries. Section 2.2, 4.2.22.	N/A	In line with AZ guidance for CSRs.
Data presentations	08 August 2022, V3.0	Section 3.3.1 (EORTC QLQ-C30 and QLQ-LC13). Update to categories of clinically meaningful changes.	N/A	To align with presentation in TFLs.
Data presentations	08 August 2022, V3.0	Section 4.1.2 (Visit window for safety and PRO assessments). Update to requirement for visit based tables and plots, to just require a minimum of 20 observations per treatment group.	N/A	Added to prevent very large tables or plots being produced that contain many cells with meaningless data.
Data presentations	08 August 2022, V3.0	Section 4.2.2 (PFS). Update to subgroup analyses for PFS to specify for the stratification factors, subgroups based on both the IVRS and eCRF data should be included (applicable for OS as well).	N/A	Added to allow for any discrepancies between the IVRS and eCRF data.

Data presentations	08 August 2022, V3.0	Section 4.2.9 (PFS2). Removal of 'Symptomatic progression in absence of objective radiological progression' as a PFS2 category.	N/A	This is not collected in the eCRF.
Data presentations	08 August 2022, V3.0	Section 4.2.11.1 (EORTC QLQ-C30 and QLQ-LC13). Additional summary for compliance tables to be presented for patients who have discontinued treatment.	N/A	To establish any differences in the compliance rates between the FAS and patients discontinuing treatment.
Data presentations	08 August 2022, V3.0	Section 4.2.13.1 (AEs). Additional summaries by AEs with max CTCAE grade 3 or 4.	N/A	Durvalumab program level update.
Data presentations	08 August 2022, V3.0	Section 4.2.13.1 (AEs). Further clarification provided on derivation of AEs with max/any CTCAE grade 3 or 4, possibly related to study medication.	N/A	Durvalumab program level update.
Data presentations	23 June 2022, V2.0	Section 4.2.19 Demographic and baseline characteristics data. Additional characteristics listed to correspond with outputs.	N/A	To be in line with data presented in TFLs.
Data presentations	20 March 2023, V4.0	To include additional plots for liver enzyme parameters, with outliers removed. Section 4.2.13.2.	N/A	To ensure plots are interpretable.
Other	23 July 2021, V2.0	Reference list updated. Section 7.	N/A	To include new references in line with SAP update.

Other	23 July 2021, V2.0	General minor formatting and text updates throughout.	Y (v4.0)	In line with CSP v4.0 update.
Other	23 July 2021, V2.0	Section 6 – Changes of analysis from protocol updated to include 3 key changes from the CSP v4.0.	N	To documents changes to the analysis plan from the CSP v4.0.
Other	23 July 2021, V2.0	Appendix A on alpha spending function and MTP updated in line with the change to the primary endpoint and statistical hypotheses. Appendix B added to detail the derivation of the 2 missed visit rule for RECIST and ePRO. Appendix C added to specify which eCRFs will be used in the determination of last known alive date for OS.	N/A	Corrections and further clarifications provided.
Other	08 August 2022, V3.0	Removal of signature pages.	N/A	Study Statistician and Lead Statistician sign-off will be applied via ANGEL.
Other	08 August 2022, V3.0	Section 2.2. Updated wording for Deviation 6.	N/A	To correspond with Oncology TA SAP.
Other	08 August 2022, V3.0	General minor formatting and some correction of section and references hyperlinks throughout.	N/A	Corrections.

Other	20 March 2023, V4.0	Update throughout the SAP to refer to the primary PFS/OS analysis as opposed to the final PFS/OS analysis.	Y (v5.0)	In line with CSP v5.0 update. As what was previously referred to as the 'final' analysis i.e. when the target number of PFS/OS events had been reached, may not be the final analysis of this endpoint i.e. longer-term follow-up analyses may be performed.
Other	20 December 2023, V5.0	Section 1.1. The secondary endpoint, TMB relative to response/efficacy outcomes, has been changed to an exploratory endpoint.	Y (v5.0)	To correct according to CSP v4.0 and later.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 1.1.2. Added OS endpoint to the PD-L1 analysis	Y (v5.0)	To correct according to CSP v4.0 and later.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 1.3. Added wording "approximately" for flexible PFS primary analysis timing.	N/A	Allows flexibility to align a single DCO for PFS primary analysis and OS IA2 if predictions for the timing of the two analyses are close.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 1.3. Added option of OS analysis at the time of PFS primary analysis with 0.01% alpha (2-sided) allocated if OS IA2 doesn't coincide with PFS primary analysis.	N/A	Allows assessment of OS at time of primary PFS analysis to support regulatory filings if OS IA2 timing does not coincide with timing of the primary PFS analysis.
Sensitivity analysis	20 December 2023, V5.0	Section 2.2. Updated the deviation bias to be performed if >10% patients	N/A	To align with AZ Oncology standards.

		overall experience the three IPD categories mentioned in the SAP		
Other	20 December 2023, V5.0	Section 4.1.2. Updated the number of observations in each treatment group to be greater than the minimum of 20 patients "dosed" to "randomized" for PRO summary tables.	N/A	Clarification to align with the analysis sets used for PRO summaries (FAS and CAS).
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.1. Added option of OS analysis at the time of PFS primary analysis with 0.01% alpha (2-sided) spent if OS IA2 doesn't coincide with PFS primary analysis.	N/A	Allows assessment of OS at time of primary PFS analysis to support regulatory filings if OS IA2 timing does not coincide with timing of the primary PFS analysis.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.2. Added duration of follow-up summary for PFS.	N/A	To be consistent across endpoints, and to provide information of duration of follow-up for PFS.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.2. Added detail that the 1% cutoff for PD-L1 specified in the subgroup analysis is based on the frequency of PD-L1 positive tumor cells (TC) and/or immune cells (IC).	N/A	To clarify the cut-off threshold.
Other	20 December 2023, V5.0	Section 4.2.2. Deleted the "programmatically derived" language for PFS BICR.	N/A	Correction.

Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.3. Updated the method to present the duration of follow-up for OS.	N/A	To clarify the methods used and to provide alignment to the broader durvalumab program.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.4. Updated to clarify that same stratification factors for ORR analyses as in the analyses for PFS endpoint are used.	N/A	Corrections and clarifications. Stratifications factors for PFS and OS are determined separately based on the number of events observed for each endpoint. The revised text clarified that for ORR, the same stratification factors as those used for the PFS endpoint will be used given that both PFS and ORR are both response/RECIST-based endpoints.
Other	20 December 2023, V5.0	Section 4.2.13.1. Deleting the time to onset summary for preferred term	N/A	To align with the broader durvalumab program.
Other	20 December 2023, V5.0	Section 4.2.13.1. Deleting the summaries for patients who has at least one AESI/AEPI possibly related to study medication (as determined by the reporting investigator)	N/A	To align with the broader durvalumab program.
Other	20 December 2023, V5.0	Section 4.2.13.2. Updated the definition of Hy's law for the narratives.	Y (v5.0)	To align with AZ Oncology standards.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.18. Added more details to the biomarker data analysis, specifically for PD-L1.	N/A	To provide more information and clarification.

Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 5.1. Added option of OS analysis at the time of PFS primary analysis with 0.01% alpha (2-sided) spent if OS IA2 doesn't coincide with PFS primary analysis.	N/A	Allows assessment of OS at time of primary PFS analysis to support regulatory filings if OS IA2 timing does not coincide with timing of the primary PFS analysis.
Other	20 December 2023, V5.0	Appendix for China cohort analysis. Updated the definition of China cohort.	N/A	Clarification.
Other	20 December 2023, V5.0	List of abbreviations. Added more abbreviations and explanation: CPAS, FPAS, TC, and IC	N/A	Clarification.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 1.3. Updated the critical value for OS endpoint.	N/A	Correction.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 2.1. Added analysis populations for PFS, OS, and ORR endpoints.	N/A	Added populations for the PD-L1 subgroup analyses outlined in Section 4.2.18.
Other	20 December 2023, V5.0	Section 2.2. Clarified the important protocol deviation categories based on different version of CSPs.	Y (v5.0)	Clarification.
Other	20 December 2023, V5.0	Section 3.2.6. Added description for sensitivity analysis of TTDM using investigator assessments.	N/A	To provide more comprehensive analyses for TTDM.

Other	20 December 2023, V5.0	Section 3.4.8. Clarified the total dose derivation for radiotherapy administered with split dosing.	N/A	Clarification.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.1 and APPENDIX A. Updated Figure 2: multiple testing procedure for primary and key secondary endpoints.	N/A	Correction.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	Section 4.2.2. Clarified that BICR new lesions will be medically reviewed and categorized.	N/A	Clarification.
Other	20 December 2023, V5.0	Section 4.2.7. Added description for sensitivity analysis of TTDM using investigator assessments.	N/A	To provide more comprehensive analyses for TTDM.
Other	20 December 2023, V5.0	Section 4.2.21. Clarified that the subsequent therapies will be categorized by medical team.	N/A	Clarification.
Other	20 December 2023, V5.0	Section 6. Updated the change in SAP about the option of an additional OS analysis compared to CSP V5.0.	N/A	Correction.
Statistical analysis method for the primary or secondary endpoints	20 December 2023, V5.0	APPENDIX A. Updated the alpha spending calculation to incorporate the additional potential OS analyses at the time of PFS primary with 0.01% alpha (2-sided) spent if OS IA2 and PFS primary don’t coincide.	N/A	Allows assessment of OS at time of primary PFS analysis to support regulatory filings if OS IA2 timing does not coincide with timing of the primary PFS analysis.

Other	20 December 2023, V5.0	APPENDIX D and Section 1.2. Clarified text for Asian sub-population analyses	N/A	Clarified text as Asian and China sub-population will only be performed if required to support country specific submissions.
Other	20 December 2023, V5.0	General minor formatting and text updates throughout.	N/A	Correction and clarification.

* Pre-specified categories are: Primary or secondary endpoints; Statistical analysis method for the primary or secondary endpoints; Derivation of primary or secondary endpoints; Derivation of safety or exploratory endpoints; Analysis sets; Study design; Multiple Testing Procedure; Data presentations; Other

1. STUDY DETAILS

1.1 Study Objectives

1.1.1 Primary objectives

Dual primary objectives:	Endpoint/variables:
To assess the efficacy of durvalumab monotherapy compared to placebo in terms of PFS	PFS using BICR assessments according to RECIST 1.1
To assess the efficacy of durvalumab monotherapy compared to placebo in terms of OS	OS

BICR Blinded Independent Central Review; OS Overall survival; PFS Progression-free survival; RECIST Response Evaluation Criteria In Solid Tumors.

1.1.2 Secondary objectives

Secondary objectives:	Endpoint/variables:
To assess the efficacy of durvalumab and tremelimumab combination compared to placebo in terms of PFS and OS	PFS using BICR assessments according to RECIST 1.1 OS
To further assess the efficacy of durvalumab monotherapy and durvalumab and tremelimumab combination therapy compared to placebo in terms of ORR, PFS18[a], PFS24[a], TTDM, OS24, OS36, and PFS2	ORR, PFS18[a], PFS24[a], and TTDM using BICR assessments according to RECIST 1.1 OS24 and OS36 PFS2
To assess the efficacy of durvalumab and tremelimumab combination therapy compared to durvalumab monotherapy in terms of PFS, OS, and ORR	PFS and ORR using BICR assessments according to RECIST 1.1 OS
To assess disease-related symptoms and HRQoL in patients treated with durvalumab monotherapy and durvalumab and tremelimumab combination therapy compared to placebo using the EORTC QLQ-C30 and QLQ-LC13	EORTC QLQ-C30 and QLQ-LC13: change in symptoms, functioning, and global health status/QoL
To assess the PK of durvalumab monotherapy and durvalumab and tremelimumab combination therapy	Concentration of durvalumab and tremelimumab in serum (such as peak concentration and trough; sparse sampling)
To investigate the immunogenicity of durvalumab monotherapy and durvalumab and tremelimumab combination therapy	Presence of ADA for durvalumab and tremelimumab (confirmatory results: positive or negative)

Secondary objectives:	Endpoint/variables:
To investigate the relationship between PD-L1 expression and spatial distribution within the tumor microenvironment and clinical outcomes with durvalumab monotherapy or durvalumab and tremelimumab combination therapy	PD-L1 expression in tumor and/or immune cells (cutoff ≥ 1%) relative to response/efficacy outcomes (PFS, OS, and ORR). Other PD-L1 cutoffs may also be analyzed

[a] Progression-free survival at 18 and 24 months following randomization (PFS18 and PFS24) is equivalent to the proportion of patients alive and progression-free at 18 and 24 months following randomization (APF18 and APF24), respectively.
ADA Anti-drug antibody; BICR Blinded Independent Central Review; EORTC European Organization for Research and Treatment of Cancer; HRQoL Health-related quality of life; ORR Objective response rate; OS Overall survival; OS24 Proportion of patients alive at 24 months from randomization; OS36 Proportion of patients alive at 36 months from randomization; PD-L Programmed cell death ligand 1; PFS Progression-free survival; PFS2 Time from randomization to second progression; PFS18 Progression-free survival at 18 months following randomization; PFS24 Progression-free survival at 24 months following randomization; PK Pharmacokinetic(s); QLQ-C30 30-item core quality of life questionnaire; QLQ-LC13 Lung cancer module; QoL Quality of life; RECIST Response Evaluation Criteria In Solid Tumors; TMB Tumor mutational burden; TTDM Time to death or distant metastasis.

1.1.3 Safety objective

Safety objective:	Endpoint/variables:
To assess the safety and tolerability profile of durvalumab monotherapy and durvalumab and tremelimumab combination therapy compared to placebo in patients with limited stage small cell lung cancer (LS-SCLC)	AEs; laboratory findings including clinical chemistry, hematology, and urinalysis; physical examinations; vital signs including blood pressure and pulse; and electrocardiograms

AE Adverse event; LS-SCLC Limited stage small cell lung cancer.

1.1.4 Exploratory objectives

Exploratory objectives*:	Endpoint/variables:
To assess treatment-related side effects in patients treated with durvalumab monotherapy and durvalumab and tremelimumab combination therapy compared to placebo using PRO-CTCAE	Change in the 9 treatment-related symptoms evaluated in this study
To assess the patients' overall impression of the severity of their cancer symptoms using PGIS	PGIS: Proportion of patients assessing current symptom severity
To describe and evaluate health resource use associated with durvalumab monotherapy and durvalumab and tremelimumab combination therapy and underlying disease	Health resource utilization measures including hospitalization, outpatient visits, or emergency department visits
To explore the impact of treatment and disease state on health state utility using the EQ-5D-5L	The EQ-5D-5L health state utility index will be used to derive health state utility based on patient-reported data

Exploratory objectives*:	Endpoint/variables:
To collect blood and tissue samples, or leverage residual samples, for analysis of peripheral and tumoral biomarkers (not applicable to China)	Exploratory biomarkers, which may include but are not limited to, DNA, RNA, and protein-based assessment within the tumor microenvironment and/or in the periphery. Evaluation of tumor-cell and/or immune-cell gene expression profiles, tumor or ctDNA-derived mutational analyses, PD-L1 expression, SCLC molecular subtypes, tumor-immune spatial profiling, etc., and association of biomarkers with response and/or resistance
To investigate the relationship between TMB measured in tumor and/or blood and efficacy outcomes with durvalumab monotherapy and durvalumab and tremelimumab combination therapy (TMB-related testing or analysis will not be conducted on samples from China)	TMB relative to response/efficacy outcomes (ORR, PFS, and OS)
To explore the relationship(s) between patient biomarker status and durvalumab PK exposure and clinical outcomes before and after treatment (TMB related testing or analysis will not be conducted on samples from China)	Biomarker status before and after treatment, durvalumab PK exposure, and relationship with clinical outcomes, efficacy, AEs, and/or safety parameters, as deemed appropriate
To explore irRECIST as assessment methodologies for clinical benefit of durvalumab monotherapy and durvalumab and tremelimumab combination therapy compared to placebo with assessment by BICR	PFS and ORR using BICR assessment according to irRECIST
To collect and store DNA from tissue and/or blood according to each country's local and ethical procedures for future exploratory research into genes/genetic variation that may influence response (i.e. distribution, safety, tolerability, and efficacy) to IPs and/or susceptibility to disease (optional, not applicable for China)	Correlation of polymorphisms with variation in PK, pharmacodynamics, safety, or response parameters observed in patients treated with durvalumab and/or susceptibility to disease
To investigate the effect of baseline colonic microbiome on response to treatment and the effect of treatment on the microbiome over time (applicable for EU and North America only)	

AE Adverse event; BICR Blinded Independent Central Review; CTCAE Common Terminology Criteria for Adverse Events; ctDNA Circulating tumor DNA; EQ-5D-5L EuroQoL 5 dimension, 5-level health state utility index; IP Investigational product; 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; PGIS Patient's Global Impression of Severity; PK Pharmacokinetic(s); PRO Patient-reported outcome; TMB Tumor mutational burden.
*Exploratory objectives may be reported outside of the CSR.

1.2 Study Design

This study is a Phase III, randomized, double-blind, placebo-controlled, multi-center study assessing the efficacy and safety of durvalumab or durvalumab and tremelimumab combination therapy versus placebo as consolidation treatment in patients with limited stage small cell lung cancer (LS-SCLC) who have not progressed following definitive, platinum-based, concurrent chemoradiation therapy (cCRT).

In order to be eligible for this study, patients must have achieved complete response (CR), partial response (PR), or stable disease (SD) and have not progressed following definitive, platinum-based, cCRT. This cCRT treatment, and prophylactic cranial irradiation (PCI) treatment if received per local standard of care, must be completed within 1 to 42 days prior to randomization and the first dose of IP (i.e. durvalumab, tremelimumab, or placebo) in this study. In addition, the baseline efficacy assessment must be performed post-cCRT as part of the screening procedures within 42 days before randomization and the first dose of IP.

Approximately 965 patients will be recruited and screened globally in order to enrol and randomize approximately 724 patients to 1 of 3 treatment groups: durvalumab monotherapy (approximately 262 patients), durvalumab and tremelimumab combination therapy (approximately 200 patients), or placebo (approximately 262 patients). Randomization will be stratified by tumor, node, and metastatic classification (TNM) stage (I/II versus III) and receipt of PCI (yes versus no).

Patients will receive 1 of the following treatments, based on their randomized treatment group assignment:

- Durvalumab monotherapy: Durvalumab (1500 mg intravenous [IV]) every 4 weeks (q4w) in combination with placebo saline solution (IV) q4w for up to 4 doses/cycles each, followed by durvalumab 1500 mg q4w. The first durvalumab monotherapy 1500 mg dose q4w will be 4 weeks after the final dose of durvalumab in combination with placebo saline solution.

- Durvalumab in combination with tremelimumab: Durvalumab (1500 mg IV) q4w in combination with tremelimumab (75 mg IV) q4w for up to 4 doses/cycles each, followed by durvalumab 1500 mg q4w. The first durvalumab monotherapy 1500 mg dose q4w will be 4 weeks after the final dose of durvalumab in combination with tremelimumab.

- Placebo: Placebo saline solution (IV) q4w in combination with a second placebo saline solution (IV) q4w for up to 4 doses/cycles each, followed by a single placebo saline solution q4w. The first placebo saline solution monotherapy dose q4w will be 4 weeks after the final dose of the 2 placebo saline solutions in combination.

Treatment in all treatment groups will be administered beginning on Day 1 until clinical/RECIST 1.1-defined radiological progression, until intolerable toxicity, or for a maximum of 24 months, whichever occurs first.

Patients who have discontinued treatment due to toxicity, symptomatic deterioration, or clinical progression will be followed up with tumor assessments until RECIST 1.1-defined radiological progression plus one follow-up scan or until death (whichever comes first) and followed for survival.

This main body of this Statistical Analysis Plan (SAP) outlines the pre-specified analyses for the overall study population (global cohort). Appendix D outlines pre-specified analyses that may be conducted for the China sub-population (China cohort) and Asia subset (patients enrolled from Asian countries) of study D933QC00001 to support ADRIATIC submission in China, which will be reported outside of the CSR.

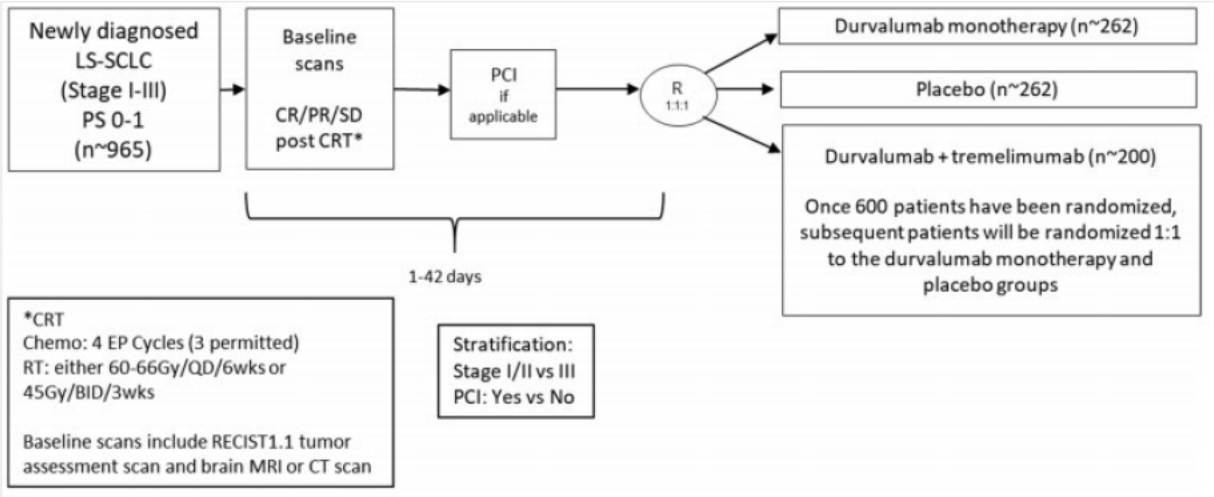
Data Monitoring Committee:

An Independent Data Monitoring Committee (IDMC) comprised of independent experts will be convened to review planned unblinded safety data. This planned safety review will take place after the first 20 patients have been randomized into each of the 3 treatment groups (i.e. after a total of 60 patients have been randomized to the study). In addition, the IDMC will review planned interim analyses on overall efficacy and safety and inform the sponsor whether the interim boundaries are met.

Details on the IDMC are provided in Section 5.2. Full details of the IDMC procedures, processes, and interim analyses can be found in the IDMC Charter.

The general study design is summarized in Figure 1.

Figure 1 Study design



BID Twice daily; CR Complete response; CRT Concurrent Chemoradiation therapy; EP Etoposide and cisplatin chemotherapy; LS-SCLC Limited stage small-cell lung cancer; PCI Prophylactic cranial irradiation; PD Progressive disease; PR Partial response; PS Performance status; QD Once daily; R Randomization; RT Radiotherapy; SD Stable disease.

1.3 Number of Subjects

The study will plan to enroll approximately 965 patients globally in order to randomize approximately 724 patients to 1 of 3 treatment groups: durvalumab monotherapy (approximately 262 patients), placebo (approximately 262 patients), or durvalumab + tremelimumab combination therapy (approximately 200 patients) over a period of approximately 38 months. Initially, patients will be randomized 1:1:1 to the 3 treatment groups. Following implementation of CSP Version 4, once 600 patients have been randomized, a further 124 patients will subsequently be randomized 1:1 to durvalumab monotherapy or placebo until a total of approximately 724 patients have been randomized.

The study has dual primary objectives which are to assess the efficacy of durvalumab monotherapy compared to placebo in terms of PFS per RECIST 1.1 as assessed by BICR, and OS.

The primary PFS analysis will occur at the earliest of:

1. When approximately 370 PFS BICR events have occurred (70.6% maturity) in the durvalumab monotherapy and placebo treatment groups
2. At OS-IA2, if OS-IA2 is statistically significant (durvalumab monotherapy vs placebo)
3. Approximately 36 months after the last patient randomized

If the true PFS HR is 0.65 for durvalumab monotherapy versus placebo, with 370 PFS BICR events, the study will have approximately 90% power to demonstrate a statistically significant difference in PFS between durvalumab monotherapy and placebo, with an overall 2-sided significance level of 0.5%. The true HR of 0.65 translates to a 5.4 month benefit in median PFS over 10 months on placebo if PFS is exponentially distributed. The smallest treatment difference that would be statistically significant is a HR of 0.743 (Critical value (CV)). At this time, approximately 309 PFS BICR events are also expected to have occurred in the durvalumab in combination with tremelimumab and placebo treatment groups. A recruitment period of approximately 38 months was assumed in the CSP for the primary PFS analysis.

One interim analysis of PFS will be performed when approximately 308 PFS BICR events have occurred across the durvalumab monotherapy and placebo treatment groups (information fraction (IF) 83.2%, maturity 58.8%). At this time, approximately 274 PFS BICR events are also expected to have occurred in the durvalumab in combination with tremelimumab and placebo treatment groups. With 308 PFS BICR events across the durvalumab monotherapy and placebo treatment groups, the study will have approximately 75% power to detect a PFS HR of 0.65 (CV=0.700) at a 0.184% significance level. The alpha level (0.5%, 2 sided) will be split between the interim and primary analyses using the Lan and DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach. The actual boundary will be calculated at the time of the interim analysis, based on the actual number of events available at the time of analysis, and assuming 370 PFS BICR events at the primary PFS analysis. The alpha level applied at the primary PFS analysis will be adjusted (using a generalized Haybittle-Peto method (SAS Institute Inc. 2018)) to account for the actual alpha

spent at the interim analysis based on the actual final total number of events, to maintain control of the overall Type I error.

The primary OS analysis will occur when approximately 348 death events have occurred (66.4% maturity) in the durvalumab monotherapy and placebo groups. At this time, approximately 276 death events are also expected to have occurred in the durvalumab in combination with tremelimumab and placebo treatment groups. If the true OS HR is 0.73 for durvalumab monotherapy versus placebo, the study will have 80% power to demonstrate a statistically significant difference in OS between durvalumab monotherapy and placebo. The true HR of 0.73 translates to an approximate 8.9 month benefit in median OS over 24 months on placebo if OS is exponentially distributed. The smallest treatment difference that would be statistically significant is a HR of 0.798.

Up to three interim analyses of OS will be performed: one at the time of the PFS interim analysis with approximately 242 death events anticipated across the durvalumab monotherapy and placebo groups (IF 69.5%, maturity 46.2%) and another with approximately 299 death events (IF 85.9%, maturity 57.1%). For these analyses the information fraction corresponds to the proportion of information i.e. events, available at the time of analysis. The alpha will be split between the interim and primary analyses using the Lan and DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach. The actual significance boundaries will be calculated at the time of each interim, based on the actual number of events available at the time of analysis, and assuming 348 death events at the primary OS analysis. In addition, an alpha of 0.01% (2-sided) will be allocated for an OS assessment at the time of PFS primary analysis if OS-IA2 does not coincide with the PFS primary analysis.

The alpha level applied at the primary OS analysis will be adjusted (using a generalized Haybittle-Peto method (SAS Institute Inc. 2018) to account for the actual alpha spent at the interim analyses based on the actual final total number of events, to maintain control of the overall Type I error. Further details on the methods for multiplicity control and the multiple testing procedure (MTP) are presented in Section 4.2.1 and the planned interim analyses, in Section 5.1.

2. ANALYSIS SETS

2.1 Definition of Analysis Sets

Definitions 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, OS	Full analysis set, Combination analysis set, Full PD-L1 analysis set, Combination PD-L1 analysis set
ORR*	Full analysis set, Combination analysis set, Full PD-L1 analysis set, Combination PD-L1 analysis set
PFS18, PFS24, TFST, TTDM, OS24, OS36, PFS2, BoR, DoR* and PRO endpoints*	Full analysis set, Combination analysis set
Study Population/Demography Data	
Demography characteristics (e.g. age, sex etc.)	Full analysis set, Combination analysis set
Baseline and disease characteristics	Full analysis set, Combination analysis set
Important deviations	Full analysis set, Combination analysis set
Medical/Surgical history	Full analysis set, Combination analysis set
Previous anti-cancer therapy	Full analysis set, Combination analysis set
Concomitant medications/procedures	Full analysis set, Combination analysis set
Subsequent anti-cancer therapy	Full analysis set, Combination analysis set
World Health Organization performance status	Full analysis set, Combination analysis set
PK Data	
PK data	PK analysis set
Safety Data	
Exposure	Safety analysis set, Combination safety analysis set
Adverse events	Safety analysis set, Combination safety analysis set
Laboratory measurements	Safety analysis set,
Vital Signs	Safety analysis set
Electrocardiograms	Safety analysis set
ADA	ADA analysis set, Durva ADA analysis set, Treme ADA analysis set

ADA Anti-drug antibody; DoR Duration of response; ORR Objective response rate; OS Overall survival; OS24 Proportion of patients alive at 24 months from randomization; OS36 Proportion of patients alive at 36 months from randomization; PFS Progression-free survival; PFS2 Time from randomization to second progression; PFS18 Progression-free survival at 18 months following randomization; PFS24 Progression-free survival at 24

months following randomization; PK Pharmacokinetic(s); PRO Patient-reported outcome; TFST Time to first subsequent therapy; TTDM Time to death or distant metastasis; Durva Durvalumab; Treme Tremelimumab.
*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. Patient evaluability for PRO endpoints is detailed in Section 3.3.

2.1.1 Full analysis set

The full analysis set (FAS) will include all randomized patients. The FAS will be used for all efficacy analyses (including PROs). Treatment groups will be compared on the basis of randomized study treatment, regardless of the treatment actually received. Patients who were randomized but did not subsequently go on to receive study treatment are included in the analysis in the treatment group to which they were randomized.

2.1.2 Combination analysis set

For analyses involving the durvalumab and tremelimumab combination treatment group (i.e. durvalumab and tremelimumab combination therapy versus placebo and durvalumab monotherapy versus durvalumab and tremelimumab combination therapy), only the first 600 patients randomized (across all 3 arms) will be included in the analyses, and all will be included in the treatment group to which they were randomized.

2.1.3 Safety analysis set

The safety analysis set will consist of all patients who received at least 1 dose of study treatment (see Section 3.4.5 for Valid Dose rule). Safety data will not be formally analyzed but summarized using the safety analysis set according to the treatment received, that is, erroneously treated patients (i.e. those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received.

Patients who receive incorrect therapy will be summarized according to treatment group as follows:

- Patients receiving only placebo infusions, at any time, will be summarized in the placebo arm.
- Patients receiving durvalumab, at any time, but not receiving tremelimumab, at any time, will be summarized in the durvalumab arm.
- Patients receiving tremelimumab, at any time, will be summarized in the durvalumab + tremelimumab arm.

2.1.4 Combination safety analysis set

The combination safety analysis set will consist of all patients from the combination analysis set who received at least 1 dose of study treatment. Data will be summarized according to the treatment they actually received, applying the same approach for those receiving incorrect therapy as outlined for the safety analysis set above.

2.1.5 Full PD-L1 analysis set

The full PD-L1 analysis set (FPAS) will consist of all patients with evaluable PD-L1 data within the full analysis set.

2.1.6 Combination PD-L1 analysis set

The combination PD-L1 analysis set (CPAS) will consist of all patients with evaluable PD-L1 data within the combination analysis set.

2.1.7 PK analysis set

All patients who receive at least 1 dose of IP per the protocol for whom any post-dose data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analyses will be included in the PK analysis set. Further details on what constitutes as a violation/deviation from the protocol can be found in Appendix E. The population will be defined by the Clinical Pharmacologist (clin pharm lead) and Statistician prior to any analyses being performed.

2.1.8 ADA analysis sets

The anti-drug antibody (ADA) analysis set includes all patients in the safety analysis set who have non-missing baseline ADA and at least 1 non-missing post-baseline ADA result of the same IP (durvalumab or tremelimumab).

The durvalumab ADA analysis set will consist of all patients in the safety analysis set who have a non-missing baseline durvalumab ADA result and at least one non-missing post-baseline durvalumab ADA result.

The tremelimumab ADA analysis set will consist of all patients in the safety analysis set who have a non-missing baseline tremelimumab ADA result and at least one non-missing post-baseline tremelimumab ADA result.

2.2 Violations and Deviations

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

- Deviation 1: Patients randomized but who did not receive study treatment.
- Deviation 2: Patients who deviate from the following key entry criteria enrolled per CSP:
    - Version 1: Inclusion criteria 5, 6, 7, 8 and 9 and exclusion criteria 1, 2, 3, 12, 15, 17, 18, 24, 25, and 26.
    - Version 2 and later: Inclusion criteria 5, 6, 7, 8 and 9 and exclusion criteria 1, 2, 3, 12, 15, 17, 18, 23, 24, and 25.

- Deviation 3: Baseline RECIST 1.1 scan > 42 days before randomization.
- Deviation 4: No baseline RECIST 1.1 assessment post-cCRT on or before date of randomization.
- Deviation 5: Received prohibited concomitant medications. Please refer to the CSP Section 6.4 for those medications that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to Database lock (DBL).
- Deviation 6: Patients randomized who received their randomized study treatment at an incorrect dose or received an alternative treatment to that which they were randomized.

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 (regarding continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader, medic, and/or statistician.

The important protocol deviations will be listed and summarized by randomized treatment group, including COVID-19 related IPDs. Deviation 1 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 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK). 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:

- Did not have the intended disease or indication or
- Did not receive intended first line chemoradiotherapy
- Did not receive any randomized therapy.

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

In addition to the programmatic determination of the deviations above, other study deviations captured from the CRF 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 (including differences in stratification factors between IVRS and eCRF).

In addition, all COVID-19 related non-important PDs and issues will be summarized and listed and included in the CSR.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST Visit Responses

For all patients, the RECIST tumor response data will be used to determine each patient's visit response according to RECIST version 1.1 (see further Appendix F of the CSP). It will also be used to determine if and when a patient has progressed in accordance with RECIST and their best objective response to study treatment.

Baseline radiological tumor assessments are to be performed post-cCRT within the 42 days before randomization and start of study treatment. Tumor assessments are then performed every 8 weeks (± 1 week) for the first 72 weeks (relative to the date of randomization), followed by every 12 weeks (± 1 week) up to 96 weeks (relative to the date of randomization), and then every 24 weeks (± 1 week) thereafter (relative to the date of randomization) until RECIST 1.1-defined radiological progression, plus one follow-up scan no earlier than 4 weeks later and no later than the next regularly scheduled imaging visit.

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 minimize 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 tumor response data will be used to determine each patient's visit response according to RECIST version 1.1. At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD or Progressive 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 no evidence of disease (NED) at baseline, then the patient will be assigned a visit response of NED if there is still no evidence of disease or PD if there is evidence of progression. If a patient has had a tumor assessment which 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 outcomes (i.e. PFS and ORR etc.) will be calculated programmatically for the BICR and site investigator data from overall visit responses.

3.1.1 Blinded Independent Central Review (BICR) assessment using RECIST 1.1

A planned BICR of all radiological imaging data will be carried out using RECIST version 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 Organization (CRO) for central analysis. The imaging scans will be reviewed by 2 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 or NE; for patients with no disease identified at baseline: PD, no evidence of disease [NED], NE). If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

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

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 up to the data cut-off will be completed before database lock for the interim analysis. Additionally, a BICR of all patients will be performed for the final database lock for PFS which will cover all the scans up to that point.

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

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

3.1.2 Site investigator assessment using RECIST 1.1: target lesions (TLs)

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 5 measurable lesions recorded at baseline with a maximum of 2 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 the date of randomization will be

used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

All other lesions (or sites of disease) not recorded as TL should be identified as NTL 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 3.1.3 for further details). If a patient does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

Table 2 TL Visit Responses (RECIST 1.1)

Visit Responses	Description
Complete Response (CR)	Disappearance of all target lesions since baseline. Any pathological lymph nodes selected as target lesions must have a reduction in short axis to <10mm.
Partial Response (PR)	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive Disease (PD)	A ≥ 20% increase in the sum of diameters of target lesions and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.
Not Evaluable (NE)	Only relevant in certain situations (i.e. if any of the target lesions 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 progressive disease criteria, progressive disease overrides not evaluable as a target lesion response.
Not Applicable (NA)	No target lesions are recorded at baseline

Rounding of TL data

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

Missing TL data

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

Note: the nadir can only be taken from assessments where all the TLs had a lesion diameter 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 recorded and this size should still be used to determine the TL visit response as normal. In the special case where all lymph nodes are < 10 mm and all other TLs are 0 mm then although the sum may be >0 mm the calculation of TL response should be over-written as a CR.

TL visit responses subsequent to CR

A CR 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 are 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 case report form and a value of 5mm 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 case report form and has entered a smaller value that can be reliably measured. If a TL response of PD results then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

A previously irradiated lesion may be selected as a Target Lesion provided it fulfils the criteria for reproducible measurability and is the only lesion available

Any TL (including lymph nodes), which has had intervention during the study (for example, 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 tumors:

- 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 if ≤ 1/3 of the TLs have missing measurements then scale up as described in the 'Scaling' section below. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then if appropriate (i.e. if ≤ 1/3 of the TLs have missing measurements), the scaled sum of diameters calculated in step 2 should be used, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or

<10mm for lymph nodes) and the lesions that have been subject to intervention have a value of 0 (or <10mm for lymph nodes) recorded. If scaling up is not appropriate due to too few non-missing measurements then the visit response will be set as NE.

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

Scaling (applicable only for lesion intervention)

If > 1/3 of target lesion measurements are treated as missing (because of intervention) then target lesion response will be NE, unless the sum of diameters of non-missing target lesion 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 target lesions has increased by 5mm from nadir).

If ≤ 1/3 of the target lesion measurements are treated as 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).

Example of scaling

Lesion 5 is missing at the follow-up visit; it had a BL measure of 29.3cm.

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 28.4cm:

$$\frac{26}{26.8} \times 29.3 = 28.4cm$$

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.

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of TLs

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

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

3.1.3 Site investigator assessment using RECIST 1.1: 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 follows:

Table 3 NTL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
Progressive Disease (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases, the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Non-CR/Non-PD	Persistence of one or more NTLs with no evidence of progression.
Not Evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For patients without TLs at baseline, this is relevant if any of the NTLs were not assessed at this visit and the progression criteria have not been met.
Not Applicable (NA)	Only relevant if there are no NTLs at baseline

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

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

A lesion 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 tumor.

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

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

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

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

3.1.4 Site investigator assessment using RECIST 1.1: overall visit response

Table 4 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 4 Overall Visit Responses

Target Lesions	Non-target lesions	New Lesions	Overall Response
CR	CR or NA	No (or NE)	CR
NA	CR	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
NA	Non CR/Non PD	No (or NE)	SD
NE	Non PD or NE or NA	No (or NE)	NE
NA	NE	No (or NE)	NE
PD	Any	Any	PD

Target Lesions	Non-target lesions	New Lesions	Overall Response
Any	PD	Any	PD
Any	Any	Yes	PD
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).

3.2 Outcome Variables

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

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.

Where applicable, in the derivation of endpoints, anti-cancer therapy does not include subsequent radiotherapy.

3.2.1 Progression-free survival (PFS)

PFS (per RECIST 1.1 as assessed by the BICR) will be defined as the time from the date of randomization until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomized therapy or receives another anti-cancer therapy prior to progression (i.e. date of event or censoring – date of randomization + 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 RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more consecutive missed visits, the patient will be censored at the time of the latest RECIST 1.1 assessment prior to the 2 missed visits (Note: NE is not considered a missed visit). A 'look-back' approach will be taken (i.e. if an event e.g. progression or death is observed, then it will be considered an event only if there is an assessment within the 2 visit window immediately before the event, irrespective of whether there were missed visits prior to that assessment). 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 i.e. ≤119 days (2 x 8 weeks plus 1 week allowing for a late assessment within the visit window), then they will be treated as an event with date of death as the event date.

Given the scheduled visit assessment scheme (i.e. q8w ± 1w for the first 72 weeks then q12w ± 1w up to 96 weeks then q24w ± 1w thereafter) the definition of 2 missed visits will change. See Appendix B for further details on the derivation of the 2 missed visit rule for RECIST assessments.

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

PFS will also be obtained using the algorithm described above for the RECIST site investigator tumor data.

RECIST 1.1 assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- For BICR assessments, the date of progression will be determined based on the earliest scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of the reviewer who read baseline first if there is no adjudication.
- For investigator assessments, the date of progression will be determined by the earliest of the RECIST 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**:** For TLs, only the latest scan date is recorded out of all scans performed at that assessment for the target lesions and similarly for NTLs only the latest scan date is recorded out of all scans performed at that assessment for the NTLs.

A sensitivity analysis of PFS will be performed using Investigator assessments according to RECIST 1.1.

For exploratory purposes, PFS may also be obtained using the irRECIST 1.1 data obtained from BICR.

Time to first subsequent therapy or death (TFST)

As a supportive summary to PFS, time to start of first subsequent therapy or death (TFST) will be derived. TFST is defined as the time from randomization to the start date of the first subsequent anti-cancer therapy after discontinuation of treatment, or death (i.e. date of first subsequent anti-cancer therapy/death or censoring – date of randomization + 1). Any patient not known to have had a first subsequent therapy or died, will be censored at the last date that the patient was known not to have received a first subsequent anti-cancer therapy. 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 randomized treatment would have TFST calculated in the same way, i.e. time from date of randomization to the subsequent therapy or death.

3.2.2 Overall survival (OS)

Overall survival is defined as the time from the date of randomization until death due to any cause (i.e. date of death or censoring – date of randomization + 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 (see Appendix C).

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

Note, 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 among the dates recorded on the case report forms (CRFs) (see Appendix C).

If any of the dates from the CRFs are after study discontinuation or after the DCO, then the minimum of the study discontinuation and DCO date will be used instead.

3.2.3 Objective response rate (ORR)

ORR (per RECIST 1.1 as assessed by the BICR) is defined as the number (%) of patients with at least 1 visit response of CR or PR (i.e. unconfirmed response) and will be based on a subset of all randomized patients. If the BICR finds any patients do not have measurable disease at baseline then the analysis of ORR for the BICR data will exclude these patients, so that the denominator is a subset of the randomized patients who have measurable disease at baseline per BICR for either reviewer. Therefore, 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 randomized treatment without progression, receive a subsequent anti-cancer therapy (note that radiotherapy is not considered a subsequent anti-cancer therapy) and then respond will not be included as responders in the ORR.

ORR will also be obtained using the algorithm described above for the RECIST site investigator tumor data. The denominator for ORR will be all randomized patients with measurable disease at baseline per the site investigator.

For exploratory purposes, ORR may also be obtained for the irRECIST 1.1 data obtained from BICR.

3.2.4 Duration of response (DoR)

DoR (per RECIST 1.1 as assessed by the 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 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 their DoR will be censored at the PFS censoring time.

DoR will not be defined for those patients who do not have documented response.

3.2.5 Progression-free survival at 18 months and 24 months (PFS18 and PFS24)

The PFS18 and PFS24 will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed by the BICR) at 18 months and 24 months, respectively.

3.2.6 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 randomization 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 or proven by biopsy. 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 RECIST 1.1 assessment. However, if the patient has distant metastasis or dies after 2 or more consecutive missed visits, the patient will be censored at the time of the latest evaluable 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.

A sensitivity analysis of TTDM may be performed using Investigator assessments according to RECIST 1.1.

3.2.7 Proportion of patients alive at 24 months and 36 months after randomization (OS24 and OS36)

The OS24 and OS36 will be defined as the Kaplan-Meier estimate of OS at 24 months and 36 months after randomization, respectively.

3.2.8 Time from randomization to second progression or death (PFS2)

PFS2 will be defined as the time from the date of randomization to the earliest of the progression event subsequent to the first subsequent anticancer therapy (excluding radiotherapy) or death (i.e. date of PFS2 event or censoring – date of randomization + 1). The date of second progression will be recorded by the investigator in the eCRF at each assessment and defined according to local standard clinical practice and may involve any of

the following: objective radiological imaging, symptomatic progression or death. The date of the PFS2 assessment and investigator opinion of progression status (progressed or non-progressed) at each assessment will be recorded in the eCRF. Patients must have had a first progression event and received subsequent systemic anticancer therapy (excluding radiotherapy) for a second progression event to be valid. Second progression status will be reviewed every 8 weeks following the progression event used for the primary variable PFS (the first progression) and status recorded. Patients alive and for whom a second disease progression has not been observed should be censored at the earliest of: date of study termination, date last known alive, DCO, or if a patient has not had a first subsequent therapy; the date last known not to have received a first subsequent therapy (TFST censoring date).

However, if the patient experiences a second progression or dies after 2 or more consecutive missed visits, the patient will be censored at the time of the last PFS2 assessment prior to the 2 missed visits.

3.2.9 Best objective response (BoR)

BoR is calculated based on the overall visit responses from each RECIST assessment. It is the best response a patient has had following randomization but prior to starting any subsequent anticancer therapy (excluding radiotherapy) and up to and including RECIST progression or the last assessment in the absence of RECIST progression and subsequent cancer therapy.

Categorization of BoR will be based on RECIST using the following response categories: CR, PR, SD, PD and NE.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 8 weeks minus 1 week, i.e. at least 49 days (to allow for an early assessment within the assessment window), after randomization (i.e. study day 50). 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 from the overall visit response using all BICR data up until the first progression event, the start of subsequent cancer therapy or the last assessment in the absence of progression/subsequent cancer therapy. The denominator will be consistent with that used in the ORR analysis. In particular, patients with no evidence of disease will be excluded.

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

For patients who die with no evaluable RECIST assessments, if the death occurs ≤119 days (i.e. 2 x 8 weeks + 1 week to allow for a late assessment within the assessment window) after randomization, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >119 days (i.e. 2 x 8 weeks + 1 week) after randomization then BoR will be assigned to the NE category.

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

3.2.10 Change in TL tumor size

For supportive purposes percentage change from baseline in tumor size will be derived at each scheduled tumor assessment visit (i.e. week 8, week 16 etc hereafter referred to as week X for convenience). Best percentage change from baseline in tumor size will also be derived as the biggest decrease or the smallest increase in tumor size from baseline.

This is based on RECIST target lesion measurements taken at baseline and at the timepoint of interest. Tumor size is defined as the sum of the longest diameters of the target lesions for the BICR data based upon RECIST assessments. Target lesions are measurable tumor lesions. Baseline for RECIST is defined to be the last assessment prior to randomization. The change in target lesion tumor size at week X will be obtained for each patient by taking the difference between the sum of the target lesions at week X and the sum of the target lesions at baseline. To obtain the percentage change in target lesion tumor size at week X the change in target lesion tumor size is divided by the sum of the target lesions at baseline and multiplied by 100 (i.e. (week X - baseline) / baseline * 100). More details on target lesions and measurements can be found in Section 3.1.

The above derivations will be programmed for the BICR data based upon RECIST 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.

3.3 Patient Reported Outcome (PRO) Variables

PRO variables will be assessed using the European Organization for Research and Treatment of Cancer (EORTC) 30-item core quality of life questionnaire (QLQ-C30) with the lung cancer module (QLQ-LC-13) (health-related quality of life [HRQoL] with lung cancer specific additional concerns), EQ-5D-5L, Patient's Global Impression of Severity (PGIS), and PRO-CTCAE. All items/questionnaires will be scored according to published scoring guidelines or the developer's guidelines, if published guidelines are not available. All PRO analyses will be based on the FAS, unless otherwise stated. PRO questionnaires will be collected as per the assessment schedule in Table 1 of the CSP, up until study termination, PFS2 (as defined in Section 3.2.8) or death. Rules for handling multiple records are detailed in Section 4.1.2.

3.3.1 EORTC QLQ-C30 and QLQ-LC13

The EORTC QLQ-C30 consists of 30 questions that are grouped into 5 multi-item functional scales (physical, role, cognitive, emotional, and social), 3 multi-item symptom scales (fatigue, pain, and nausea/vomiting), 5 single items assessing additional symptoms commonly reported

by cancer patients (dyspnea, insomnia, appetite loss, constipation, and diarrhea), a 2-item global measure of health status/quality of life (GHS/QoL), and a single item on the financial impact of the disease. The GHS/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 EORTC QLQ-C30 will be scored according to the EORTC QLQ-C30 scoring manual (Fayers et al 2001).

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

All items are scored between 1 ("not at all") to 4 ("very much") with the exception of the 2 GHS/QoL items which are scored 1 ("very poor") to 7 ("excellent"). An outcome variable consisting of a transformed score from 0 to 100 will be derived for each of the symptom scales/symptom items, the functional scales and the GHS/QoL scale in the QLQ-C30 and for each of the symptom scales/items in the QLQ-LC13 according to the EORTC QLQ-C30 Scoring Manual (Fayers et al 2001) and EORTC QLQ-LC13 instructions.

Higher scores on the GHS/QoL and functional scales indicate better health status/function, but higher scores on symptom scales/items represent greater symptom severity. For each subscale, if <50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers et al 2001). If at least 50% of the items are missing, then that subscale 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 minimized. Rules for handling multiple records are detailed in Section 4.1.2.

Functional and symptom scale derivations

The EORTC QLQ-C30 functional and symptom scales, individual symptom items and global health status are derived as follows.

1. Calculate the average of the items that contribute to the scale or take the value of an individual item, i.e. the raw score (RS):

$$RS = (I_1 + I_2 + \ldots + I_n) / n,$$

where $I_1 + I_2 + \ldots + I_n$ are the items included in a scale and n is the number of items in a scale.

2. Use a linear transformation to standardize the raw score, so that scores range from 0 to 100, where a higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.

   Functional scales: Score = (1 – [RS – 1] / *range*) x 100

   Symptom scales/items; global health status: Score = ([RS – 1] / *range*) x 100,

where *range* is the difference between the maximum and the minimum possible value of RS.

The scoring approach for the EORTC QLQ-LC13 is identical to that for the symptom scales / single items of the QLQ-C30. The number of items and item range for each scale/item are displayed in Table 5.

Table 5 Scoring of the EORTC QLQ-C30 and QLQ-LC13

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
QLQ-C30				
Global health status/QoL	QL2	2	6	29, 30
Functional scales				
Physical	PF2	5	3	1-5
Role	RF2	2	3	6, 7
Cognitive	CF	2	3	20, 25
Emotional	EF	4	3	21-24
Social	SF	2	3	26, 27
Symptom scales				
Fatigue	FA	3	3	10, 12, 18
Pain	PA	2	3	9, 19
Nausea/ vomiting	NV	2	3	14, 15
Symptom items				

Scale/ item	Scale/ item abbreviation	Number of items (n)	Item range	Item numbers
Dyspnea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhea	DI	1	3	17
Financial difficulties*	FI	1	3	28
QLQ-LC13				
Symptom scales / items				
Dyspnoea	LCDY	3	3	3,4,5
Coughing	LCCO	1	3	1
Haemoptysis	LCHA	1	3	2
Sore mouth	LCSM	1	3	6
Dysphagia	LCDS	1	3	7
Peripheral neuropathy	LCPN	1	3	8
Alopecia	LCHR	1	3	9
Pain in chest	LCPC	1	3	10
Pain in arm or shoulder	LCPA	1	3	11
Pain in other parts	LCPO	1	3	12

EORTC European Organisation for Research and Treatment of Cancer; QLQ-C30 30-item core quality-of-life questionnaire; QoL Quality of life.

*Not reported on in the analysis.

Baseline will be defined as described in Section 4.1.2.

The following 5 PRO measures have been identified as primary from the EORTC QLQ-C30 and EORTC QLQ-LC13, namely:

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

Furthermore, physical functioning, role functioning, and GHS/QoL domains of the EORTC QLQ-C30 are pre-specified endpoints of interest (FDA, 2021).

Definition of clinically meaningful changes

Changes in score compared with baseline will be evaluated. 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 (). For example, a clinically relevant deterioration in chest pain (as assessed by QLQ-LC13) is defined as an increase in the score from baseline of ≥10. A clinically relevant improvement in fatigue (as assessed by QLQ-C30) 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 categorized as improvement, no change, or deterioration as shown in Table 6.

Table 6 Visit responses for symptoms and HRQoL

Score	Change from baseline	Visit response
QLQ-C30/QLQ-LC13 symptom scales/items	≥+10	Deterioration
	≤-10	Improvement
	>-10 to <+10	No change
QLQ-C30 functional scales and GHS/QoL	≥+10	Improvement
	≤-10	Deterioration
	>-10 to <+10	No change

GHS/QoL Global health status/Quality of life; HRQoL Health-related quality of life; QLQ-C30 30-Item core quality of life questionnaire; QLQ-LC13 Lung cancer module.

For the visit level summaries of Improvement/Deteriorated/No change then all patients with a baseline and post-baseline score will be included thus the denominator may differ from the time to deterioration and improvement rate endpoints derived below.

3.3.1.1 Time to QLQ-C30 and QLQ-LC13 symptom deterioration

For each of the symptoms scales/items in the QLQ-C30 and QLQ-LC13, time to symptom deterioration will be defined as the time from randomization until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) that is confirmed at the next available assessment at least 14 days apart or death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration (i.e. date of symptom deterioration event or censoring – date of randomization + 1). 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 (and where there were no previous 2 consecutive missed visits) where the symptom change could be evaluated. Patients with a single deterioration and no further assessments will be treated as deteriorated in the analysis.

Patients whose symptoms (as measured by QLQ-C30 and QLQ-LC13) have not shown a clinically meaningful deterioration 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, or prior to 2 consecutive missed visits if 2 missed visits are observed. Also, if symptoms deteriorate after 2 or more missed PRO assessment visits or the patient dies after 2 or more consecutive 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 . 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. Confirmation of deterioration will first be determined, then the censoring rules will be applied to assess if any determined deterioration can be used or if it needs to be censored earlier due to 2 missed visits..

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

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

In this analysis, RECIST 1.1 progression will not be considered as symptom deterioration and data will not be affected by RECIST 1.1 progression.

3.3.1.2 Time to QLQ-C30 GHS/QoL/functional deterioration

For GHS/QoL and functional deterioration, time to deterioration will be defined as the time from the date of randomization until the date of the first clinically meaningful deterioration (a

decrease in the functional scales or the GHS/QoL from baseline of ≥10) that is confirmed at the next available assessment at least 14 days apart, or death (by any cause) in the absence of a clinically meaningful deterioration , regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to GHS/QoL and functional deterioration (i.e. date of GHS/QoL/functional deterioration event or censoring – date of randomization + 1). 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 (and where there were no previous 2 consecutive missed visits) where the GHS/QoL/function change could be evaluated (prior to the 2 missed assessment visits, Appendix B). Patients with a single deterioration and no further assessments will be treated as deteriorated in the analysis.

Patients whose GHS/QoL/function (as measured by QLQ-C30) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the GHS/QoL/function could be evaluated, or prior to 2 missed visits if 2 missed visits are observed. Also, if GHS/QoL/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 the symptom 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. 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. Confirmation of deterioration will first be determined, then the censoring rules will be applied to assess if any determined deterioration can be used or if it needs to be censored earlier due to 2 missed visits.

See Appendix B 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 analysis of time to GHS/QoL/function deterioration will include a subset of the FAS who have baseline scores ≥ 10.

In this analysis, RECIST 1.1 progression will not be considered as GHS/QoL/function deterioration and data will not be affected by RECIST 1.1 progression.

3.3.1.3 QLQ-C30 and QLQ-LC13 symptom improvement rate

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 from baseline score ≥10 for symptom scales/items) 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.

When the assessment schedule is weekly i.e. QLQ-LC13 weeks 1 to 8, an improvement needs to be observed at each assessment that falls within the 14 day interval to be classed as an improvement. For example:

V1 (Week 1)	V2 (Week 2)	V3 (Week 3)	V4 (Week 4)	Outcome
IMP	IMP	IMP	NC	Improvement
IMP	IMP	NC	DT	No improvement*
IMP	MV	IMP	IMP	No improvement*
NC	IMP	NC	IMP	No improvement*

DT Deterioration; IMP Improvement; MV Missed visit. NC No change.
*Assuming no further assessments.

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

3.3.1.4 QLQ-C30 GHS/QoL/functional improvement rate

The GHS/QoL/functional improvement rate will be defined as the number (%) of patients with 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (an increase from baseline score ≥10 for functional scales and GHS/QoL) 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 a subset of the FAS who have a baseline score ≤90.

3.3.2 EQ-5D-5L

The EQ-5D-5L is a standardized 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 EQ-5D-5L index comprises six questions that cover five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) and a global health status

visual analog scale (VAS). 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 unable to/extreme problems). A unique EQ-5D-5L 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-5L 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 values sets are not available, the EQ-5D-5L to EQ-5D-3L crosswalk (Oemar and Janseen 2013) will be applied. 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.

The evaluable population will comprise a subset of the FAS who have a baseline EQ-5D-5L assessment.

3.3.3 PGIS

The PGIS item is included to assess how a patient perceives his/her overall current severity of cancer symptoms. Patients will choose from response options including: “no symptoms”, “very mild”, “mild”, “moderate”, “severe” and “very severe”. For summaries of the “overall” score, “overall” will present the best post-baseline response over all time points, ranked from best (i.e. “no symptoms”), to worst (i.e. “very severe”).

3.3.4 PRO-CTCAE

The PRO-CTCAE is included to address tolerability from the patient’s perspective. It was developed by the National Cancer Institute (NCI). The PRO-CTCAE will only be administered in English, German, Spanish and Japanese. In countries where a linguistically validated version does not exist, the English version is administered for patients who signed English informed consent. The PRO-CTCAE was developed in recognition that collecting treatment-related symptom data directly from patients using PRO tools can improve the accuracy and efficiency of symptomatic AE data collection. This was based on findings from multiple studies demonstrating that physicians and nurses underestimate treatment-related symptom onset, frequency, and severity in comparison with patient ratings (Basch et al 2009, Litwin et al 1998, Sprangers and Aaronson 1992). These treatment-related symptoms have been converted to patient terms (e.g. the CTCAE term “myalgia” has been converted to “aching muscles”). For several symptoms, like fatigue and pain, additional questions are asked about the frequency, severity, and interference with usual activities. The items included in the PRO-CTCAE have undergone extensive qualitative review among experts and patients. These items have been extensively evaluated by cancer subjects to be clear, comprehendible, and measure the symptom of interest. For this study, 9 symptoms are considered relevant for this cancer treatment: swallowing, nausea, heartburn, diarrhea, wheezing, rash, itchy skin, skin burns from radiation and headache (see CSP Appendix H).

3.3.5 PRO compliance rates

Summary measures of overall compliance and compliance over time will be derived for the EORTC QLQ-C30, LC13 and EQ-5D-5L respectively. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.

- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. Date of study discontinuation, PFS2 or death will be mapped to the nearest visit date to define the number of expected forms.

- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that is non-missing.

- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all 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 randomized 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 by100.

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.

3.4 Safety

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

'On treatment' will be defined as assessments between date of the first dose and 90 days following last dose of the study treatment or up to and including the date of initiation of the first subsequent anticancer therapy (excluding radiotherapy) (whichever occurs first).

The safety analysis set will be used for reporting of safety data.

3.4.1 Adverse events (AEs)

AEs and SAEs will be collected throughout the study, from the date of informed consent until 90 days after the last dose of the study treatment. 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 4.03). A treatment emergent adverse event (TEAE) is an AE with an onset date/time or a pre-existing AE worsening (taking the last grade prior to dosing as the reference) on or following the first dose of study treatment through to 90 days after the last dose of the study treatment or initiation of first subsequent therapy (whichever occurs first). All AEs will be listed however only TEAEs will be summarized. For each treatment group, in the unlikely event of the two components being administered separately then date of first dose/last dose will be considered as the earliest/latest dosing date of either component.

Adverse events that have missing causality (after data querying) will be assumed to be related to study drug ("durvalumab or placebo" or "tremelimumab or placebo"), with the exception of causality assessments that are not applicable for the patient, e.g. causal relation to durvalumab for patients in the placebo arm.

AEs of special interest (AESI) and AEs of potential interest (AEPI)

Some clinical concepts (including some selected individual preferred terms [PTs] and higher level terms [HLTs]) have been considered "AEs of special interest" (AESI) to the durvalumab program. AESIs represent pre-specified risks which are of importance to a clinical development 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 durvalumab and requiring more frequent monitoring and/or interventions such as systemic 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 Adverse Events of Possible Interest (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 durvalumab 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 are identified as a list of categories provided by the clinical team. Other categories may be added as necessary or existing terms may be merged. A further review will take place prior to DBL to ensure any further terms not already included are captured within the categories.

Immune-mediated Adverse Events (imAE)

To fully characterize the AESI (excluding AESI group Infusion related/ Hypersensitivity/ Anaphylactic reactions) during which systemic corticosteroids, endocrine therapy, or other immunosuppressants were administered), the Sponsor will classify AESIs as immune-mediated AEs (imAEs) or not imAEs. Further details are provided in an imAE Charter.

The imAEs will be determined by a programmatic algorithm that requires specific treatment for AESIs to be considered imAEs such as systemic corticosteroids, immunosuppressants, and/or endocrine therapy. The same specific treatment is required for AEPIs as well, though only to AEPIs that have been retained for further treatment consideration after identifying those that have been assessed by the investigator as possibly related to any study treatment and/or as an imAE.

Other significant adverse events (OAE)

During the evaluation of the AE data, an AstraZeneca medically qualified expert will review the list of AEs that were not reported as SAEs and 'Discontinuation of Investigational Product due to Adverse Events' (DAEs). Based on the expert's judgement, significant adverse events of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered other significant adverse events (OAEs) and reported as such in the CSR. A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs.

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

3.4.2 Laboratory data

Laboratory data will be collected throughout the study, from screening to the follow-up visits as described in the CSP. Blood and urine samples for determination of hematology, 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 4.1.2 will be used.

Change from baseline in hematology 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 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: absolute Lymphoctyes, Potassium, Sodium, Magnesium, Glucose and Corrected calcium.

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

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

Calculated creatinine clearance (CrCl) will be derived according to the Cockroft-Gault formula (using most recent body weight measurement if available, otherwise subsequent weight can be used) (Cockcroft and Gault 1976).

Males:

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

Females:

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

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-dose 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 needs only to have 1 post dose-value recorded.

3.4.3 ECGs

ECG data obtained up until the 30 days from date of last dose of study treatment will be used for reporting. Resting 12-lead electrocardiograms (ECGs) are recorded at screening and as clinically indicated thereafter. Categorical summaries of change from baseline in overall ECG assessments (recorded as “abnormal” and “normal”) will be created if sufficient number of ECG assessments are recorded. Where triplicate values are recorded, the average will be used in summary tables, where relevant. Individual values will be reported in listing.

3.4.4 Vital signs

Vital signs data obtained up until the 30 days from date of last dose of study treatment will be used for reporting. 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 4.1.2 will be used.

3.4.5 Treatment exposure

Exposure will be defined as follows and calculated separately for "durvalumab or placebo" and "tremelimumab or placebo". Exposure in each treatment group is the maximum of calculated values from the two components.

Total (or intended) exposure

- Total (or intended) exposure = the earliest of (date of last dose date of study drug where valid dose is considered + 27 days, death date or DCO) – first dose date + 1 day.

Actual exposure

- Actual exposure is defined as intended exposure, but excluding total duration of dose interruptions and cycle delays

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

Exposure will also be measured by the number of cycles received. If a cycle is prolonged due to toxicity, this should still be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

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

- Duration of dose delays will be calculated as follows for each dose and summed over the entire dosing period to obtain a total duration of dose delay:

  If Date of the dose – Date of the previous dose > 31 days (28 + 3 days) then

  Duration of dose delay= Date of the dose - Date of previous dose – 31 days

  Otherwise duration of dose delay will be zero.

Valid dose rule

A valid dose of treatment will be classed as:

- At least one record with "DURVALUMAB/PLACEBO" or "TREMELIMUMAB/PLACEBO" and where the start date is non-missing.
- Volume of infusion must be greater than 0 (before – after)
- Study Drug Dose per administration must not be 0 or missing.

Dose unit conversion

Dose of durvalumab and tremelimumab is collected in mL. To be reported in units of mg using the following conversions:

Durvalumab:

$$dose\ in\ mg = dose\ in\ mL\ x\ 50$$

e.g. 30mL × 50 = 1500mg

Tremelimumab:

$$dose\ in\ mg = dose\ in\ mL\ x\ 19.73684211$$

e.g. 3.8mL × 19.73684211 = 75mg

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.

3.4.6 Dose intensity

Dose intensity will be derived for each treatment group. 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 to the actual last day of dosing of the respective treatment (durvalumab or tremelimumab) and D is the intended cumulative dose up to the actual last day of dosing of the respective treatment. 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 may be added to the date of last dose to reflect the protocol allowed window for dosing.

- The dosing of "durvalumab or placebo" and "tremelimumab or placebo" will be considered together in the derivation, i.e. actual and intended cumulative dose in the first 4 cycles taken as the sum of doses from the two components.

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

3.4.7 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 + 90 days 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.

3.4.8 Concurrent Chemoradiation Therapy (cCRT)

cCRT must be completed within 1 to 42 days prior to randomization and the first dose of IP (i.e. durvalumab, tremelimumab, or placebo) in this study.

Where radiotherapy is administered with split dosing, the total dose must be derived (i.e. the absolute values from each split dose are to be added together, regardless of the dose per fraction with which they were delivered, and used in the summary of the radiotherapy data.

Where radiotherapy administration is split by differing frequency (once daily with a total dose of 60 to 66 Gy or twice daily with a total dose of 45 Gy), the total dose must still be derived as above but for all other radiotherapy parameters, the frequency with ≥95% dose received will be used in the analysis.

3.5 Pharmacokinetic and Immunogenicity Variables

3.5.1 Pharmacokinetic analysis

The actual sampling times will be used in the PK calculations. PK concentration data and summary statistics will be tabulated and listed. Individual and mean blood concentration-time profiles will be generated. Methods for handling samples below the lower limit of quantification (LLOQ) are described in Section 4.2.15.

3.5.2 Immunogenicity analysis

Serum samples for ADA assessments will be conducted utilizing 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 neutralizing antibody (nAb) will be tested for all ADA positive samples using a ligand binding assay. The

nAb results will be reported as positive or negative. 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.

For each subject, following responses variables will be evaluated for durvalumab and tremelimumab:

- ADA prevalence, defined as the proportion of study population having drug reactive antibodies at any point in time, baseline or post-baseline
- ADA incidence (treatment-emergent ADA), defined as the sum of both treatment-induced and treatment-boosted ADA.
- 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 (regardless of Baseline result), or an ADA positive result at the last available assessment (regardless of Baseline result).
- Transiently positive ADA, defined as having at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive (regardless of Baseline result).
- nAb positive at any visit

3.6 Biomarker Variables

Programmed cell death ligand 1 (PD-L1) expression level in tumor cells (TC) and in immune cells (IC) as measured by immunohistochemistry (IHC) staining will be collected. A flag will be derived in the reporting database to identify cases that were outside the recommended cut slide stability, i.e. stained date – sectioned date > 90 days.

Tumor Mutational Burden (TMB) in both tumor tissue and blood samples will be measured.

3.7 Health Care Resource Use

To investigate the impact of treatment and disease on health care resource, the following variables will 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 the patient presents with
- Length of hospital stay
- Length of any time spent in an intensive care unit (ICU)

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

4. ANALYSIS METHODS

4.1 General Principles

All summaries will be presented in tabular format by treatment group, unless otherwise specified. Data will be presented in data listings by treatment group and patient number.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days.

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

- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarized 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, median, minimum and maximum. Categorical variables will be summarized by frequency counts and percentages for each category.
- Unless otherwise stated, percentages will be calculated out of the population total and for each treatment group.
- For continuous data, the mean and median will be rounded to 1 additional decimal place compared to the original data. 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.
- P-values will be rounded to 4 decimal places (with the exception of those in the MTP). P-values less than 0.00005 (e.g. 0.00002) will not be rounded to 4 decimal places (e.g. 0.0000) but instead be displayed as <0.0001. P-values output as <0.0001 by statistical software will not be rounded and will be displayed in the same way ('<0.0001'). P-value presented for the MTP will be rounded to 5 decimal places.
- SAS® version 9.4 will be used for all analyses.

Efficacy and PRO data will be summarized and analyzed on the FAS, or combination analysis set where stated. Safety and treatment exposure data will be summarized based on the safety analysis set, or combination safety analysis set where stated. Study population and demography data will be summarized based on the FAS, or combination analysis set where stated.

4.1.1 Statistical hypotheses

The dual primary objectives are to assess the efficacy of durvalumab monotherapy compared to placebo in terms of PFS per RECIST 1.1 as assessed by BICR and OS. The study will be considered positive (i.e. a success) if either of the below null hypotheses are rejected based on the primary analysis of PFS or OS in the FAS.

The statistical hypotheses for primary PFS:

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

The statistical hypotheses for primary OS are:

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

The following statistical hypotheses for the key secondary endpoints of PFS and OS comparing durvalumab in combination with tremelimumab versus placebo will be tested if the null hypothesis for both dual primary endpoints (PFS and OS, durvalumab monotherapy versus placebo) is rejected (see Section 4.2.1):

- H0: No difference between durvalumab in combination with tremelimumab and placebo

- H1: Difference between durvalumab in combination with tremelimumab and placebo

Analyses will be stratified by disease stage (I/II versus III) based on the TNM classification and receipt of PCI (yes versus no).

Statistical analyses of the durvalumab in combination with tremelimumab group will be performed in the combination analysis set to ensure no bias is introduced from an imbalance in the follow-up and recruitment period.

In general, for efficacy and PRO endpoints the last observed measurement prior to randomization will be considered the baseline measurement. However, if an evaluable assessment is only available after randomization but before the first dose of randomized treatment then this assessment will be used as baseline. For safety endpoints, the last observation before the first dose of study treatment will be considered the baseline measurement unless otherwise specified. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as 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 protocol 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 Visit window for safety and PRO assessments

Time windows will need defining for any presentations that summarize 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 summarized. Inclusion within the time window should be based on the actual date of assessment and not the intended date of the visit. For laboratory assessments, if date of assessment is missing, visit date will be used.

- All unscheduled visit data should have the potential to be included in the summaries.

- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls 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 (with 4 weeks between scheduled assessments) 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

  Note that visits up to 28 days after the last dosing date will be considered as being on treatment for the purposes of visit windowing and may be assigned to an on-treatment visit. Visits after this will be considered as follow-up and may be assigned accordingly.

- 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 non-EPRO 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.
- For ePRO based visit summaries, if there are multiple records within the same analysis visit window, then the closest assessment to the target date should be used (by date, time and sequence). If there are multiple records on the same day prior to the target day, then the latest assessment will be used (by time and sequence). If there are multiple records on the same day after the target day, then the earliest assessment will be used.
- 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 summarized if the number of observations is greater than the minimum of 20 patients dosed for safety summaries and 20 patients randomized for PRO summaries.

- 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 of safety data will be defined as the last non-missing measurement prior to the first dose of study treatment. For laboratory data, 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 summarized over time, study day will be calculated in relation to date of first treatment. When PRO data are summarized over time, study day will be calculated in relation to date of randomization.

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.

4.1.3 Handling of missing and incomplete dates

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 randomization.

During the creation of the reporting database, partial start dates including, but not limited to: AEs, medications and subsequent therapy, will be imputed as follows:

- If only day is missing: impute day as the first day of the month (unless month and year are the same as month and year of first dose of study treatment, in which case day should be imputed as date of first dose);

- If day and month are missing: impute day and month as the first day of the year (unless year is the same as year of first dose of study treatment, in which case day and month should be imputed as date of first dose);

- If the start date is missing, then the analysis start date will not be imputed.

Partial end dates will be imputed as follows:

- If only day is missing: impute day as the earlier of either the DCO or the last day of the month;

- If day and month are missing: impute day and month as the earlier of either the DCO or the last day of the year;

- If the end date is missing, then the analysis end date will not be imputed.

If either both the start and end date of an AE or medication are missing, or the start date of an AE or medication is missing, but the end date is complete or imputed and on or after the date of first dose, then the AE or medication is considered treatment emergent or concomitant.

Partial death dates will be imputed as follows:

- If only day is missing: impute day as the latest of the first day of the month or the last known alive date + 1;

- If day and month are missing: impute day and month as the latest of the first day of the year or the last known alive date + 1;

- If the year is missing, then the analysis date will not be imputed.

4.2 Analysis Methods

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

Table 7 details which endpoints are to be subject to formal statistical analysis, together with pre-planned sensitivity analyses making clear which analysis is regarded as primary for that endpoint. Note, all efficacy endpoints (including PROs) comparing durvalumab monotherapy with placebo will be in all randomized patients (FAS). Efficacy analyses comparing durvalumab and tremelimumab combination therapy with placebo, and durvalumab monotherapy with durvalumab and tremelimumab combination therapy, will be in the combination analysis set, unless otherwise indicated.

Table 7 Pre-planned statistical and sensitivity analyses to be conducted

Endpoints Analyzed	Notes
Progression-free survival	Stratified log-rank tests for: • Primary analysis using BICR assessments (RECIST 1.1) for durvalumab monotherapy versus placebo • Sensitivity analyses using BICR assessments (RECIST 1.1) 1) Interval censored analysis – evaluation time bias 2) Analysis using alternative censoring rules – attrition bias • Sensitivity analysis using site Investigator assessments (RECIST 1.1) – ascertainment bias • Secondary analysis using BICR assessments (RECIST 1.1) for durvalumab + tremelimumab versus placebo and durvalumab monotherapy versus durvalumab + tremelimumab) Subgroup analysis using Cox proportional hazards models Additional analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate Additional analysis using Cox proportional hazards models to determine the consistency of treatment effect between subgroups via the approach of Gail and Simon 1985.
Overall survival	Stratified log-rank tests for: • Primary analysis (durvalumab monotherapy versus placebo) • Secondary analysis (durvalumab + tremelimumab versus placebo and durvalumab monotherapy versus durvalumab + tremelimumab) Sensitivity analysis using a Kaplan-Meier plot of time to censoring where the censoring indicator of the primary analysis is reversed – attrition bias Subgroup analysis using Cox proportional hazards models Additional analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate Additional analysis using Cox proportional hazards models to determine the consistency of treatment effect between subgroups via the approach of Gail and Simon 1985.

Endpoints Analyzed	Notes
Objective response rate	Cochran-Mantel-Haenszel (CMH) test for difference in proportions using BICR assessments (RECIST 1.1) adjusting for the same factors as the primary endpoint for: • Secondary analysis (durvalumab monotherapy and durvalumab + tremelimumab combination therapy versus placebo and durvalumab monotherapy versus durvalumab + tremelimumab) Sensitivity analysis using CMH test repeated using the site Investigator assessments (RECIST 1.1)
Duration of response	Kaplan-Meier estimates and confidence intervals
PFS at 18 months and 24 months	Kaplan-Meier estimates and confidence intervals
Time to death or distant metastasis	Stratified log-rank test using BICR assessments (RECIST 1.1)
OS at 24 and 36 months	Kaplan-Meier estimates of survival at 24 months and 36 months and analyses following the method described by Klein et al (Klein et al 2008, Klein et al 2007)
Time from randomization to second progression	Stratified log-rank test using site Investigator assessments
Change from baseline in key symptoms (EORTC QLQ-C30 and QLQ-LC13)	Mixed-model repeated measures analysis
GHS/QoL/Function improvement rate (EORTC QLQ-C30 endpoints)	Logistic regression
Symptom improvement rate (EORTC QLQ-C30 and QLQ-LC13 endpoints)	Logistic regression
Time to GHS/QoL/Function deterioration (EORTC QLQ-C30 endpoints)	Stratified log-rank test
Time to symptom deterioration (EORTC QLQ-C30 and QLQ-LC13 endpoints)	Stratified log-rank test
Treatment-related symptoms (PRO-CTCAE and PGIS)	Presented using summaries and descriptive statistics

BICR Blinded Independent Central Review; CTCAE Common Terminology Criteria for Adverse Event; EORTC European Organisation for Research and Treatment of Cancer; GHS/QoL Global health score/Quality of life; ORR Objective response rate; OS Overall survival; PFS Progression-free survival; PGIS Patient's Global Impression of Severity; PRO Patient-reported outcome; QLQ-C30 30-Item core quality of life questionnaire; QLQ-LC13 Lung cancer module.

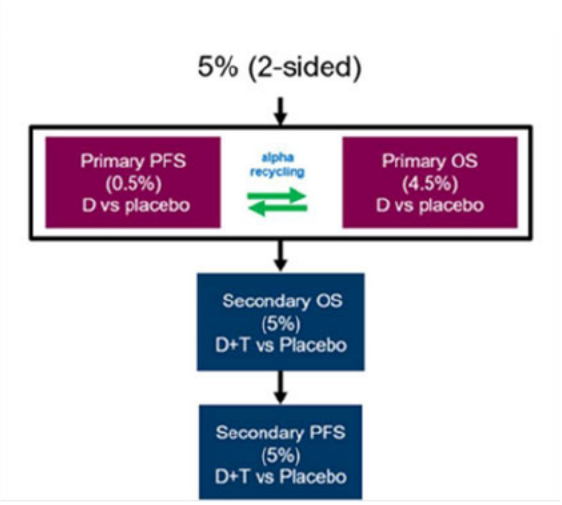
4.2.1 Methods for multiplicity control

The following multiple testing procedure shown in Figure 2 will be used to strongly control the family-wise type I error rate (alpha) at 5% (2-sided) for testing the following primary and key secondary endpoints:

- Dual primary endpoints: PFS and OS for durvalumab monotherapy versus placebo.
- Key secondary endpoint: PFS and OS for durvalumab in combination with tremelimumab versus placebo

The testing procedure is hierarchical in that it starts with testing the 2 dual primary endpoints of PFS (per BICR) and OS as outlined in Figure 2. The overall 5% type I error (2-sided) is split among the 2 dual primary endpoints. An alpha level of 0.5% is allocated to PFS and will be split between 2 potential analysis timepoints (1 interim plus 1 primary analysis). An alpha level of 4.5% is allocated to the OS and will be split between 4 potential analysis timepoints (2 interim analyses, 1 primary analysis, plus 1 potential additional assessment of OS at time of primary PFS analysis if OS-IA2 timing does not coincide with the primary PFS analysis timing, as described below). The 0.5% alpha level for PFS is controlled at the time of the interim and primary analyses using the Lan-DeMets spending function that approximates an O'Brien Fleming approach as described in Section 5. An alpha of 0.01% (2-sided) will be allocated to the potential additional OS analysis that will be conducted at the time of PFS primary analysis if OS-IA2 timing does not coincide with the primary PFS analysis timing. The remaining 4.49% alpha level for OS will be controlled at the time of planned OS-IA1, OS-IA2 and the primary analysis using the Lan-DeMets spending function that approximates an O'Brien Fleming approach as described in Section 5.

Figure 2 Multiple testing procedure for primary and key secondary endpoints



D durvalumab; OS overall survival; PFS progression-free survival; T tremelimumab.

If any of the PFS analyses are significant, then the allocated test mass (0.5%) can be recycled to OS giving a total test mass of 5% for OS. In this case the 5% will be split between OS-IA1, OS-IA2, and primary OS analyses using the LanDeMets spending function with an alpha of 0.01% (2-sided) allocated to the additional analysis of OS that may occur at the time of the PFS primary analysis if OS-IA2 timing does not coincide with the primary PFS analysis timing. If neither of the PFS analyses are significant at the 0.5% level but OS is significant at the 4.5% level, at either an interim or primary analysis, then the 4.5% can be recycled to test PFS at 5% (split between the interim and primary analyses). PFS can only be re-tested using the information available at the time of the original PFS analysis.

The alpha levels at each of the planned interim analyses will be determined at the time of each analysis using the Lan-DeMets spending functions. This will be derived based on the information fraction i.e. actual number of events available at the time of the analysis, out of the total final number of events expected for PFS and OS respectively. The alpha level applied at the primary PFS and OS analyses will be adjusted (using a generalized Haybittle-Peto method (SAS Institute Inc. 2018)) to account for the actual alpha spent at the interim analyses based on the actual final total number of events, to maintain control of the overall Type I error.

If both PFS and OS are statistically significant for the primary endpoint analyses comparing durvalumab monotherapy versus placebo, then the 5% alpha can be carried down to test durvalumab in combination with tremelimumab versus placebo. OS will be tested at 5% and then, if significant PFS will be tested at 5%.

A detailed calculation of alpha allocation at interim and primary analyses is provided in Appendix A.

4.2.2 Progression-free survival

The primary PFS analysis will be based on RECIST 1.1 using BICR tumor assessments. The analysis will be performed in the FAS using a stratified log-rank test (using the TEST statement in PROC LIFETEST) adjusting for TNM stage (I/II versus III) and receipt of PCI (yes versus no) for generation of the p-value and using the Efron approach for handling ties (Efron, 1977). The effect of durvalumab monotherapy and durvalumab and tremelimumab combination therapy versus placebo treatment will be estimated by the hazard ratio (HR) together with its corresponding confidence interval (CI) (95% and [1-adjusted alpha] × 100%) 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.

In order to ensure there are at least 5 events within each strata; if there are too few events observed in the TNM Stage I/II stratification level then TNM stage may be excluded from the stratified models leaving receipt of PCI as the sole stratification factor.

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

Secondary PFS analysis will be performed using the same methodology as for the primary analysis.

Supportive summaries/graphs

Kaplan-Meier plots of PFS will be presented by treatment group. 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 together with its corresponding 95% CI for each treatment group. The CI for median progression-free survival will be derived based on Brookmeyer-Crowley method with log-log transformation (SAS Institute Inc. 2018).

The assumption of proportionality will be assessed for each treatment comparison firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time-dependent covariate to assess the extent to which this represents random variation. If lack of proportionality is found, this may be a result of treatment-by-covariate interactions.

If non-proportional hazards are observed, this may be investigated further. Under non-proportional hazards, the HR from the primary analysis can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. However, under non-proportional hazards the variation in treatment effect may be described by presenting piecewise HR calculated over distinct time-periods: 0-6m, 6-12m, 18-24m, >24m, using a Cox regression model with time-dependent covariates, stratified by the same factors as the primary analysis. Under non-proportional hazards, the Restricted Mean Survival Time (RMST) may also be analysed up to the minimum of the largest observed event time in each of the two arms, using the pseudovalues approach (Andersen et al. 2004) an area-under-the-curve approach (Kaplan-Meier method), with standard error, for each treatment group, along with the estimate of difference in means (and ratio of means) between treatment groups, with 95% confidence intervals and p-value. The standard error of the difference will be estimated using Generalised Estimating Equations (PROC GENMOD, Klein et al. 2008), with covariate adjustment for stratification factors. The standard error for the RMST ratio (unadjusted) may be estimated from the treatment group RMSTs using the delta method. In addition, an area-under-the-curve approach (Kaplan-Meier method) the pseudovalues approach (Andersen et al. 2004) and Royston-Parmar model (Royston and Parmar 2011, 2013) may also be used; all RMST analyses [except RMST ratio] will control for the stratification factors used in the primary analysis. Additional analyses such as the MaxCombo test may also be performed.

The treatment status at progression of patients at the time of analysis will be summarized. 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. 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 summarized by treatment group. 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 tumor assessment interval plus 2 weeks (10 weeks if time period between randomization and DCO for that patient is 72 weeks or less; 14 weeks if that time period is 72 weeks onwards up to 96 weeks; 26 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 summarized using median time from randomization to date of censoring (date last known to be non-progressor) 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 assessment prior to progression will be presented for each treatment group.

Summaries of the number and percentage of patients who miss 2 or more consecutive RECIST assessments will be presented for each treatment group.

All of the collected RECIST 1.1 data will be listed for all randomized patients. In addition, a summary of new lesions (i.e. sites of new lesions) will be produced. All BICR new lesions are to be medically reviewed and categorized for the summary.

Sensitivity analyses

The following sensitivity analyses will be performed between durvalumab monotherapy and durvalumab and tremelimumab combination therapy versus placebo.

- Evaluation-Time bias

A sensitivity analysis will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled time points. The midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment) will be analyzed 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 group. This approach will use the BICR RECIST assessments.

- Attrition bias

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

absence of progression immediately following two, or more, non-evaluable tumor assessments will be included. In addition, patients who take subsequent therapy (note radiotherapy is not considered a subsequent anti-cancer therapy) prior to their last RECIST assessment or progression or death will be censored at their last assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier plot of the time to censoring from the primary analysis where the censoring indicator of the PFS analysis is reversed. This approach will use the BICR RECIST assessments.

- Ascertainment bias

Ascertainment bias will be assessed by analyzing the site investigator data. The stratified log rank test will be repeated on the programmatically derived PFS using the site investigator data based upon RECIST. The HR and CI will be presented.

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

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

- Deviation bias

Deviation bias may be assessed by repeating the PFS analysis excluding patients with deviations that may affect the efficacy of trial therapy (see Section 2.2).

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided for each treatment comparison to compare the primary and sensitivity analyses of progression-free survival.

- Duration of follow-up

Duration of follow-up will be summarized using medians (as well as minimum and maximum):

- Time from randomization to the date of PFS event for patients who experience PFS events or to the date of censoring summarized in all patients regardless of treatment arm, as well as by treatment arm;
- Time from randomization to the date of censoring in censored patients only, presented by treatment arm

Subgroup analyses

Subgroup analyses will be conducted comparing PFS (per RECIST 1.1 using BICR assessments) between durvalumab monotherapy and durvalumab and tremelimumab combination therapy versus placebo in the following subgroups of the FAS and combination analysis set respectively (but not limited to):

- TNM stage (Stage I/II, III) – based on both IVRS and eCRF data
- Receipt of PCI (yes, no) – based on both IVRS and eCRF data
- Time from end date of cCRT to randomization in this study (<14 days (2 weeks), ≥14 to <28 days (4 weeks), ≥28 days)
    - End date of cCRT is defined as the last date of last cycle of chemotherapy (last date of platinum chemotherapy dosing + cycle length) **or** the last dose of radiotherapy, whichever is later
- Time from last dose of radiotherapy to randomization in this study (<28 days (4 weeks), ≥28 to <56 days (8 weeks), ≥56 days to <84 days (12 weeks), ≥ 84 days
- Prior platinum chemotherapy (carboplatin, cisplatin)
    - This will be determined from the coded text of "Cancer therapy agent" (CAPRX module) on the eCRF at screening and be based on cycle 1 of treatment.
- Prior radiotherapy regimen (daily, twice daily)
    - This will be determined from the response to "Frequency of fraction doses per day" (CAPRXR module) on the eCRF at screening (see Section 3.4.8 for split dosing rules).
- Best response to cCRT (CR, PR, SD)
    - This will be determined from the response to "Best response" of the last cCRT entry ("Concomitant chemoradiotherapy" =Yes) prior to randomization (CAPRX module) on the eCRF at screening.
- Sex (male, female)

- Age at randomization (<65, ≥65 years of age)
  - This will be determined from the date of birth (BRTHDAT in the DM module) and date of randomization (IERNDDAT in the IE module) on the eCRF at screening. Patients with a missing age value (after imputation in Section 4.1.3) will be included using the mean age (overall FAS) and categorized accordingly.
- PD-L1 status (TC and IC < 1%, TC or IC ≥ 1%)
  - This will be derived from third party vendor data.
- Smoking status (smoker [current/former smoker], non-smoker [never smoker])
  - This will be determined from the response to "Has the subject ever used the substance" (SU_NIC module) on the eCRF at screening. Patients with a missing smoking status will be included in the "smoker" category.
- Race (White, Black/African-American, Asian, Other [Native Hawaiian/Pacific Islander or American Indian/Alaska Native or Others])
  - This will be determined from the response to "Race" (DM module) on the eCRF at screening.
- Geographic region (Asia, Europe, South America, North America)
  - This will be determined from the center number (CENTRE). If there are less than 20 events across both treatment groups in the "South America" category, these patients will be combined with those in North America.
- WHO/ECOG performance status at baseline (Normal activity [PSTAT=0], Restricted activity [PSTAT=1])
  - This will be determined from the response to "Performance status" (PSTAT module) on the eCRF at screening. Patients with a missing performance status will be included in the "Restricted activity" category.

The subgroup analyses for the stratification factors will be based on the values entered into the IVRS and repeated for the values recorded on the eCRF. All other factors will be based on values recorded on the eCRF as indicated above. Patients with missing data for a subgroup variable will be excluded from the analysis for that subgroup only.

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

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

For each subgroup, the HR (durvalumab monotherapy or durvalumab and tremelimumab combination therapy: placebo) and 95% CI will be calculated from an unstratified Cox proportional hazards with treatment as the only covariate. 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.

For each treatment comparison, these hazard ratios and associated two-sided 95% CIs will be summarized and presented on a forest plot, 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 across both treatment groups within a subgroup), the relationship between that subgroup and PFS will not be formally analyzed. In this case, only descriptive summaries will be provided.

Effect of covariates on the HR estimate

Cox proportional hazards modelling will be employed to assess the effect of covariates on the HR estimate for the treatment comparisons of primary analysis (durvalumab monotherapy and durvalumab and tremelimumab combination therapy versus placebo). Before embarking on more detailed modelling, an initial model will be constructed, containing treatment and the stratification factors alone, to ensure any output from the Cox modelling is likely to be consistent with the results of the stratified log-rank test.

The result from the initial model and the model containing additional covariates will be presented.

Additional covariates for this model will be sex, age at randomization, smoking status, WHO/ECOG performance at baseline, region, race, time from last dose of cCRT to randomization, prior platinum chemotherapy, prior radiotherapy regimen and best response to cCRT.

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

Consistency of treatment effect between subgroups

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

A global interaction test will be performed by comparing the fit of a Cox proportional hazards model including treatment, the stratification variables and stratification variables by treatment

interactions, with the fit of a model excluding the interaction terms, and will be assessed at the 2-sided 10% significance level. Ties will be handled using the Efron method.

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 factor will also be tested to rule out any qualitative interaction using the approach of Gail and Simon (Gail and Simon 1985).

This test will be performed for each treatment comparison of the primary analysis (durvalumab monotherapy and durvalumab and tremelimumab combination therapy versus placebo) separately.

Time to first subsequent therapy or death (TFST)

For supportive purposes, the time to the start of subsequent therapy will be analyzed 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, a Kaplan-Meier plot of the time to the start of subsequent therapy will be presented by treatment group. This will be summarized per treatment group, 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 cancer therapy, a summary table of first subsequent cancer therapies by treatment group will be provided; response to first subsequent therapy by treatment group will be included in a listing.

The number of patients who receive a second subsequent anti-cancer therapy (3rd line therapy) will be reviewed and if sufficient data, summaries of time from randomization to second subsequent therapy may be produced.

Radiotherapy will not be considered a subsequent therapy.

Exploratory analysis

An exploratory analysis of PFS using the irRECIST 1.1 data obtained from the BICR may be performed using the same methodology and model as that used for the primary analysis of PFS. The stratified log-rank test will be repeated on PFS using the BICR based upon the irRECIST 1.1 data. The HR and CI will be presented. No sensitivity or subgroup analyses will be performed and results will be reported outside of the CSR.

4.2.3 Overall survival

The primary analysis of OS in the FAS will be analyzed using stratified log-rank tests using the same methodology as described for the PFS endpoint.

The treatment effect will be estimated by the HR together with its corresponding CI (95% and [1-adjusted alpha] × 100%) from a stratified Cox proportional hazards model.

Kaplan-Meier plots of OS will be presented by treatment group. 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 together with its corresponding 95% CI for each treatment group. The CI for median overall survival will be derived based on Brookmeyer-Crowley method with log-log transformation (SAS Institute Inc. 2018).

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

Sensitivity analysis

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

The number of patients prematurely censored will be summarized by treatment group. 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 summarized using medians (as well as minimum and maximum):

- Time from randomization to the date of death (i.e. overall survival) or to the date of censoring (date last known to be alive, for censored patients) summarized in all patients regardless of treatment arm, as well as by treatment arm.
- Time from randomization to the date of censoring in censoring patients only, presented by treatment arm.

Duration of OS follow-up will also be summarized using median (and interquartile range), estimated by analyzing time to censoring with the reverse Kaplan-Meier method (i.e. OS censoring and event flags reversed), presented by treatment arm and overall.

Subgroup analyses and a forest plot will be generated comparing OS between treatments of primary analysis in the same way as previously specified for PFS.

No adjustment to the significance level for testing will be made since all these subgroup analyses will be considered supportive of the primary analysis of OS.

The effect of covariates upon the HR estimate and the consistency of treatment effect between subgroups will be analyzed for OS with regards to the treatment comparison of primary analysis, using the same methods as those described for PFS.

4.2.4 Objective response rate

The ORR will be based on the programmatically derived RECIST 1.1 using BICR assessments for unconfirmed and confirmed responses. The analysis will be performed using a Cochran-Mantel-Haenszel (CMH) test, stratified using the same stratification factors as for the PFS endpoint. The results of the analysis will be presented in terms of the difference in the proportion of patients with a response, together with associated 95% CIs and p-value (2-sided). The confidence intervals for the difference in proportions between treatment groups will be computed using Miettinen and Nurminen's (MN) stratified confidence limits. This analysis will be performed in a subset of the FAS, including all patients with measurable disease at baseline.

A sensitivity analysis will be performed, repeating the analysis of ORR using the site Investigator assessments based on RECIST 1.1, for patients with confirmed and unconfirmed responses.

An exploratory analysis of ORR using the irRECIST 1.1 data obtained from the BICR assessments may be performed where the above analysis will be repeated.

Summaries will be produced that present the number and percentage of patients with a tumor response (CR/PR) based upon the number of patients with measurable disease at baseline per BICR/investigator as appropriate.

For each treatment group, BoR will be summarized by n (%) for each category (CR, PR, SD, PD, and NE). This will be produced for the BICR assessment of RECIST only. No formal statistical analyses are planned for BoR. Analyses will be repeated for patients with confirmed and unconfirmed responses.

Overall visit response data will be listed for all patients in the FAS.

4.2.5 Duration of response

Descriptive data will be provided for the DoR in responding patients (i.e. median duration of response and 95% CIs) by treatment group, including the associated Kaplan-Meier curves (without any formal comparison of treatment groups or p-value attached). Analyses will be repeated for patients with confirmed and unconfirmed responses.

4.2.6 Progression-free survival at 18 months and 24 months

The PFS18 and PFS24 (where 1 month equates to 30.4375 days) will be summarized (using the Kaplan-Meier curve) and presented by treatment group. For each treatment group, the PFS18 and PFS24 based on Kaplan-Meier method will be presented, along with their 95% CIs. The computation of the CIs will be based on a log(-log(.)) transformation.

4.2.7 Time to death or distant metastasis

TTDM will be analyzed using identical methods as outlined for the analysis of PFS (i.e. a stratified log-rank test) and adjusting for the same set of covariates, but no subgroup analyses will be performed. Medians and Kaplan Meier plots will be presented to support the analysis. The sensitivity analyses outlined in Section 4.2.2 will not be repeated for TTDM. A sensitivity analysis of TTDM may be performed using Investigator assessments according to RECIST 1.1.

4.2.8 Proportion of patients alive at 24 and 36 months

The OS24 and OS36 (where 1 month equates to 30.4375 days) will be summarized (using the Kaplan-Meier curve) and presented by treatment group. For each treatment group, the OS24 and OS36 based on Kaplan-Meier method will be presented, along with their 95% CIs. The computation of the CIs will be based on a log(-log(.)) transformation.

4.2.9 Time from randomization to second progression

PFS2 in the FAS will be analyzed using a stratified log-rank test, using the same methodology as described for the PFS endpoint. The treatment effect will be estimated by the HR together with its corresponding 95% CI from a stratified Cox proportional hazards model. Medians and Kaplan-Meier plots will be presented by treatment group.

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

4.2.10 Change in tumor size

The absolute values, change in TL tumor size from baseline and percentage change in TL tumor size from baseline will be summarized using descriptive statistics and presented at each timepoint and by treatment group. The best change in target lesion tumor size from baseline, (where best change in target lesion size is the maximum reduction from baseline or the minimum increase from baseline in the absence of a reduction) will also be summarized and presented by randomized treatment group.

Tumor size will also be presented graphically using waterfall plots for each treatment group, to present each patient's best percentage change in tumor size as a separate bar, with the bars ordered from the largest increase to the largest decrease. Reference lines at the +20% and -30% change in TL tumor size level will be added to the plots, which correspond with the definitions of progression and 'partial' response. All progressions will be marked with a '●'. The scale in these plots will be fixed to be from -100 to 100 to avoid presenting extreme values. Values that are capped as a result of this restriction to the scale are marked with '#'. Values are ordered in descending order with the imputations due to death appearing first followed by a gap followed by all other patients. On each of the waterfall plots the TNM stage (I/II versus III) and receipt of PCI (yes versus no) of each patient will be indicated. Additional waterfall plots showing percentage change in tumor size at specific timepoints may be produced if it is felt that these are warranted to provide greater clarity.

The above outputs will be programmed for the BICR data based upon RECIST assessments.

4.2.11 Patient reported outcomes

4.2.11.1 EORTC QLQ-C30 and QLQ-LC13

The PRO endpoints identified as primary are the following 5 lung cancer symptoms:

- Dyspnea: 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, role functioning and GHS/QoL domains of the EORTC QLQ-C30 are furthermore pre-specified endpoints of interest.

Mixed-model repeated measures (MMRM) analysis

Change from baseline, using the MMRM analysis, will be examined for dyspnea, cough, and chest pain scores as assessed by the EORTC QLQ-LC13, and GHS/QoL, physical functioning, role functioning, fatigue and appetite loss as assessed by the EORTC QLQ-C30 and as detailed above. The MMRM analysis will examine the change from baseline in the scores for each assessment time point up to PD, death or 24 months. The analysis will be to compare the average treatment effect from the point of randomization until PD, death or 24 months unless there is excessive missing data (defined as missing in >75% of randomized patients in any treatment arm).

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 GHS/QoL 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 analyzed using the above method. Each visit will be assigned a sequential number. The time from randomization to each of these will be derived to select only those visits occurring within the first 24 months of randomization.

As an example, say a patient X attends the first 4 scheduled visits of a weekly schedule and then discontinues treatment, whilst patient Y discontinues treatment after the first scheduled visit, the first 6 generic visits would be as shown in Table 8:

Table 8 Generic visit variable for MMRM analysis

Generic visit	Study day (week)	
	Patient X	Patient Y
Baseline	Baseline	Baseline
1	8	7
2	15	11 (discontinuation)
3	22	22
4	29	29
5	33 (discontinuation)	36
6	43	43

The MMRM model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, TNM stage (I/II versus III) and receipt of PCI (yes versus no) as well as the continuous fixed covariate of baseline score and the 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 and p-value.

An unstructured covariance matrix will be used to model the within-patient 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.

Time to deterioration

Time to symptom, functioning and GHS/QoL deterioration will be analyzed for each of the symptom scales/items, functional scales, and GHS/QoL in EORTC QLQ-C30 and QLQ-LC13. This will be achieved by comparing between treatment groups 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 GHS/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 group.

Symptom and function and GHS/QoL 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 improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and GHS/QoL will be produced.

Symptom, functional and GHS/QoL improvement rates will be analyzed by comparing between treatment groups using a logistic regression model adjusting for the same factors as the PFS endpoint. The results of the analysis will be presented in terms of an odds ratio 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). The odds ratio and 95% CI for each scale/item will be presented graphically on a forest plot. If there are very few responses in 1 treatment group, a Fisher's exact test will be considered.

Change from baseline

Summaries of original and change from baseline values of each symptom scale/item, the GHS/QoL score, and each functional domain will be reported by assessment timepoint for each treatment group. 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 improvement, no change, and deterioration as defined in Table 6) will also be produced for each treatment group.

A summary of compliance rate and evaluability rate will be provided for each treatment group, by assessment time point and also for overall. The summary will be provided for the FAS, combination analysis set, and patients who have discontinued treatment.

4.2.11.2 EQ-5D-5L

Descriptive statistics will be calculated for each scheduled time point in the study, for each treatment group and as a total. These will report the number of patients, the number of EQ-5D-5L questionnaires completed at each visit, the number and proportion responding to each dimension of the EQ-5D-5L. Additionally, summary statistics will be reported for the EQ-5D-5L index score and the EQ-5D-5L VAS score, and the change from baseline for the index and VAS scores.

A summary of compliance rate and evaluability rate will be provided for each treatment group, by assessment time point and overall.

To support submissions to payers, additional analyses may be undertaken, and these will be outlined in a separate Payer Analysis Plan.

4.2.11.3 PGIS

The number and percentage of patients in each category of the PGIS responses will be summarized by treatment group, at each assessment time point and overall.

4.2.11.4 PRO-CTCAE

The number and percentage of patients in each category of the responses for each PRO-CTCAE item will be summarized by treatment group and assessment time point.

4.2.12 Health care resource use

The potential impact the disease and treatment have on health care resource use will be analyzed for the purposes of submissions to payers. Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) will be provided for each treatment group on the different types of hospital admissions, the length of stay of people admitted in to hospital for at least one overnight stay and length of stay of people admitted to intensive care / high dependency units, as well as the primary sign or symptom the patient presents with. To support submissions to payers, additional analyses may be undertaken, and these will be outlined in a separate Payer Analysis Plan.

4.2.13 Safety data

Safety and tolerability data from all cycles of treatment will be combined and will be presented by treatment group using the safety population. Safety summaries will be descriptive only. No formal statistical analyses will be performed on the safety data.

The following sections describe the planned safety summaries for AEs, laboratory parameters, vital signs, ECG and WHO/ECOG performance status. However, additional safety summaries (not specified in this SAP) may need to be produced to aid interpretation of the safety data.

4.2.13.1 Adverse events

All AEs, both in terms of current MedDRA preferred term and CTCAE grade, will be summarized descriptively by count (n) and percentage (%) for each treatment group. The current MedDRA dictionary at the time of database lock will be used for coding wherever possible. The majority of the AE summaries, unless stated otherwise, will be based on TEAEs. Any AE occurring before study treatment (i.e. before the administration of the first dose 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 is ongoing and increases in severity after the first dose will be regarded as treatment emergent and thus will be included in the majority of summary tables.

AEs observed up until 90 days following discontinuation of the study treatment or until the initiation of the first subsequent anti-cancer therapy following discontinuation of study treatment (whichever occurs first) will be used for reporting of all the AE summary tables. 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 may also be produced containing AEs observed up until 90 days following discontinuation of the study treatment (i.e. without taking subsequent therapy into account).

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 study drug. Frequencies and percentages of patients reporting each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from on the episode level summaries which may be produced).

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

- All AEs
- All AEs possibly related to study medication (as determined by the reporting investigator)
- AEs with a maximum CTCAE grade 3 or 4
- AEs of any CTCAE grade 3 or 4
- AEs with a maximum CTCAE grade 3 or 4, possibly related to study medication (as determined by the reporting investigator) – causality to be determined first, then max grade 3 or 4
- AEs with any CTCAE grade 3 or 4, possibly related to study medication (as determined by the reporting investigator) – causality to be determined first, then maximum grade 3 or 4
- AEs with outcome of death
- AEs with outcome of death possibly related to study medication (as determined by the reporting investigator)
- All SAEs
- All SAEs possibly related to study medication (as determined by the reporting investigator)
- AEs leading to discontinuation of study medication

- AEs leading to discontinuation of study medication, possibly related to study medication (as determined by the reporting investigator)
- AEs leading to hospitalization
- AEs leading to dose interruption of study medication
- Immune mediated AEs (as determined by the reporting investigator)
- Infusion reaction AEs (as determined by the reporting investigator)

An overall summary of the number and percentage of patients in each category will be presented. In addition, a truncated AE table of most common AEs and another table showing most common AEs with CTCAE grade 3 or 4 (maximum grade and any), showing all events that occur in at least 5% of patients overall will be summarized by preferred term, 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 summarized by preferred term within each system organ class. For each preferred term, the event rate is defined as the number of patients with that AE divided by the total duration (days) 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, system organ class, preferred term and treatment group.

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

Deaths

A summary of all deaths will be provided with number and percentage of patients by treatment group in the FAS, categorized 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 treatment, or

up to the date of initiation of the first subsequent therapy (whichever occurs first)

- AE onset after start of subsequent therapy, which includes AEs with onset date >90 days following the last dose of study treatment 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 (as defined above)
  - AE onset after start of subsequent therapy (as defined above)
- Death after end of safety follow up period (last dose of study treatment + 90 days) and not due to AE or disease under investigation (*)
- Unknown reason for death
- Other deaths (deaths not captured in the earlier categories) e.g. death occurring after randomization but prior to receiving treatment and which is not related to disease under investigation

This summary will be produced twice; firstly, including all deaths and, secondly, including all deaths on-treatment or within 90 days of the last dose of study treatment. Hence the category marked (*) will only appear in the first summary.

Adverse events of special interest and possible interest

Preferred terms used to identify AESI and AEPI 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 grouping. For each grouped term, the number (%) of patients experiencing any of the specified terms will be presented by maximum CTCAE grade. Time to onset of first AESI for each grouped term will also be produced. 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.

Additional summaries of the above-mentioned grouped AE categories will include number (%) of patients who have:

- At least one AESI/AEPI presented by outcome
- At least one AESI/AEPI leading to discontinuation of study medication

A summary of total duration (days) of AESI/AEPI will be provided for events which have an end date, and these may be supported by summaries of ongoing AESIs at death and, separately, at DCO.

Additionally, there will be several summaries of AESIs/AEPIs requiring concomitant treatment, and particularly the relationship of AESIs/AEPIs to the use of immunosuppressive agents (i.e. depicting which AESI/AEPI triggered immunosuppressive use), endocrine treatment and systemic corticosteroids, and, separately, to the use of systemic corticosteroids at high doses.

4.2.13.2 Laboratory assessments

Summaries of laboratory assessments will include the variables listed in Tables 8, 9 and 10 of the CSP (Version 4, 13 November 2020). Calculated creatinine clearance values reported by the site (e.g. at the screening visit) will be summarized separately from programmatically derived CrCl.

Data obtained up until the 90 days following discontinuation of study treatment or until the initiation of the first subsequent therapy 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 a number of 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, summaries of laboratory data may also be produced containing data collected up until 90 days following discontinuation of the study treatment. Any data post 90 days last dose will not be summarized.

Data summaries will be provided in preferred units.

Scatter plots (shift plots) of baseline to maximum value/minimum value (as appropriate) on treatment (i.e. on-treatment is defined as data collected between the start of treatment and the relevant follow-up period following the last dose of study treatment) may be produced for certain parameters if warranted after data review.

For continuous laboratory assessments, absolute value and change from baseline will be summarized 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 increased and decreased directionality of change will be produced. The laboratory parameters for which CTCAE grade shift outputs will be produced are:

- Hematology: Hemoglobin - increased and decreased, Leukocytes - increased and decreased, Absolute Lymphocyte Count – increased and decreased, Absolute Neutrophils Count - decreased, Platelets - decreased

- Clinical chemistry: ALT - increased, AST - increased, ALP - increased, Total bilirubin - increased, Albumin - increased, Magnesium – increased and decreased, Sodium –

increased and decreased, Potassium – increased and decreased, Corrected calcium – increased and decreased, Glucose – increased and decreased, Creatinine - increased

Additional summaries will include a shift table for urinalysis (Bilirubin, Blood, Glucose, Ketones, Protein) comparing baseline value to maximum on-treatment value.

Liver enzyme elevations and Hy's law

The following summaries 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 concurrent or preceding the date of Total Bilirubin elevation
- Narratives will be provided in the CSR for patients who have ALT ≥ 3x ULN plus Total bilirubin ≥ 2x ULN or AST ≥ 3x ULN plus Total bilirubin ≥ 2x ULN, where the onset date of ALT or AST elevation should be concurrent or preceding the date of Total Bilirubin elevation.

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) (at any time) will be plotted. Individual patient data where ALT or AST (i.e. ≥ 3x ULN) plus Total bilirubin (i.e. ≥ 2x ULN) are elevated at any time will be listed also.

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 in the vertical axis.

Plots of ALT, AST, ALP and Total bilirubin, baseline vs. maximum observation on treatment by treatment group will also be produced. Plots will be repeated removing any outliers, as defined by the medical team.

Assessment of thyroid function test results

For the thyroid function tests (TSH, T3 and T4), will be based on data up to 90 days after the last dose of study medication or date of initiation of subsequent therapy (whichever occurs first).

Absolute value and change from baseline will be summarized using descriptive statistics at each scheduled assessment time point by actual treatment group.

Shift tables showing baseline to maximum and baseline to minimum will also be produced for TSH, T3 and T4.

4.2.13.3 ECGs

ECG data obtained up until the 30-day safety follow-up visit will be included in the summary tables.

Overall evaluation of ECG is collected at each visit in terms of normal or abnormal, and the relevance of the abnormality is termed as “clinically significant” or “not clinically significant”. A shift table of baseline evaluation to worst evaluation will be produced, if a sufficient number of ECG assessments are recorded.

4.2.13.4 Vital signs

Vital signs data obtained up until the 30-day safety follow-up visit will be included in the summary tables.

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

4.2.13.5 Physical examination

All individual physical examination data will not be summarized.

4.2.13.6 Other safety data

Data from positive pregnancy tests will not be summarized.

4.2.14 WHO/ECOG performance status

All WHO/ECOG performance status will be summarized over time for the FAS.

4.2.15 Pharmacokinetic data

All measured and collected PK concentration data will be listed for each patient and each dosing day by treatment group. PK concentration data will also be summarized by treatment group in tables and figures based on the PK analysis set.

At a time point where less than or equal to 50% of the concentration values are NQ, all NQ values will be set to the LLOQ, and all descriptive statistics will be calculated accordingly.

At a time point where more than 50% (but not all) of the values are NQ, the gmean, gmean ± gSD and gCV% will be set to NC. The maximum value will be reported from the individual data, and the minimum and median will be set to NQ.

If all concentrations are NQ at a time point, no descriptive statistics will be calculated for that time point. The gmean, minimum, median and maximum will be reported as NQ and the gCV% and gmean ± gSD as NC.

4.2.16 Immunogenicity data

The number and percentage of patients who develop detectable ADA to durvalumab and to tremelimumab within each ADA response category listed in Section 3.5.2 will be summarized by treatment group based on the relevant ADA analysis set. Median and range of maximum titer for each category will be provided. The presence of nAb will be reported for samples confirmed positive for the presence of ADA.

The effect of immunogenicity as well as the effect of its neutralizing properties on PK, pharmacodynamics, efficacy, and safety may be evaluated by descriptive summaries, which will be presented by ADA status (positive/negative) in each treatment group, if the data allow.

All measured and collected immunogenicity results will be listed for each patient and each dosing data by treatment group. ADA titer and nAb data will be listed for samples confirmed positive for the presence of ADA. AEs in ADA positive patients by ADA positive category will be listed.

4.2.17 Pharmacokinetic/pharmacodynamic relationships

If the data are suitable, the relationship between PK exposure and efficacy/safety parameters may be investigated graphically or using an appropriate data modeling approach. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee and outside of the scope of CSR so will be reported separately if applicable.

4.2.18 Biomarker data

The relationship of PD-L1 expression (secondary objective) and, if applicable, of exploratory biomarkers (such as TMB) to clinical outcomes (including but not restricted to) of PFS, OS, and ORR will be assessed in the evaluable populations for each biomarker.

PFS and OS will be analyzed in the FPAS and CPAS using the same methods as for the primary PFS and OS analyses. Analyses of ORR using BICR assessments for a confirmed response will be performed in the FPAS and CPAS using a stratified Cochran-Mantel-Haenszel test. The CIs for the difference in the proportion of patients with a response between treatment groups will be computed using Miettinen and Nurminen's stratified confidence limits.

Subgroup analyses will also be performed to estimate treatment effect of PFS, OS and ORR in PD-L1 expression high (TC or IC ≥ 1%) and low (TC and IC <1%) subgroups. These analyses will be performed using the same methodologies as the analyses conducted in FPAS and

CPAS for PD-L1 but unstratified, meaning that treatment will be the only covariate in all the models.

A separate SAP will be prepared for more comprehensive biomarker analyses.

4.2.19 Demographic and baseline characteristics data

The following will be summarized for all patients in the FAS (unless otherwise specified) by treatment group:

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations
- Inclusion in analysis populations
- Demographics (age, age group [<50, ≥50-< 65, ≥ 65 - <75 and ≥ 75 years], sex, race and ethnicity)
- Patient characteristics at baseline (height, weight, weight group [<70, ≥70 - ≤90, >90], BMI, BMI group [<18.5, ≥18.5 - <25, ≥25 - <30, ≥30])
- Patient recruitment by region, country and center
- Previous treatment modalities
- Number of regimens of previous chemotherapy at baseline
- Previous lung cancer chemotherapy
- Pre-treatment radiotherapy
- Disease characteristics at baseline (WHO/ECOG performance status, primary tumor location, histology type, AJCC disease stage, best response to cCRT, TNM disease classification and PD-L1 status)
- Extent of disease at baseline
- Medical history (past and current)
- Relevant surgical history
- Time from last dose of cCRT to randomization
- Disallowed concomitant medications
- Allowed concomitant medications

- Post-discontinuation cancer therapy
- Nicotine use, categorized (never, current, former)
- Stratification factors at randomization as per IVRS and eCRF data
- Stratification factors at randomization by IVRS versus eCRF
- WHO/ECOG performance status

The WHO drug dictionary (WHODD) will be used for concomitant medication coding.

4.2.20 Treatment exposure

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

- Total exposure of each treatment group.
- Actual exposure of each treatment group.
- Total number of cycles received.
- Reasons for dose delays and infusion interruptions. Dose delays and infusion interruptions will be based on investigator initiated dosing decisions.
- Number of dose delays and duration of delays. In addition, delays due to AEs and due to reasons other than AEs will be summarized separately.
- Number of infusions received.
- RDI of each treatment group.

For patients on study treatment at the time of the PFS and OS analysis, the DCO date will be used to calculate exposure.

4.2.21 Subsequent Therapy

Subsequent systemic anticancer therapies received after discontinuation of study treatment will have summaries produced by treatment group. Radiotherapy will not be considered a subsequent therapy and a separate summary and listing will be produced for radiotherapy received after discontinuation of study treatment. Therapy received on the same day as discontinuation of study treatment will be considered to be subsequent therapy. The subsequent therapies are to be categorised by medical team.

4.2.22 Coronavirus Disease 2019 (COVID-19)

A listing of all subjects affected by the COVID-19 pandemic, and subjects with reported issues in the Clinical Trial Management System due to COVID-19 pandemic will be generated. In addition, all COVID-19 related non-important PDs and issues will be summarized and listed. Additional analyses might be conducted to investigate the impact of COVID-19 on study endpoints.

5. INTERIM ANALYSES

5.1 Analysis Methods

Up to four interim analyses are planned: 1 for PFS and 3 for OS (or 2 for OS if OS-IA2 timing coincides with the timing of the PFS primary analysis).

The planned PFS-IA will occur when approximately 308 PFS BICR events have occurred in the durvalumab monotherapy and placebo treatment groups (58.8% maturity). OS-IA1 will occur at the same time when it is anticipated that approximately 242 death events in the durvalumab monotherapy and placebo treatment groups (46.2% maturity) will have occurred.

The planned OS-IA2 will occur when approximately 299 death events have occurred in the durvalumab monotherapy and placebo treatment groups (57.1% maturity).

If OS-IA2 does not coincide with the PFS primary analysis, an additional analysis of OS will occur at the same time as the PFS Primary with 0.01% alpha (2-sided) allocated as described in section 4.2.1.

The interim analyses will be assessed by an IDMC (further details are given in the IDMC charter). If none of the PFS and OS interim analysis results (primary analysis method per BICR) meet the efficacy boundary for superiority, the study will remain blinded and continue to be followed for PFS and survival. The recommendations from the IDMC will not reveal the results of the analysis but will take the form of "Continue/Modify/Recommend Early Submission/Stop."

A data cut-off will be applied programmatically at each analysis with the date of the cut-off shown in all summaries.

PFS interim analysis

For the comparison of durvalumab monotherapy versus placebo, approximately 308 PFS BICR events (58.8% maturity) will be available for the interim analysis. The Lan-DeMets spending function that approximates an O'Brien Fleming approach will be used to account for multiplicity introduced by including an interim analysis for superiority (Lan and DeMets 1983s).

At the time of the PFS-IA, the significance level will be calculated based on the information fraction i.e. the actual number of events observed as a proportion of the planned primary

number of events (n = 370). If for example 308/370 (IF=83.2%) of the number of PFS BICR events required at the time of the primary PFS analysis are available at the time of PFS-IA, the 2-sided significance level to be applied for the PFS-IA would be 0.184%, and the 2-sided significance level to be applied for the primary PFS analysis would be 0.444% (controlled at an overall alpha level of 0.5%). The alpha level applied at the primary analysis will be adjusted (using a generalized Haybittle-Peto method) to account for the actual alpha spent at the interim analysis based on the actual final total number of events, to maintain control of overall Type I error.

OS interim analyses

For the comparison of durvalumab monotherapy versus placebo, approximately 242 and 299 death events (46.2% and 57.1% maturity) will be available for the first and the second interim OS analyses, respectively. An additional OS analysis will occur at the time of PFS primary analysis if OS IA2 doesn't coincide with PFS primary analysis. The Lan-DeMets spending function that approximates an O'Brien Fleming approach will be used to account for multiplicity introduced by including an interim analysis for superiority (Lan and DeMets 1983).

At the time of each interim analysis, the significance level will be calculated based on the information fraction i.e. the actual number of events observed as a proportion of the planned primary number of events (n = 348). If for example 242/348 (IF=69.5%) and 299/348 (IF=85.9%) of the number of death events required at the time of the primary OS analysis are available at the time of each interim analysis respectively, the 2-sided significance level to be applied for the first and second OS interim analyses would be 1.239% and 2.392%, respectively, and the 2-sided significance level to be applied for the primary OS analysis would be 3.608% (controlled at an overall alpha level of 4.5% and allocating an alpha of 0.01% (2-sided) to the OS analysis that may occur at the time of PFS primary analysis.)

The alpha level applied at the primary analysis will be adjusted (using a generalized Haybittle-Peto method) to account for the actual alpha spent at the 2 interim analyses based on the actual final total number of events, to maintain control of overall Type I error.

The key secondary comparison of durvalumab in combination with tremelimumab versus placebo will similarly use a LanDeMets spending function to define significance boundaries for PFS and OS (with a generalized Haybittle-Peto method applied to the primary analysis). These will be based on the number of events for the comparison and are therefore distinct from those defined for the boundaries for the primary comparison of durvalumab monotherapy versus placebo.

5.2 Independent Data Monitoring Committee

The safety of all AstraZeneca clinical studies is closely monitored on an ongoing basis by AstraZeneca representatives in consultation with Patient Safety. Issues identified will be addressed; for instance, this could involve amendments to the study protocol and letters to Investigators.

A data monitoring committee will be utilized for this study. Appendix A5 of the CSP provides more details on the rationale for and the remit of the committee.

An IDMC comprised of independent experts will be convened to confirm the safety and tolerability of the proposed dose and schedule and for the planned interim analyses. The safety review will take place after the first 20 patients have been randomized into each of the 3 treatment groups (i.e. after a total of 60 patients have been randomized to the study). In addition, the IDMC will review planned interim analyses and inform the Sponsor whether the interim boundaries specified in Section 5.1 are met. The recommendations from the IDMC will not reveal the results of the analyses but will take the form of "Continue/Modify/Recommend early submission/Stop."

The study may also be stopped based on the findings of the interim safety analysis conducted by the IDMC.

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

6. CHANGES OF ANALYSIS FROM PROTOCOL

- Table 13 in CSP v5 states that the ORR analysis will use Investigator RECIST 1.1. assessments. The main analysis of ORR will use BICR RECIST 1.1 assessments. A further sensitivity analysis of ORR will be conducted using Investigator RECIST 1.1. assessments, as detailed in Section 4.2.4 of the SAP. This will be updated in any future versions of the CSP.
- Section 9.2 of the CSP v5 states that the first and second OS interim analyses would provide "48% power to detect a PFS HR of 0.73 (CV=0.725)" and "68% power to detect a PFS HR of 0.73 (CV=0.770)" respectively. This should be "OS HR" rather than "PFS HR" and will be updated in any future versions of the CSP.
- Section 9.6 of the CSP v5 states that 3 OS analyses are planned, including one primary analysis and two interim analyses. This has been updated in multiple sections (Section 1.3, Section 4.2.1 etc.) that a potential additional OS analysis will occur at the time of PFS primary analysis with 0.01% alpha (2-sided) spent if OS IA2 doesn't coincide with PFS primary analysis. The provided alpha boundaries also have been updated in the SAP to reflect this change and might not align with the alpha boundaries in Section 9.6 of CSP V5.
- Section 9.3.6 of the CSP defines the ADA analysis set. This has been updated in Section 2.1.6 of the SAP to include further subsets for the durvalumab and tremelimumab.

7. REFERENCES

Andersen et al 2004

Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. Lifetime Data Analysis 2004; 10, 335-350.

Basch et al 2009

Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationship with clinical outcomes. J Natl Cancer Inst 2009; 101:1624-32.

Clopper and Pearson 1934

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934; 26; 404-413.

Cockcroft and Gault 1976

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

Cox 1972

Cox DR. Regression models and life-tables. J Royal Stat Society Series B 1972; 34(2):187-220.

Efron 1977

Efron, B. The Efficiency of Cox's Likelihood Function for Censored Data. Journal of the American Statistical Association 72:557–565.EORTC QLQ-C30 Scoring Manual, Third Edition

EORTC QLQ-C30 Scoring Manual

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

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.

FDA PRO Guidance Document 2021

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials

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.

Klein et al 2008

Klein JP, Gerster M, Andersen PK, Tarima S, Perme MP. SAS and R functions to compute pseudo-values for censored data regression. Comput Mehods Programs Biomed 2008 Jan; 89(3): 289-300.

Klein et al 2007

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

Lan and DeMets 1983

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

Litwin et al 1998

Litwin MS, Lubeck DP, Henning JM, et al. Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: results of the CaPSURE database. J Urol 1998; 159:1988-92.

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.

Royston and Parmar 2013

Royston P and Parmar MKB. Restricted mean survival time: an alternative to hazard ratio for the design and analysis of randomized trials with time-to-event outcome. BMC Med Res Methodol 2013 Dec; 7;13:152.

Royston and Parmar 2011

Royston P and Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when proportional hazards assumption is in doubt. Stat Med 2011; 13;2409-2421.

SAS Institute Inc. 2018

SAS Institute Inc. 2018. SAS/STAT® 15.1 User's Guide. Cary, NC: SAS Institute Inc

Sprangers and Aaronson 1992

Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. J Clin Epidemiol 1992; 45:743-60.

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.

8. APPENDICES

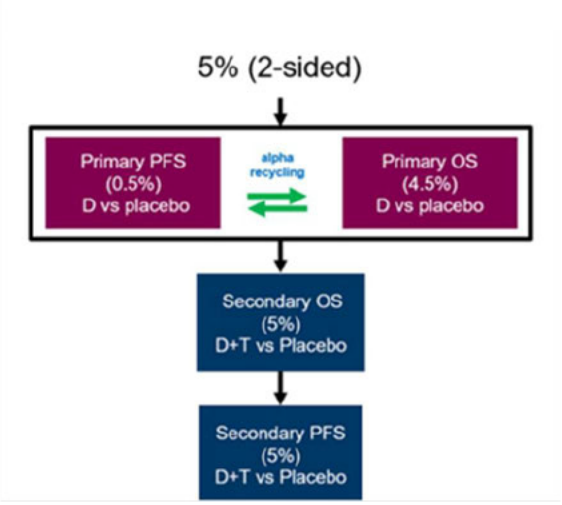
APPENDIX A. ALPHA SPENDING

INTRODUCTION

As per the MTP (Figure 2) and repeated below: At the first level of the MTP, the alpha is split between the tests of PFS and OS for durvalumab monotherapy (D) vs placebo. The D vs placebo PFS test has an overall alpha 0.5%, the D vs placebo OS test has overall alpha 4.5%. If both tests are found significant then durvalumab plus tremelimumab (D+T) vs placebo will be tested sequentially at 5%, OS first, and if this is significant, PFS will be tested.

One interim analysis is planned for PFS and up to three interim analyses are planned for OS.

We use the Lan and DeMets spending function that approximates an O'Brien Fleming approach to calculate the alpha splits for the interim and primary analyses of PFS and OS. The alpha calculation will use the actual observed number of events in the corresponding population at the time of analysis as a proportion of those expected at the primary analysis.



D Durvalumab; OS Overall survival; PFS Progression-free survival; T Tremelimumab

Instructions for Alpha Level Calculation:

A separate spending function will be used to control alpha for the tests of PFS D vs placebo and D+T vs placebo, and OS D vs placebo and D+T vs placebo.

Software: SAS PROC SEQDESIGN

Confidence Interval of HR:

All formal tests included in the MTP will summarize the corresponding HR CIs at the assigned alpha level, i.e. 2 CIs;

- 1 – 2-sided alpha level **with** alpha recycling
- 1 – 2-sided 5% level (i.e. 95% CI).

Table 9 shows the significance boundaries for PFS if the number of events are observed exactly match those predicted by the protocol.

Table 9 Significance boundaries for PFS analyses*

Test	Analysis timepoint	Events	Information Fraction	Boundary controlled at 0.5%	Boundary controlled at 5%
PFS D vs placebo	PFS-IA	308	308/370 = 83.2%	0.00184	0.02805
	PFS-Primary	**370**	100%	0.00444	0.04194
PFS D+T vs placebo	PFS-IA	274	274/309 = 88.7%	NA	0.03460
	PFS-Primary	**309**	100%	NA	0.04042

*At the time of the PFS-IA and PFS primary analyses, significance boundaries will be calculated based on the actual number of events at the time of analysis (with no rounding). For statistical significance to be declared, the p-value for the HR for the treatment effect must be <2-sided significance boundary (rounded to 5 decimal places).

Table 10 shows the significance boundaries for OS if the number of events are observed exactly match those predicted by the protocol and an alpha of 0.01% (2-sided) is allocated to OS analysis that may occur at the time of PFS primary analysis. The remaining 4.49% alpha level for OS will be controlled at the time of planned OS-IA1, OS-IA2 and the primary analysis using the Lan-DeMets spending function that approximates an O'Brien Fleming approach.

Table 10 Significance boundaries for OS analyses*

Test	Analysis timepoint	Events	Information Fraction	Boundary controlled at 4.49%	Boundary controlled at 4.99%
OS D vs placebo	OS-IA1	242	242/348 = 69.5%	0.01239	0.01434
	OS-IA2	299	299/348 = 85.9%	0.02392	0.02690
	OS-Primary	**348**	100%	0.03608	0.03994

Test	Analysis timepoint	Events	Information Fraction	Boundary controlled at 5%
OS D+T vs placebo	OS-IA1	204	204/276 = 73.9%	0.01826
	OS-IA2	242	242/276 = 87.7%	0.02808
	OS-Primary	**276**	100%	0.03925

*At the time of the OS-IA1, OS-IA2 and OS primary analyses, significance boundaries will be calculated based o n the actual number of events at the time of analysis (with no rounding). For statistical significance to be declared , the p-value for the HR for the treatment effect must be <2-sided significance boundary (rounded to 5 decimal pl aces).

Example SAS Code

```
title1 "PFS: Durva mono vs Placebo";
title2 "Control at 0.5% (2-sided)";
ods output errspend = errspend boundary  = i_test1;
  proc seqdesign  ERRSPEND boundaryscale=pvalue;
    twosidedobf:  design  nstages=2  method = errfuncobf
    info=cum(%SYSEVALF(308/370) 1) alpha=0.005;
  run;
ods output close;

data pvalout1 (keep = _stage_ _infoprop_ bound_la Pval) ;
      set i_test1 ;
      Pval = round(2*100000*bound_la)/100000;
run;

proc print data=pvalout1;
run;


title1 "OS: Durva mono vs Placebo ";
title2 "Control at 4.5% (2 sided)";
ods output errspend = errspend boundary  = i_test2;
```

```
  proc seqdesign  ERRSPEND boundaryscale=pvalue;
    twosidedobf:  design  nstages=3  method = errfuncobf
    info=cum(%SYSEVALF(242/348) %SYSEVALF(299/348) 1) alpha=0.0449;
  run;
ods output close;

data pvalout2 (keep = _stage_ _infoprop_ bound_la Pval) ;
      set i_test2;
      Pval = round(2*100000*bound_la)/100000;
run;

proc print data=pvalout2;
run;
```

APPENDIX B. 2 MISSED VISIT RULES

RECIST 1.1. Assessments

For RECIST 1.1 assessments, we will apply a 'look-back' approach to the 2 missed visit rule. That is, if an event (e.g. progression or death) is observed, then it will be considered an event only if there is an assessment within the 2 visit window immediately before the event, irrespective of whether there were missed visits prior to that assessment. If there are 2 consecutive missed visits prior to the event, the event will be censored as the last valid RECIST 1.1 assessment, prior to the 2 missed visits. For example:

Week 8	Week 16	Week 24	Week 32	Week 40	Censor for 2 missed visit rule?
Event?	Event?	Event?	Event?	Event?	
No	Missed visit	Missed visit	No	Yes	No
No	No	Missed visit	Missed visit	Yes	Yes – censor at week 16

Given the scheduled visit assessment for RECIST (i.e. q8w ± 1w for the first 72 weeks then q12w ± 1w up to 96 weeks then q24w ± 1w thereafter) the definition of 2 missed visits window will change:

Length of 2 missed visit window for RECIST assessments

Missing baseline or within 2 visits of baseline:

- A patient without baseline will be censored at day 1 unless they die ≤ day 119 (week 17).
- Any scan ≤ day 119 (week 17) is within 2 visits of baseline and therefore cannot be censored for missing visits.

Schedule: q8w:

- For any scan > day 119 until ≤ day 511 (week 73) the visit schedule is q8w ±1 week and thus the maximum time between two consecutive scans is 1+8+8+1=18 weeks (126 days).

Schedule: change from q8w to q12w:

- For any scan > day 511 until ≤ day 595 (week 85) the visit schedule is q12w ±1 week and the visit schedule for the cycle before was q8w ±1 week and thus the maximum time between two consecutive scans is 1+8+12+1=22 weeks (154 days).

Schedule: q12w:

- For any scan > day 595 until ≤ day 679 (week 97) the visit schedule is q12w ±1 week and the visit schedule for the cycle before was q12w ±1 week thus the maximum time between two consecutive scans is 1+12+12+1=26 weeks (182 days).

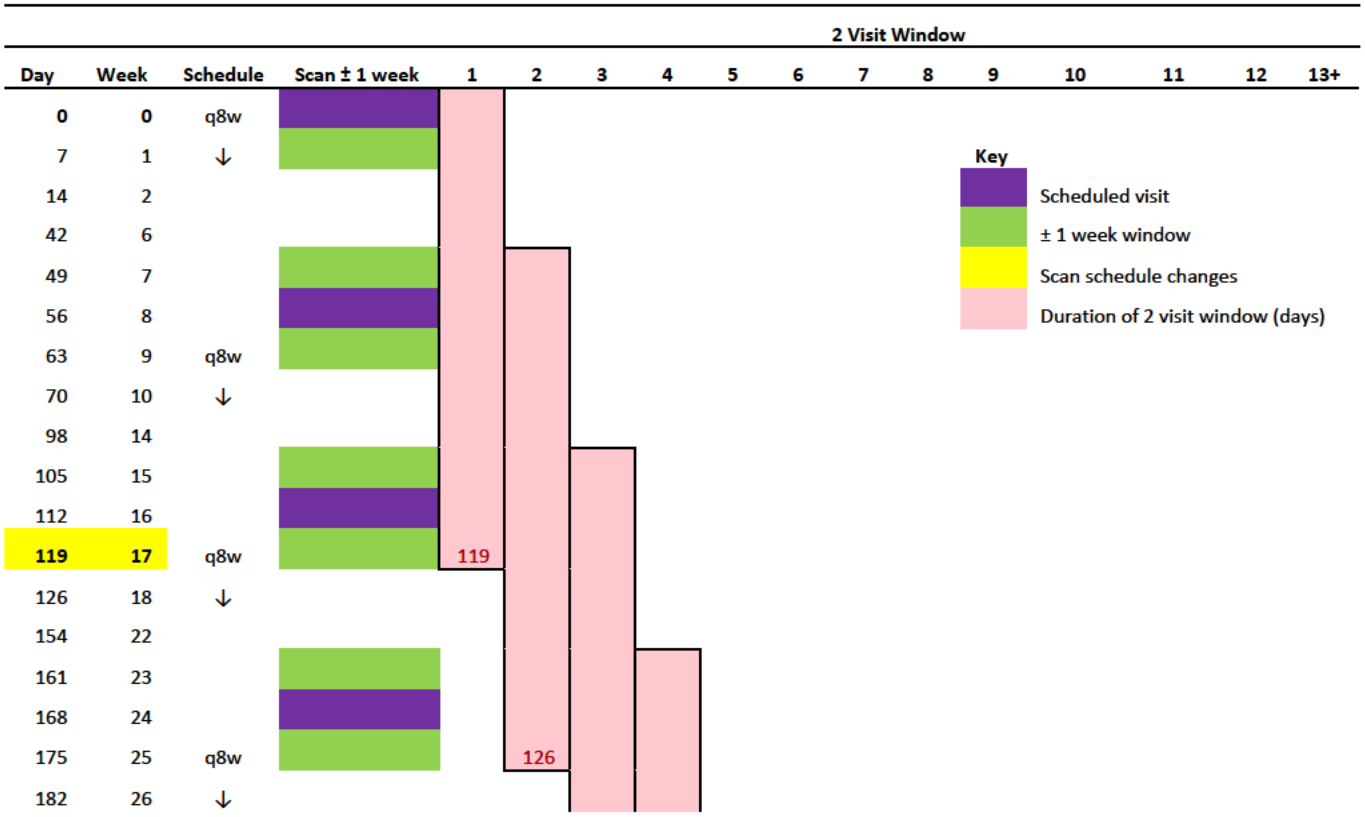
Schedule: change from q12w to q24w:

- For any scan > day 679 until ≤ day 847 (week 121) the visit schedule is q24w ±1 week and the visit schedule for the cycle before was q12w ±1 week and thus the maximum time between two consecutive scans is 1+12+24+1=38 weeks (266 days).

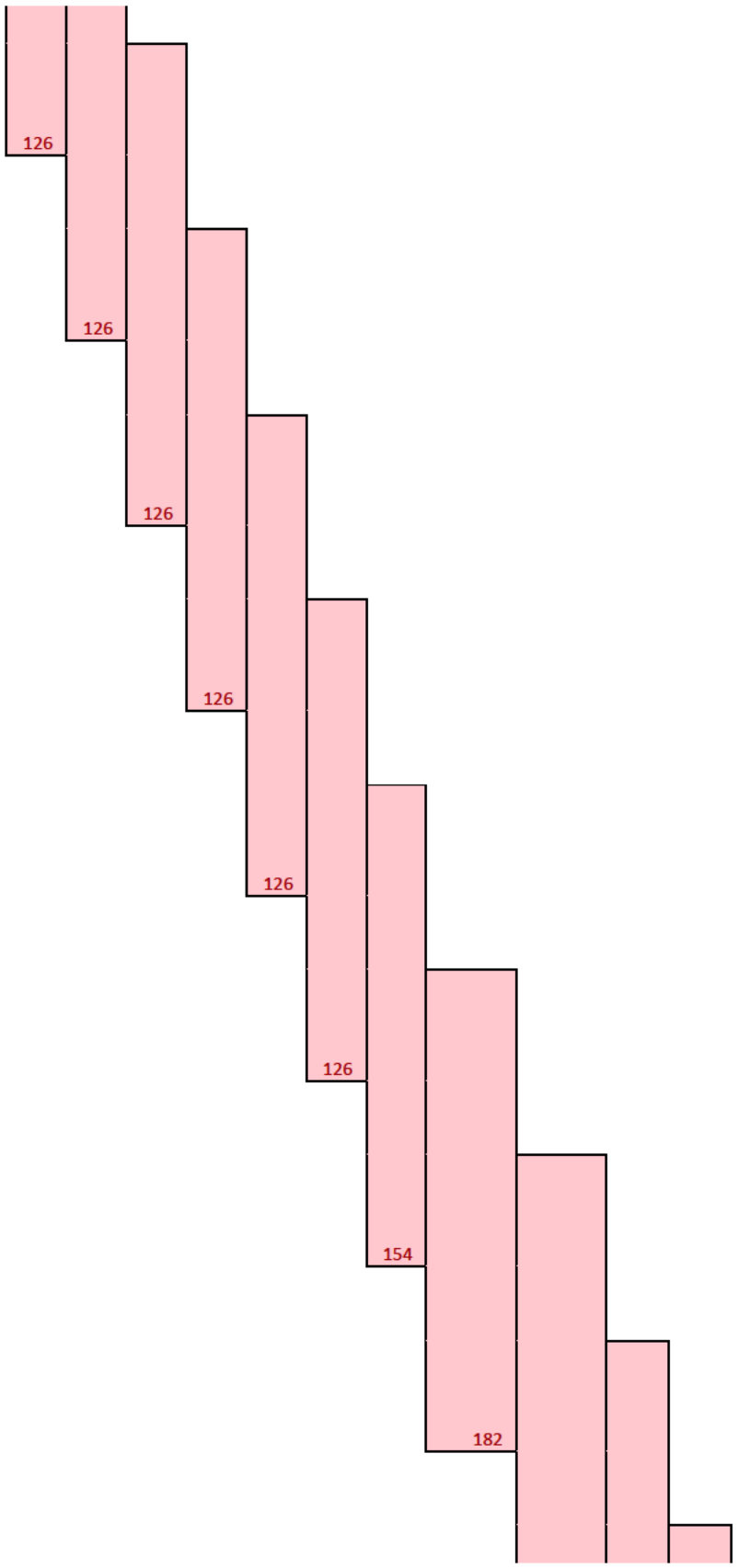
Schedule: q24w:

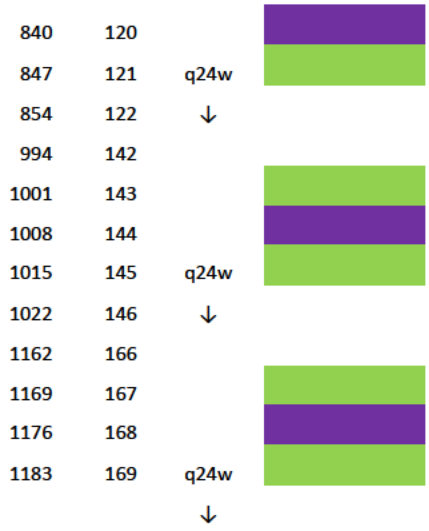
- For any scan > day 847 (week 121) the visit schedule is q24w±1 week and thus the maximum time between two consecutive scans is 1+24+24+1=50 weeks (350 days).

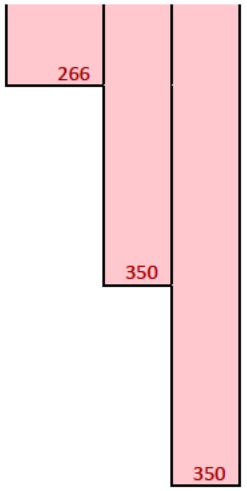
Table 10 2 Missed Visits – RECIST



210	30		
217	31		
224	32		
231	33	q8w	
238	34	↓	
266	38		
273	39		
280	40		
287	41	q8w	
294	42	↓	
322	46		
329	47		
336	48		
343	49	q8w	
350	50	↓	
378	54		
385	55		
392	56		
399	57	q8w	
406	58	↓	
434	62		
441	63		
448	64		
455	65	q8w	
462	66	↓	
490	70		
497	71		
504	72		
511	**73**	q12w	
518	74	↓	
574	82		
581	83		
588	84		
595	85	q12w	
602	86	↓	
658	94		
665	95		
672	96		
679	**97**	q24w	
686	98	↓	
826	118		
833	119		




840
120
847
121
q24w
854
122
994
142
1001
143
1008
144
1015
145
q24w
1022
146
1162
166
1169
167
1176
168
1183
169
q24w


266
350
350

ePRO Assessments

As per Sections 3.3.1.1 and 3.3.1.2, confirmation of deterioration will first be determined, then the censoring rules will be applied to assess if any determined deterioration can be used or if it needs to be censored earlier due to 2 consecutive missed visits. 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 (due to 2 missed visits)
	MV	DT	NC	MV	MV	DT	DT	NC	Event at V6	No event, censor at V3 (due to 2 missed visits)
	NC	DT	NC	MV	IMP	DT	NC	DT	Event not confirmed, bit if this is the last DT prior to DCO and then it is counted as confirmed at V8	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, if last visit prior to DCO	No event, censor at V4 (due to 2 missed visits)
	MV	MV	NC	MV	MV	DT	DT	NC	Event at V6	No event, censor at

										baseline (due to 2 missed visits)
	NC	DT	MV	MV	DT	NC	IMP	DT	Event at V2	Event at V2

DT Deterioration; IMP Improvement; MV Missed visit. NC No change.

*Missed visits are allowed in between assessments confirming deterioration and confirmation of deterioration requires at least 14 days between assessments (which should take into account the weekly assessment schedule for the QLQ-LC13 for the first 8 weeks).

Length of 2 missed visit window for EORTC QLQ-LC13

The scheduled visit assessments and visit windows for the QLQ-LC13 are weekly (±1 day on weeks 1, 2, 3, 5, 6 and 7 and ±3 days on weeks 4 and 8) for the first 8 weeks from randomization and then every 4 weeks relative to randomization until study termination, PFS2 or death.

Missing baseline or within 2 visits of baseline:

- A patient without baseline will be censored at day 1 unless they die ≤ day 15.
- Any assessment ≤ day 15 (2 weeks + 1 day) is within 2 visits of baseline and therefore cannot be censored for missing visits.

Schedule: q1w ±1 day:

- For any assessments > day 1 where the visit schedule is q1w ±1 day, the maximum time between two consecutive assessments is 1+7+7+1 = 16 days

Schedule: q1w ±3 days:

- For any assessments where the visit schedule changes from q1w ±1 day to q1w ±3 days, the maximum time between two consecutive assessments is 1+7+7+3 = 18 days

Schedule: q4w ±3 days:

- For any assessments where the visit schedule changes from q1w ±3 days to q4w ±3 days, the maximum time between two consecutive assessments is 3+7+28+3 = 41 days
- For any assessments where the visit schedule is q4w ±3 days, the maximum time between two consecutive assessments is 3+28+28+3 = 62 days

Length of 2 missed visit window for EORTC QLQ-C30

The scheduled visit assessments and visit windows for the QLQ-C30 are 4 weekly (±3 days) relative to randomization until study termination, PFS2 or death.

Missing baseline or within 2 visits of baseline:

- A patient without baseline will be censored at day 1 unless they die ≤ day 59.
- Any assessment ≤ day 59 is within 2 visits (8 weeks + 3 days) of baseline and therefore cannot be censored for missing visits.

Schedule: q4w:

- For any assessment > day 1 the visit schedule is q4w ±3 days and thus the maximum time between two consecutive assessments is 3+28+28+3=62 days (8 weeks + 6 days).

APPENDIX C. ECRFS USED TO DETERMINE LAST KNOWN ALIVE DATE

The following eCRFs will be used to determine the last known alive date in the absence of a record in the SURVIVE module:

CONSENT, CONSENT1, DISPOSITION, CONSWD, DOSDISC, DOSDISC1, CAPRX, CAPRX1, CAPRXR, CAPRXR1, CAPRXR_PCI, CM, DM, SERAE, AE, AESI, EG, EG2, EG3, EX, EX1, OVERDOSE, DISEXT, PATHGEN, LIVERDI, ILDIS, TTSCAPRX, HOSPAD, LB, LB1, LB2, LB3, LB4, LB7, PREG, LB5, LB6, MH, HISS, LIVERRF, LIVERSS, PSTAT, ePRO Assessments, 1RECIST1, 2RECIST1, 3RECIST1, 4RECIST1, RECIST2_2, PFS2, SU_NIC, VS, VS1, VS2, VS3, VS4, VS5, VS6, VS7, VS8, VS9, CONPRO.

APPENDIX D. ANALYSIS OF DATA FROM CHINA AND ASIA

Introduction

Per 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 in Chinese and Asian populations may be required to facilitate the benefit-risk assessment for Chinese patients.

This appendix outlines the pre-specified analyses that may be conducted for the China cohort and Asia subset of study D933QC00001 to support ADRIATIC submission in China.

The China cohort will include all patients randomized at sites located in the mainland China.

The Asia subset will include all patients randomized at sites in Asia countries/ regions e.g. China, Japan, South Korea, Taiwan and Vietnam, and claiming themselves as Asian.

Definition of analysis sets

China full analysis set

The China full analysis set will include all patients in the China cohort. The FAS will be used for all efficacy analyses (including PROs). Treatment groups will be compared on the basis of randomized study treatment, regardless of the treatment actually received. Patients who were randomized but did not subsequently go on to receive study treatment are included in the analysis in the treatment group to which they were randomized.

China combination analysis set

For analyses involving the durvalumab and tremelimumab combination treatment group, only the first 600 patients randomized (across all 3 arms) and in the China cohort will be included in the analyses, and all will be included in the treatment group to which they were randomized.

China safety analysis set

The China safety analysis set will consist of all patients in China cohort who received at least 1 dose of study treatment (see Section 3.4.5 for Valid Dose rule). Safety data will not be formally analyzed but summarized using the safety analysis set according to the treatment received, that is, erroneously treated patients (i.e. those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received. Patients who receive incorrect therapy will be summarized according to treatment group outlined in Section 2.1.3.

China combination safety analysis set

The China combination safety analysis set will consist of all patients from the China combination analysis set who received at least 1 dose of study treatment. Data will be summarized according to the treatment they actually received, applying the same approach for those receiving incorrect therapy as outlined for the safety analysis set in Section 2.1.3.

China pharmacokinetic analysis set

All patients in China cohort who receive at least 1 dose of IP per the protocol for whom any post-dose data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analyses will be included in the China PK analysis set. Further details on what constitutes as a violation/deviation from the protocol can be found in Appendix E. The population will be defined by the Study Clinical Lead, Pharmacokineticist, and Statistician prior to any analyses being performed.

China ADA analysis set

The China anti-drug antibody (ADA) analysis set includes all patients in the China safety analysis set who have non-missing baseline ADA and at least 1 non-missing post-baseline ADA result of the same IP (durvalumab or tremelimumab).

The durvalumab China ADA analysis set will consist of all patients in the safety analysis set who have a non-missing baseline durvalumab ADA result and at least one non-missing post-baseline durvalumab ADA result.

The tremelimumab China ADA analysis set will consist of all patients in the safety analysis set who have a non-missing baseline tremelimumab ADA result and at least one non-missing post-baseline tremelimumab ADA result.

Asia full analysis set

The Asia full analysis set will include all patients in the Asia subset. The FAS will be used for all efficacy analyses (including PROs). Treatment groups will be compared on the basis of randomized study treatment, regardless of the treatment actually received. Patients who were randomized but did not subsequently go on to receive study treatment are included in the analysis in the treatment group to which they were randomized.

Asia combination analysis set

For analyses involving the durvalumab and tremelimumab combination treatment group, only the first 600 patients randomized (across all 3 arms) and in the Asia subset will be included in the analyses, and all will be included in the treatment group to which they were randomized.

Asia safety analysis set

The Asia safety analysis set will consist of all patients in Asia subset who received at least 1 dose of study treatment (see Section 3.4.5 for Valid Dose rule). Safety data will not be formally analyzed but summarized using the safety analysis set according to the treatment

received, that is, erroneously treated patients (i.e. those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received. Patients who receive incorrect therapy will be summarized according to treatment group outlined in Section 2.1.3.

Asia combination safety analysis set

The Asia combination safety analysis set will consist of all patients from the Asia combination analysis set who received at least 1 dose of study treatment. Data will be summarized according to the treatment they actually received, applying the same approach for those receiving incorrect therapy as outlined for the safety analysis set in Section 2.1.3.

Asia pharmacokinetic analysis set

All patients in Asia subset who receive at least 1 dose of IP per the protocol for whom any post-dose data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analyses will be included in the Asia PK analysis set. Further details on what constitutes as a violation/deviation from the protocol can be found in Appendix E. The population will be defined by the Study Clinical Lead, Pharmacokineticist, and Statistician prior to any analyses being performed.

Asia ADA analysis set

The Asia anti-drug antibody (ADA) analysis set includes all patients in the Asia safety analysis set who have non-missing baseline ADA and at least 1 non-missing post-baseline ADA result of the same IP (durvalumab or tremelimumab).

The durvalumab Asia ADA analysis set will consist of all patients in the safety analysis set who have a non-missing baseline durvalumab ADA result and at least one non-missing post-baseline durvalumab ADA result.

The tremelimumab Asia ADA analysis set will consist of all patients in the safety analysis set who have a non-missing baseline tremelimumab ADA result and at least one non-missing post-baseline tremelimumab ADA result.

Primary, secondary and exploratory variables for China and Asia analysis

All efficacy, safety, PRO and PK variables for the China cohort and Asia subset will be derived in the same way as detailed in the SAP Section 3.

Analysis methods

The same analysis methods described in Section 4 of the SAP will be applied to the China cohort and Asia subset accordingly.

All analyses detailed in Section 4 of the SAP will be repeated for the China cohort and Asia subset where possible, using the analysis sets described above.

However, the analyses for the efficacy endpoints for the China cohort and Asia subset include the following differences compared to those described in the SAP (for the Global cohort):

- All statistical analyses will be considered exploratory in the China cohort/Asia subset, and only performed if sufficient numbers of events or patients provide meaningful analyses with associated results, otherwise only descriptive statistics will be presented.
- The calculated p-values are considered as nominal only.
- No adjustment for multiplicity will be made and the multiple testing procedure (MTP) detailed in Section 4.2.1 will not be applied.
- The subgroup analyses for PFS and OS will not be performed for Race and Geographic region.
- No sensitivity analysis for the primary endpoints will be conducted except the one for ascertainment bias.
- Analyses results for primary and key secondary endpoints will be presented in a forest plot which will include results of the Global cohort, China cohort and Asia subset.
- No summary for Healthcare resource use will be produced.
- No analysis for biomarker will be conducted.

The same DCOs will be applied to the China cohort and Asia subset as in the global cohort.

No separate listings of individual patients in the China cohort and Asia subset will be produced.

APPENDIX E. RULES FOR EXCLUSION FROM THE PK ANALYSIS SET

Rule #	Category	Programming rules (applies to both Durva and Treme PK samples)
	General	All measured PK concentration data should be reported in the listings. Any assessments / subjects who meet the exclusion criteria below should be excluded from the summary tables and figures only.
1	PK sampling assessment conducted outside of the Schedule of Assessments.	E.g. a PK sample assessment scheduled at cycle 2 (C2) (week 4) but instead conducted at C3 (week 8), to exclude this particular PK concentration from the tables and figures.
2	IV infusion - dose not fully (100%) received	If dose not fully (100%) received at >1 cycle, to exclude all PK concentration data for that particular subject. If dose not fully (100%) received at C1, exclude that particular PK concentration (as a post dose assessment), as well as all subsequent PK concentration data. If dose not fully (100%) received at any cycles post C1, exclude all subsequent PK concentration data as these are pre dose assessments which will be impacted.
3	IV infusion - dose interruption	EX dose interruption to the IV dose infusion. If dose interrupted at >1 cycle, to exclude all PK concentration data for that particular subject. If dose interrupted at C1, exclude that particular PK concentration (as post dose assessment), as well as all subsequent PK concentration data.

		If dose interrupted at any cycles post C1, exclude all subsequent PK concentration data as these are pre dose assessments which will be impacted.
4	Timing of sample - post-dose PK sampling conducted 1 hour after the end of infusion.	C1 only (as post dose) - exclude PK concentration data where sampling was conducted >1 hour after the end of IV infusion.

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: d933qc00001-sap-ed-5		
Document Title:	Statistical Analysis Plan Edition 5	
Document ID:	Doc ID-004048124	
Version Label:	5.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
20-Dec-2023 14:00 UTC	[redacted]	Content Approval
20-Dec-2023 14:01 UTC	[redacted]	Author Approval

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