AstraZeneca

Statistical Analysis Plan

Study Code	D419AC00002
Edition Number	3.0
Date	11 Jun 2021

A Phase III Randomized, Open-Label, Multi-Center Study of Durvalumab Versus Standard of Care Platinum-Based Chemotherapy as First Line Treatment in Patients with PD-L1-High Expression Advanced Non Small-Cell Lung Cancer (NSCLC) (PEARL)

A Phase III Randomized, Open-Label, Multi-Center Study of Durvalumab Versus Standard of Care Platinum-Based Chemotherapy as First Line Treatment in Patients with PD-L1-High Expression Advanced Non Small-Cell Lung Cancer (NSCLC) (PEARL)

Study Statistician

PPD (AstraZeneca) Date

A Phase III Randomized, Open-Label, Multi-Center Study of Durvalumab Versus Standard of Care Platinum-Based Chemotherapy as First Line Treatment in Patients with PD-L1-High Expression Advanced Non Small-Cell Lung Cancer (NSCLC) (PEARL)

Global Product Statistician

PPD (AstraZeneca) Date

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
ALK	Anaplastic lymphoma kinase
APF12	Proportion of patients alive and progression free at 12 months from first dose
AUC	Area under the curve
Baseline	Refers to the most recent assessment of any variable prior to dosing with study treatment/randomization (as appropriate)
BDRM	Blinded data review meeting
BoR	Best objective response
BSA	Body surface area
CI	Confidence Interval
COVID	Coronavirus Disease
CR	Complete Response
CRF/eCRF	Case Report Form (electronic)
CSP	Clinical study protocol
CSR	Clinical Study Report
CTC/CTCAE	Common Terminology Criteria for Adverse Event (National Institutes of Health, National Cancer Institute)
DCO	Data cutoff
DCR	Disease Control Rate
DoR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EDoR	Expected Duration of Response
EGFR	Epidermal growth factor receptor
EM	Early Mortality
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire
EORTC QLQ-LC13	European Organisation for Research and Treatment of Cancer 13-item lung cancer-specific quality of life questionnaire. Module used as a supplement to EORTC QLQ-C30

Abbreviation or special term	Explanation
CCI	CCI
FAS	Full analysis set
GFR	Glomerular Filtration Rate
GGT	Gamma-glutamyltransferase
HR	Hazard ratio
HRQoL	Health-related Quality of Life
ICU	Intensive care unit
IDMC	Independent data monitoring committee
IHC	Immunohistochemistry
imAE	Immune-mediated Adverse Event
IP	Investigational Product
ITT	Intention to Treat
IVRS	Interactive Voice Response System
LSMEAN	Least squares mean
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed model for repeated measures
MRI	Magnetic Resonance Imaging
MTP	Multiple testing procedure
nAb	Neutralizing Antibody
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NLR	Neutrophil/lymphocyte ratio
NSCLC	Non-small cell lung cancer
NTL	Non-target lesions
OAE	Other Significant Adverse Event
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive disease
PD-L1	Programmed death ligand 1
PFS	Progression-free survival
PFS2	Time from randomization to second progression

Abbreviation or special term	Explanation
PR	Partial Response
PRO	Patient Reported Outcome
QoL	Quality of Life
QTcF	QT interval (corrected for heart rate using Fridericia's correction)
RDI	Relative dose intensity
RECIST 1.1	Response Evaluation Criteria In Solid Tumours, Version 1.1
RS	Raw score
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Safety analysis set
SD	Stable disease
SoC	Standard of care
TL	Target lesions
WHO	World Health Organization

AMENDMENT HISTORY

Date	Brief description of change
11 Jun 2021	Sections 2.2, 4.2.2, 4.2.12 and 4.2.20 updated to add details of additional summaries to assess impact of COVID-19. Section 6 revised to clarify that there's no changes of analysis from protocol (CSP v9.0) now.

Date	Brief description of change
29 Jun 2020	Updated to reflect the changes in protocol amendment #4, #5, #6, #7, #8. These changes are reflected in • Section 1 • Section 2.1 • Sections 3.1, 3.2.1, 3.2.2 • Sections 4.1, 4.2.1, 4.2.2, 4.2.3 • Sections 5.1, 5.2 Additional changes are listed as below. Section 1.3 added the details regarding secondary endpoints. Section 2.1, added ADA evaluable analysis set and ADA evaluable and low risk of early mortality analysis set. Added details regarding cut-off value for the prognostic model developed by AstraZeneca and how to handle erroneously treated patients. Section 2.2, revised deviation categories 4 and 5. Clarified that no deviation will lead to exclusion from FAS. Section 3.2.2.4 revised censoring rule of PFS2. Section 3.2.2.8 removed “apply a window around the week X visit”. Sections 3.3.1 and 3.3.2 added derivation rules and clarifications of EORTC QLQ-C30 and EORTC QLQ-LC13 questionnaires. Section 3.3.4 added PRO compliance rate. Section 3.4.1 removed other significant adverse events, revised the wording from AESI to including both AESI and AEPI, added imAE. Section 3.4.2 added actual exposure definition and calculation of duration of dose delays for SoC. Section 3.4.3 added details regarding dose intensity calculation in SoC arm. Section 3.4.4 added calculation details of creatinine clearance. Section 3.6.1 added details regarding immunogenicity analysis. Section 3.7 minor revision of outcome variables of health resource use. Section 4.1 revised baseline definition, and provided details regarding handling of missing/incomplete dates. Section 4.2.1, following by the revision of dual primary endpoints in CSP #8, full MTP details including recycling to secondary endpoints are provided. Section 4.2.2 added details of primary OS analyses, revised sensitivity analyses and supportive summaries, revised the subgroup analyses and list of covariates to be included in the multivariate analysis, added details regarding the interaction test. Section 4.2.3 revised the supportive summaries.

Date	Brief description of change
29 Jun 2020	Section 4.2.8 added/clarified supportive analyses. Section 4.2.10, added details regarding analyses of improvement rates. Revised/clarified the analysis of MMRM of change from baseline in PRO symptoms. Removed PRO-CTCAE. Section 4.2.12 revised the planned AE summary and summary of death tables, revised the analyses of AESI/AEPI, added analyses of infection AEs and imAEs. Revised the analyses of ECG/urinalysis. Added the analyses of creatinine clearance and additional supportive analyses of lab parameters used in the prognostic model developed by AstraZeneca. Minor revision of other safety data. Section 4.2.14 added the analysis of AE in ADA positive patients. Section 4.2.15 and Appendix B: added the analysis due to PD-L1 dispenser issue. Section 4.2.16 revised the planned analyses of demographic and baseline characteristics data. Section 4.2.17 removed the summary of number of infusion received, added summary of number of patients that switched treatment and exposure by pemetrexed maintenance therapy status in SoC arm. Section 4.2.18 removed the summary of number of regimens received. Section 4.2.19 clarified that analysis of dual primary, secondary and safety endpoints will be repeated for China subgroup. Section 4.2.20 added the analyses due to COVID-19. Section 6 added table of changes of analysis from protocol. Appendix A: added the alpha level calculation details.
25 Apr 2017	Initial Version

1. STUDY DETAILS

1.1 Study objectives

The primary objectives of this study are to assess the efficacy of durvalumab compared with SoC in terms of overall survival (OS) in advanced NSCLC patients with EGFR and ALK wild-type and PD-L1-high expression (TC ≥ 25%) tumors, and in the PD-L1 TC ≥ 25% and low risk of early mortality population.

1.1.1 Primary objectives

Primary Objectives:	**Outcome Measures:**
To assess the efficacy of durvalumab compared to SoC in patients with PD-L1 TC ≥ 25% (all randomized patients)	OS
To assess the efficacy of durvalumab compared to SoC in patients with PD-L1 TC ≥ 25% and low risk of early mortality[a]	OS

a The population at low risk of early mortality consists of patients identified by a prognostic model developed by AstraZeneca as having low risk of early mortality.

1.1.2 Secondary objectives

Secondary Objectives:	**Outcome Measures:**
To assess the efficacy of durvalumab compared to SoC in terms of OS	OS in patients with • PD-L1 TC ≥ 50% • PD-L1 TC ≥ 50% and low risk of early mortality
To further assess the efficacy of durvalumab compared to SoC in terms of PFS, ORR, DoR, OS18, OS24, APF12, and PFS2	PFS, ORR, DoR, APF12 using Investigator assessments according to RECIST 1.1, PFS2 using local standard clinical practice, OS18 and OS24 respectively in patients with • PD-L1 TC ≥ 25% • PD-L1 TC ≥ 25% and low risk of early mortality • PD-L1 TC ≥ 50% • PD-L1 TC ≥ 50% and low risk of early mortality

Secondary Objectives:	Outcome Measures:
To assess disease-related symptoms and HRQoL in patients treated with durvalumab compared to SoC using the EORTC QLQ-C30 v3 and the LC13 module	EORTC QLQ-C30, EORTC QLQ-LC13, and changes in Eastern Cooperative Oncology Group (ECOG) performance status in patients with • PD-L1 TC ≥ 25% • PD-L1 TC ≥ 25% and low risk of early mortality
To investigate the immunogenicity of durvalumab	Presence of ADAs for durvalumab in patients with • PD-L1 TC ≥ 25% • PD-L1 TC ≥ 25% and low risk of early mortality

1.1.3 Safety objective

Safety Objective:	Outcome Measures:
To assess the safety and tolerability profile of durvalumab compared to SoC	AEs, physical examinations, laboratory findings, and vital signs

1.1.4 Exploratory objectives

CCI [redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

Note: Exploratory objective analyses may be reported separately from the main clinical study report.

1.2 Study design

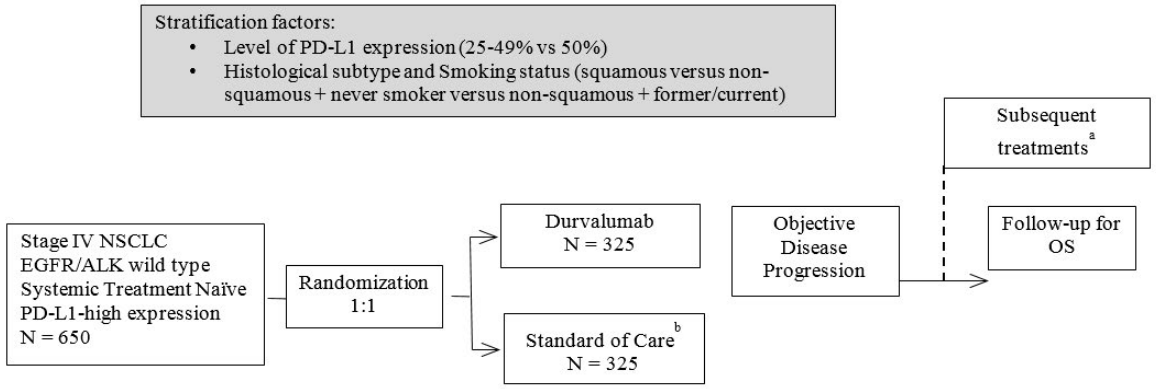
This is a randomized, open-label, multi-center Phase III study to determine the efficacy and safety of durvalumab versus platinum-based SoC chemotherapy in the first-line treatment of advanced NSCLC patients who are EGFR and ALK wild-type and have PD-L1-high expression. A schematic diagram of the overall study design is shown in Figure 1, and a detailed study flow chart is shown in Study flow chart Figure 2.

The study will randomize approximately 650 patients from selected global sites that include China and other countries who will be randomized in a 1:1 ratio to 2 treatment arms (durvalumab or SoC therapy) in a stratified manner according to level of PD-L1 expression (25-49% versus ≥50%), histology and smoking status (squamous versus non-squamous + never smoker versus non-squamous + former/current smoker). This is to ensure the uniform distribution of main predictive factors of response, e.g. PD-L1 expression intensity, histological subtype and smoking status in each treatment group.

Patients will provide a tumour tissue sample at enrolment (newly acquired or archived sample <3 months old) to determine PD-L1 expression status (CCI [redacted] in which ≥ 25% PD-L1–membrane expression in tumoral tissue is considered as high expression.

Doses and treatment regimens are described in Section 7.2 of the clinical study protocol (CSP). Assessments will be conducted as indicated in Table 2, Table 3, and Table 4 of the CSP.

Figure 1 Overall study design

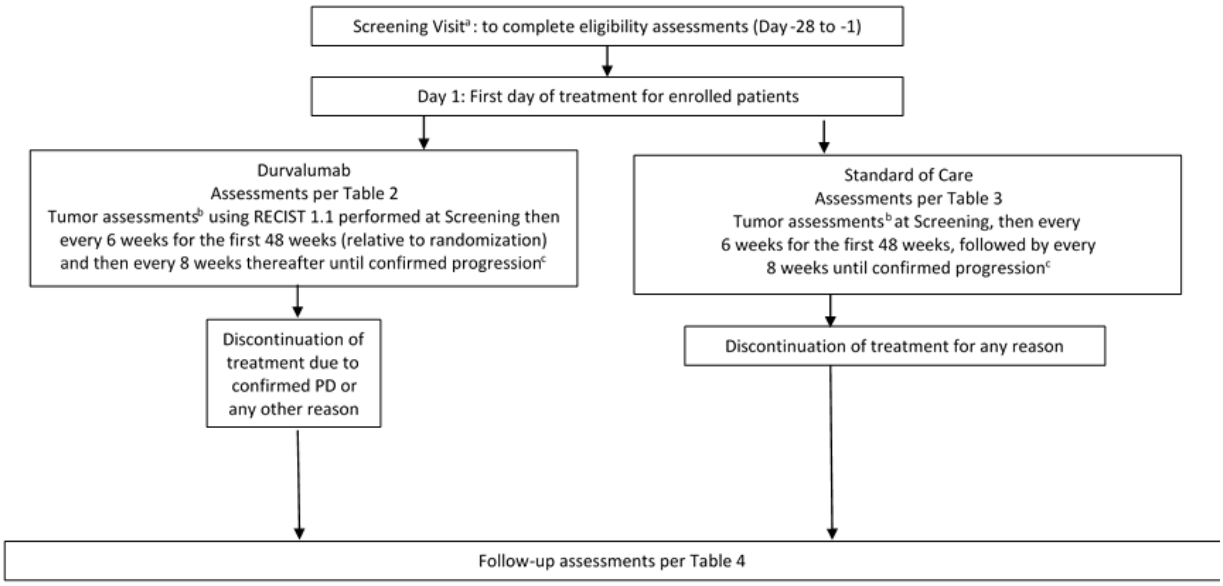


[a] Offer of subsequent therapy per Investigator discretion.

[b] Standard of Care is an Investigator choice from the following: paclitaxel + carboplatin, gemcitabine + cisplatin (or carboplatin) (squamous only), pemetrexed + cisplatin (or carboplatin) (non-squamous only), and for eligible patients, pemetrexed maintenance (non-squamous only following pemetrexed/platinum induction).

Figure 2 Study flow chart

Tables referred to in this figure are in reference to the CSP.



[a] Informed consent of study procedures and tumor biopsy sample may be obtained prior to the 28-day screening window, if necessary, in order to permit tumor biopsy sample acquisition and analysis (PDL1 status and, if unknown, EGFR and ALK status) prior to randomization.
[b] Tumour assessments were performed using RECIST 1.1
[c] A confirmatory scan is always required following the initial demonstration of PD and preferably at the next scheduled visit (in the absence of clinically significant deterioration).

Independent Data Monitoring Committee (IDMC)

An IDMC comprising independent experts will be convened and will meet approximately 6 months after the study has started or after the first 30 patients have been randomized and received at least 2 cycles of treatment, whichever occurs first, to review safety assessments and make recommendations to continue, amend, or stop the study based on safety findings. The committee will meet approximately every 6 months thereafter until a decision to unblind the study is made. In addition, the IDMC will review the unblinded interim analysis summaries of efficacy data.

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

1.3 Number of patients

The study will randomize approximately 650 eligible patients 1:1 to durvalumab or SoC.

The primary objectives of this study are to assess the efficacy of durvalumab compared with standard of care (SoC) in terms of overall survival (OS) in advanced NSCLC patients with EGFR and ALK wild-type and PD-L1-high expression (TC ≥ 25%) tumors, and in the PD-L1 TC ≥ 25% and low risk of early mortality population. To strongly control for type I error at a 5% level (2 sided), an alpha of 4% (2 sided) will be used for the OS comparison of

durvalumab versus SoC in the PD-L1 TC ≥ 25% population, and an alpha of 1% (2 sided) will be used for the OS comparison of durvalumab versus SoC in the PD-L1 TC ≥ 25% and low risk of early mortality population. The study will be considered to have met its primary objective if either of the dual primary OS results are statistically significant.

There will be 2 data cut-offs in the study, interim analysis and final analysis. The alpha will be split between the 2 analyses using the Lan and DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, with the boundaries for the treatment comparison derived based upon the exact number of OS events at the time of analysis to strongly control the overall type I error at a 5% level (2 sided).

One interim analysis to assess the superiority of the durvalumab group (compared to SoC group) in terms of OS will be performed when all of the following conditions have been met

- approximately 85% of the target 521 OS events (approximately 68% maturity) have been achieved across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% population
  And
- approximately 85% of the target 414 OS events (approximately 64% maturity) have been achieved across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% and low risk of early mortality population
  And
- a minimum 12 months follow-up from last patient randomized to the study.

The final (primary) analysis of OS will be performed when all of the following conditions have been met

- approximately 521 OS events (approximately 80% maturity) have occurred across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% population
  And
- approximately 414 OS events (approximately 76% maturity) have occurred across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% and low risk of early mortality population.

Durvalumab versus SoC (OS in the PD-L1 TC ≥ 25% population)

The sizing assumes a 6 months delay in separation of the OS curves between the two groups, hence the use of an average hazard ratio. If OS at 18 months was 46% with durvalumab and 36% with SoC (with a 12.9-month median OS [Chen, et al 2014, Ciuleanu et al 2009, Paz-Ares et al 2013, Scagliotti et al 2008] , i.e., Weibull distribution was assumed with shape parameter = 1.164459) and assuming the true average OS HR is 0.75, an estimated 521 death events (approximately 80% maturity) are expected to have occurred at 47 months from the time the first patient has been randomized with a 25-month recruitment period. With an estimated 521 deaths, the trial will have approximately 86% power to demonstrate statistical significance at the 2-sided alpha level of 3.334% (with overall alpha for OS 4%) for the

comparison of durvalumab versus SoC, allowing for 1 interim analysis conducted at approximately 85% of the target events. The smallest treatment difference that could be statistically significant will be an average HR of 0.83.

Durvalumab versus SoC (OS in the PD-L1 TC ≥ 25% and low risk of early mortality population)

The sizing assumes a 3 months delay in separation of the OS curves between the two groups, hence the use of an average hazard ratio. If OS at 18 months was 52% with durvalumab and 39% with SoC (with a 13.8-month median OS, i.e., Weibull distribution was assumed with shape parameter = 1.164459) and assuming the true average OS HR is 0.69, an estimated 414 death events (approximately 76% maturity) are expected to have occurred at 46 months from the time the first patient has been randomized with a 25-month recruitment period. With an estimated 414 deaths, the trial will have approximately 87% power to demonstrate statistical significance at the 2-sided alpha level of 0.862% (with overall alpha for OS 1%) for the comparison of durvalumab versus SoC, allowing for 1 interim analysis conducted at approximately 85% of the target events. The smallest treatment difference that could be statistically significant will be an average HR of 0.77.

Secondary OS analyses in PD-L1 TC ≥ 50% population and PD-L1 TC ≥ 50% and low risk of early mortality population will be included in the MTP, as described in Section 4.2.1.

Durvalumab versus SoC (OS in the PD-L1 TC ⩾ 50% population)

If the average true OS HR is 0.75, the study will have 75% power to demonstrate a statistically significant difference at the 5% level (using a 2-sided test) for the OS comparisons when approximately 374 OS events have been observed across the durvalumab and SoC treatment groups, allowing for 1 interim analysis conducted at approximately 85% of the target events.

Durvalumab versus SoC (OS in the PD-L1 TC ⩾ 50% and low risk of early mortality population)

If the average true OS HR is 0.69, the study will have 87% power to demonstrate a statistically significant difference at the 5% level (using a 2-sided test) for the OS comparisons when approximately 298 OS events have been observed across the durvalumab and SoC treatment groups, allowing for 1 interim analysis conducted at approximately 85% of the target events.

Non uniform accrual of patients is assumed when estimating the analysis times. The total proportion of patients randomized at time t following the start of the study is assumed to be $(t/D)^{1.5}$, where D is the accrual duration.

2. ANALYSIS SETS

2.1 Definition of analysis sets

Full analysis set (Intention to treat (ITT))

The full analysis set (FAS) will include all randomized patients (i.e., patients with PD-L1 TC ≥ 25%). The full analysis set 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 FAS will include the subset of patients in the FAS that were recruited from sites located in mainland China.

PD-L1 TC ≥ 25% and low risk of early mortality analysis set

The PD-L1 TC ≥ 25% and low risk of early mortality population consists of patients identified by a prognostic model developed by AstraZeneca as having low risk of early mortality. The model is based on tumor type and six routine laboratory parameters at baseline including neutrophil to lymphocyte ratio (NLR,) neutrophils, albumin, lactate dehydrogenase, gamma glutamyltransferase and aspartate aminotransferase. All six parameters are required in the model. The model produces a score representing a probability of death occurring in ≤ 12 weeks for each patient. The score is then converted to a status that assigns patients to prognostic categories of high or low risk. The cut-off point is a score of 0.649. Patients with a score above the cut-off (> 0.649) are identified as high risk of early mortality (EM), and patients at or below the cut-off (≤ 0.649) are identified as low risk of EM. If any of the six parameters is missing, then the patient will be identified as unknown risk of EM, and as unknown prognostic category of EM (i.e., not included in the PD-L1 TC ≥ 25% and low risk of early mortality analysis set).

China PD-L1 TC ≥ 25% and low risk of early mortality analysis set will include the subset of patients in the PD-L1 TC ≥ 25% and low risk of early mortality analysis set that recruited from sites located in mainland China.

PD-L1 TC ≥ 50% analysis set

The PD-L1 TC ≥ 50% analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1 TC ≥ 50% as defined by the Ventana SP263 PD-L1 IHC assay (i.e., ≥ 50% of tumor cells with membrane positive for PD-L1).

China PD-L1 TC ≥ 50% analysis set will include the subset of patients in the PD-L1 TC ≥ 50% analysis set that recruited from sites located in mainland China.

PD-L1 TC ≥ 50% and low risk of early mortality analysis set

The PD-L1 TC ≥ 50% and low risk of early mortality population consists of patients with PD-L1 TC ≥ 50% and identified by a prognostic model developed by AstraZeneca as having low risk of early mortality.

China PD-L1 TC ≥ 50% and low risk of early mortality analysis set will include the subset of patients in the PD-L1 TC ≥ 50% and low risk of early mortality analysis set that recruited from sites located in mainland China .

Safety analysis set

The safety analysis set (SAS) will consist of all patients who received at least 1 dose of study treatment. Safety data will not be formally analyzed but summarized using the safety analysis set, according to the treatment received, that is, erroneously treated patients (e.g., those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received. Patients in the durvalumab arm who received only SoC chemotherapy and no doses of durvalumab will be summarized in the SoC arm. If a patient received any dose of durvalumab, they will be reported under the durvalumab arm.

China SAS will consist of the subset of patients in the safety analysis set that recruited from sites located in mainland China.

PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set

The PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set will include patients identified by a prognostic model developed by AstraZeneca as having low risk of early mortality and who received at least 1 dose of study treatment.

China PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set will include the subset of patients in the PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set that recruited from sites located in mainland China.

ADA evaluable analysis set

ADA evaluable analysis set consists of all patients in the safety analysis set who have a non-missing baseline ADA and at least one non-missing post-baseline ADA result.

China ADA evaluable analysis set will include the subset of patients in the ADA evaluable analysis set that recruited from sites located in mainland China.

ADA evaluable and low risk of early mortality analysis set

ADA evaluable and low risk of early mortality analysis set consists of all patients in the safety analysis set who have a non-missing baseline ADA and at least one non-missing post-baseline ADA result, and identified by a prognostic model developed by AstraZeneca as having low risk of early mortality.

China ADA evaluable and low risk of early mortality analysis set will include the subset of patients in the ADA evaluable and low risk of early mortality analysis set that recruited from sites located in mainland China.

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	Population
Efficacy data	
OS (primary)	Full analysis set (ITT population, i.e., PD-L1 TC ≥ 25%)
OS (primary)	PD-L1 TC ≥ 25% and low risk of early mortality analysis set
OS	PD-L1 TC ≥ 50% analysis set
OS	PD-L1 TC ≥ 50% and low risk of early mortality analysis set
PFS, ORR, DoR, OS18, OS24, APF12, PFS2, PROs, and symptom endpoints	Full analysis set (ITT population, i.e., PD-L1 TC ≥ 25%)
PFS, ORR, DoR, OS18, OS24, APF12, PFS2, PROs, and symptom endpoints	PD-L1 TC ≥ 25% and low risk of early mortality analysis set
PFS, ORR, DoR, OS18, OS24, APF12, PFS2	PD-L1 TC ≥ 50% analysis set
PFS, ORR, DoR, OS18, OS24, APF12, PFS2	PD-L1 TC ≥ 50% and low risk of early mortality analysis set
Demography	
Demography	Full analysis set (ITT population, i.e., PD-L1 TC ≥ 25%)
Demography	PD-L1 TC ≥ 25% and low risk of early mortality analysis set
Immunogenicity Data	
Immunogenicity	ADA evaluable analysis set
Immunogenicity	ADA evaluable and low risk of early mortality safety analysis set
Safety Data	
Exposure	Safety analysis set
	PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set
AEs	Safety analysis set
	PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set

Outcome variable	Population
Laboratory measurements	Safety analysis set
	PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set
Vital signs	Safety analysis set
	PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set

2.2 Violations and deviations

The important protocol deviations will be listed and summarized by randomized treatment group. IPDs related to COVID-19 will also be summarized. COVID-19 PDs that are not classed as IPDs will also be listed. Deviation 1, below, 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. A per-protocol analysis excluding patients with important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis will 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 any randomized therapy.

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

Eligibility criteria deviations are deviations from the protocol inclusion and exclusion criteria. Post-entry deviations are deviations from the protocol that occurred after the patient was assigned to the study.

The following general categories will be considered important deviations and be listed and discussed in the CSR as appropriate for the study. If a 'deviation bias' sensitivity analysis is conducted then patients with these deviations will be excluded from the sensitivity analysis:

- Deviation 1: Patients randomized but who did not receive study treatment.
- Deviation 2: Patients who deviate from key entry criteria as per the CSP. These are inclusion criteria 3, 4, 5 and exclusion criteria 1, 4, 6.
- Deviation 3: Baseline RECIST scan > 42 days before date of randomization.
- Deviation 4: No baseline RECIST 1.1 assessment on or before date of randomization.

- Deviation 5: Received prohibited concomitant systemic anti-cancer medications. Please refer to the Clinical Study Protocol (CSP) section 7.7 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 to identify those likely have an impact on efficacy.

- Deviation 6: Patients randomized who received treatment other than that to which they were randomized to.

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

In addition to the programmatic determination of the deviations above, monitoring notes or summaries will be reviewed to determine any important post entry deviations that are not identifiable via programming, and to check that those identified via programming are correctly classified. The final classification of deviations will be made at the blinded data review meeting (BDRM) prior to database lock or data freeze. Decisions made at the BDRM will be documented and approved by AstraZeneca prior to analysis.

Misrandomizations in terms of errors in treatment dispensing, in addition to incorrect stratifications, will also be summarised and listed separately to the important protocol deviations. A misrandomization is when a patient is not randomized or treated according to the randomization schedule. It is envisaged that there will be 2 subcategories of this:

- Patients who receive no treatment whatsoever for a period of time due to errors in dispensing of medication. Note, this is not due to tolerability issues where patients may stop taking drug.

- The patient receives a treatment pack with a different code to their randomization code. However, the actual treatment may still match the randomized treatment. For example, a patient is given randomization code 0001, which according to the randomization schedule is durvalumab. However, at the randomization visit they are given treatment pack 0003, but this still contains durvalumab.

The summary will include all patients with a dispensing error but will also include information on how many of those patients received at least one dose of the wrong treatment at any time. 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 misrandomizations will be made on an individual basis with written instruction from the study team leader and/or statistician.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST 1.1 visit responses

For all patients, the RECIST version 1.1 (see further details in Appendix D in the CSP) tumour response data will be used to determine each patient's visit response. It will also be used to determine if and when a patient has progressed and also their best objective response.

Baseline assessments should be performed no more than 28 days before randomization, and ideally should be performed as close as possible to the start of investigational product. Efficacy for all patients will be assessed by objective tumour assessments every 6 weeks for the first 48 weeks (relative to the date of randomization; see Section 4 of the CSP), then every 8 weeks thereafter until confirmed objective disease progression as defined by RECIST 1.1 (irrespective of the reason for stopping treatment/or subsequent therapy). If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits. This schedule is to be followed in order to minimise any unintentional bias caused by some patients being assessed at a different frequency than other patients.

For patients who discontinue study drug due to toxicity in the absence of confirmed objective progression, objective tumour assessments should be continued every 6 weeks for 48 weeks (relative to the date of randomization) then every 8 weeks until confirmed objective disease progression.

Radiologic progression (PD by RECIST 1.1) requires confirmation with a subsequent scan, and the confirmatory scan should occur no earlier than 4 weeks after the prior assessment of progression of disease (PD) in the absence of clinical deterioration. Radiologic progression would be considered confirmed if the following criteria are met:

- $\geq$ 20% increase in sum diameters of target lesions (TLs) compared to the nadir at 2 consecutive visits (with an absolute increase of at least 5 mm)

- And/or significant progression (worsening) of non-target lesions (NTLs) at the confirmatory scan time-point compared with the immediate prior time-point

- And/or significant progression (worsening) of pre-existing new lesions (NLs) at the confirmatory scan time-point compared with the immediate prior time-point (Note: Pre-existing new lesions from earlier time-points are evaluated as NTLs at the confirmatory scan time-point)

- And/or additional new unequivocal lesions (NLs) at the confirmatory scan time-point.

In the absence of significant clinical deterioration, treatment with durvalumab or SoC may continue between the initial assessment of radiologic progression and confirmation of progression. If radiologic progression is not confirmed, then the patient should continue with regularly scheduled imaging assessments. If a patient discontinues treatment (and/or receives

a subsequent cancer therapy) prior to progression, then the patient should still continue to be followed until confirmed objective disease progression.

Categorization of objective tumour response assessment will be based on the RECIST 1.1 criteria of response: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Target lesion progression will be calculated in comparison to when the tumour burden was at a minimum (i.e., smallest sum of diameters previously recorded on study). In the absence of progression, tumour response (CR or PR) and SD will be calculated in comparison to the baseline tumour measurements obtained before starting treatment.

Following confirmed progression, patients should continue to be followed up for survival every 1 to 2 months as outlined in the study plan (Table 4 in the CSP). Subsequent anticancer therapy information will be collected at the time points indicated in Table 4 in the CSP.

3.1.1 Investigator RECIST 1.1-based assessments: Target lesions (TLs)

At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, or PD depending on the status of their disease compared with baseline and previous assessments. Baseline will be assessed within the 28 days prior to randomization. If a patient has had a tumour assessment that cannot be evaluated, then the patient will be assigned a visit response of not evaluable (NE; unless there is evidence of progression in which case the response will be assigned as PD). Endpoints (of PFS) will be derived from the scan dates.

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (except lymph nodes which must have short axis ≥ 15 mm) with CT or 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 suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). If more than one baseline scan is recorded then measurements from the one that is closest and prior to 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 non-target lesions (NTL) at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

Measurable disease (i.e., at least one TL) is one of the entry criteria for the study. However, if a patient with non-measurable disease is enrolled in the study, the evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions. If a patient does not have measurable disease at baseline then the TL visit response will be not applicable (NA).

Table 2 TL visit responses

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

Lymph nodes

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

TL visit responses subsequent to CR

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., 0mm or < 10mm 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 < 10mm.

- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e., 0mm or < 10mm 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 a value of 5mm will be entered into the database and used in TL calculations, unless the radiologist has indicated and 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

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

Any TL (including lymph nodes), which has had 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 tumours:

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

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and 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 subject would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then, if appropriate, a scaled sum of diameters will be calculated (as long as ≤ 1/3 of the TLs have missing measurements), treating the lesion with intervention as missing, 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 also have a value of 0 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 where appropriate (as per step 2 above).

Scaling (applicable only for irradiated lesions/lesion intervention)

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

If ≤ 1/3 of the TL 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	Longest diameter at nadir visit	Longest diameter at follow-up visit
1	7.2	7.1
2	6.7	6.4
3	4.3	4.0
4	8.6	8.5
5	2.5	Intervention
Sum	**29.3**	**26**

Lesion 5 is missing at the follow-up visit.

The sum of lesions 1-4 at the follow-up is 26 cm. The sum of the corresponding lesions at 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 0cm.

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.

3.1.2 Investigator RECIST 1.1-based assessments: 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 tumour burden has increased sufficiently to merit a determination of disease progression. A modest ‘increase’ in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.

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

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

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

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

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

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

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

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

3.1.3 Investigator RECIST 1.1-based assessments: 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 visit response
CR	CR or NA	No (or NE)	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	Non-PD or NE or NA	No (or NE)	PR
SD	Non-PD or NE or NA	No (or NE)	SD
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NE	Non-PD or NE or NA	No (or NE)	NE
NA	CR	No (or NE)	CR
NA	Non-CR/Non-PD	No (or NE)	SD
NA	NE	No (or NE)	NE

3.2 Outcome variables

OS will be evaluated as a primary endpoint from all-cause mortality. The analysis of secondary endpoints, PFS, ORR, DoR, and APF12, will be based on Investigator tumour assessments according to RECIST 1.1. In addition, time to secondary progression (PFS2) will be defined by local clinical practice. Survival rate (OS18 and OS24) will also be analyzed.

All RECIST assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues study treatment or receives another anticancer therapy.

3.2.1 Primary endpoint - overall survival

OS 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 (SUR_DAT, recorded within the SURVIVE module of the eCRF).

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 data cut off), and 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 that for any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment (since the SURVIVE module is only completed for patients off treatment if a survival sweep is not performed).

3.2.2 Secondary endpoints

Investigator RECIST 1.1 based endpoints

All RECIST assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues study treatment or receives another anticancer therapy.

At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, or PD depending on the status of their disease compared with baseline and previous assessments. Baseline will be assessed within the 28 days prior to randomization.

Please refer to Appendix D of the CSP for the definitions of CR, PR, SD, and PD.

3.2.2.1 Progression-free survival

PFS (per RECIST 1.1 using Investigator assessments) 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 anticancer therapy prior to progression (i.e., date of PFS 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 evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more missed

visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the two consecutive missed visits (Note: NE visit is not considered a missed visit).

Given the scheduled visit assessment scheme (i.e. six-weekly for the first 48 weeks then eight-weekly thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 288 (i.e. week 41) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to eight-weekly this will equate to 16 weeks (i.e. take the average of 6 and 8 weeks which gives 7 weeks and then apply same rationale, hence 2 x 7 weeks + 1 week for an early assessment + 1 week for a late assessment = 16 weeks). The time period for the previous RECIST assessment will be from study days 288 to 330 (i.e. week 41 to week 47). From week 47 onwards (when the scheduling changes to eight-weekly assessments), two missing visits will equate to 18 weeks (i.e. 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks).

If the patient has no evaluable visits or does not have baseline data, they will be censored at day 1 unless they die within 2 visits of baseline (2*6 weeks for tumour assessments + 7 days for permitted visit window).

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

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

- The date of progression will be determined based on the earliest of the scan dates of the component that triggered the progression.

- When censoring a patient for PFS, the patient will be censored at the latest of the dates contributing to a particular overall visit assessment.

Note: For target lesions, only the latest scan date is recorded out of all scans performed at that assessment for the target lesions, and similarly for non-target lesions, only the latest scan date is recorded out of all scans performed at that assessment for the non-target lesions.

3.2.2.2 Objective response rate

ORR (per RECIST 1.1 using Investigator assessments) is defined as the percentage of patients with an unconfirmed response of CR or PR. If any patients do not have measurable disease at baseline then the analysis of ORR will exclude these patients, so that the denominator is a subset of the Intent-to-Treat (ITT) population who have measurable disease at baseline. Data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who go off treatment without progression, receive a subsequent therapy, and then respond will not be included as responders in the ORR.

Confirmed ORR which is defined as the percentage of patients with a confirmed response of CR or PR will be derived. A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 4 weeks after the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit.

3.2.2.3 Duration of response

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

The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of CR or PR.

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

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

3.2.2.4 Time from randomization to second progression (PFS2)

Time from randomization to second progression (PFS2) is defined as the time from the date of randomization to the earliest of the progression events (subsequent to that used for the primary variable PFS) or death (i.e., date of PFS2 event or censoring – date of randomization + 1). The date of the first progression will be programmatically determined from investigator assessed data (See Section 3.2.2.1 for details.). The date of second progression will be recorded by the Investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: objective radiological imaging, symptomatic progression or death. RECIST assessments will not be collected for assessment of PFS2. The date of the PFS2 assessment and investigator opinion of progression status (progressed or non progressed) at each assessment will be recorded in the electronic case report form (eCRF). The site will be asked whether the patient has had a second progression event on a regular basis (every 6 weeks for the first 48 weeks relative to the date of randomization and then every 8 weeks thereafter) following the first progression event used for the secondary variable PFS (the first progression) and the status recorded. Actual timing of assessments for a second progression event will, as mentioned, be according to local standard practice. Patients alive and for whom a second disease progression has not been observed should be censored at the last time known to be alive and without a second disease progression, that is, censored at the latest of the PFS or PFS2 assessment date if the patient has not had a second progression or death.

The analysis of PFS2 should include all randomized patients. Patients who have a first progression event but no second event are censored at last available PFS2 assessment. Patients who died as a first PFS are then considered events for PFS2 at the date of death.

Patients who have a first progression event and then die subsequently will have their PFS2 event at date of death. Patients without any first progression event will be censored at their last available scan.

3.2.2.5 Proportion of patients alive at 18 months and 24 months from randomization

The proportion of patients alive at 18 months (OS18) and 24 months (OS24) will be defined as the Kaplan-Meier estimate of OS at 18 and 24 months.

3.2.2.6 Proportion of patients alive and progression free at 12 months

The proportion of patients alive and progression free at 12 months (APF12) will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed using Investigator assessments) at 12 months.

3.2.2.7 Best objective response

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in Appendix D in the CSP. It is the best response a patient has had following randomization but prior to starting any subsequent cancer therapy up until RECIST 1.1 progression or the last evaluable assessment in the absence of RECIST 1.1 progression.

Categorization of BoR will be based on RECIST 1.1 (Appendix D in the CSP) using the following response categories: CR, PR, SD, PD, and NE.

BoR will be determined programmatically based on RECIST 1.1 using all Investigator assessments up until the first progression event, the start of any subsequent cancer therapy or the last evaluable assessment in the absence of progression. For patients whose progression event is death, BoR will be calculated based upon all evaluable RECIST 1.1 assessments prior to death.

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 6 weeks minus 1 week, (i.e., at least 35 days (to allow for an early assessment within the assessment window), after randomization (i.e., study day 36). 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.

The denominator will be consistent with that used in the ORR analysis.

For patients who die with no evaluable RECIST 1.1 assessments, if the death occurs ≤91 days (i.e., 2*(6 weeks) + 1 week) after randomization, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST 1.1 assessments, if the death occurs >91 days (i.e., 2*(6 weeks) + 1 week) after the date of randomization then BoR will be assigned to the NE category.

BoR based on confirmed response will also be derived.

3.2.2.8 Change in tumour size

For supportive purposes percentage change from baseline in tumour size will be derived at each scheduled tumour assessment visit (i.e., week 6, week 12 etc. hereafter referred to as week X for convenience). Best percentage change from baseline in tumour size will also be derived as the biggest decrease or the smallest increase in tumour size from baseline.

This is based on RECIST 1.1 target lesion measurements taken at baseline and at the time point of interest. Tumour size is defined as the sum of the longest diameters of the target lesions for the Investigator data based upon RECIST 1.1 assessments. Target lesions are measurable tumour lesions. Baseline for RECIST 1.1 is defined to be the last evaluable assessment prior to randomization. The change in target lesion tumour 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 tumour size at week X the change in target lesion tumour 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 Investigator data based upon RECIST 1.1 assessments.

3.3 Patient-reported outcome (PRO) variables

Patient reported outcome (PRO) questionnaires will be assessed using the EORTC QLQ-C30 with the QLQ-LC13 module (HRQoL and lung cancer specific symptoms), and CCI
All items/questionnaires will be scored according to published scoring guidelines or the developer's guidelines, if published guidelines are not available.

3.3.1 EORTC QLQ-C30

The EORTC QLQ-C30 consists of 30 questions that can be combined to produce 5 functional scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, pain, and nausea/vomiting), 5 individual items (dyspnea, insomnia, appetite loss, constipation, and diarrhea), and a global measure of health status. The EORTC QLQ-C30 will be scored according to the EORTC QLQ-C30 scoring manual (Fayers et al 2001). An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales/symptom items, each of the functional scales, and the global health status scale in the EORTC QLQ-C30 according to the EORTC QLQ-C30 Scoring Manual. Higher scores on the global health status and functioning scales indicate better health status/function, but higher scores on symptom scales/items represent greater symptom severity.

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: $Score = (1 - [RS - 1] / range) * 100$

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

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

The number of items and item range for each scale/item are displayed in Table 5.

Table 5 Scoring the EORTC QLQ-C30

Scale/item	Scale/item abbreviation	Number of items	Item range	Item numbers
Global health status/ QoL	QL2	2	6	29; 30
Functional scales				
Physical	PF2	5	3	1 to 5
Role	RF2	2	3	6, 7
Cognitive	CF	2	3	20, 25
Emotional	EF	4	3	21 to 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				
Dyspnea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13

Scale/item	Scale/item abbreviation	Number of items	Item range	Item numbers
Constipation	CO	1	3	16
Diarrhea	DI	1	3	17

For each subscale, if <50% of the subscale items are missing, then the subscale score will be derived by ignoring any items with missing values when applying the standard equations (i.e., the raw score will be divided by the number of non-missing items instead of the total number of items on the subscale) (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. If there is evidence that the missing data are systematic, missing values will be handled to ensure that any possible bias is minimized.

Baseline will be defined as the last non-missing assessment prior to randomization.

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 (Osoba et al 1998). For example, a clinically meaningful improvement in physical function (as assessed by EORTC QLQ-C30) is defined as an increase in the score from baseline of ≥10, whereas a clinically meaningful deterioration is defined as a decrease in the score from baseline of ≥10. At each post-baseline assessment, the change in symptoms/functioning from baseline will be categorized as improvement, no change or deterioration as shown in Table 6.

Table 6 Mean change and visit response in health-related quality of life

Score	Change from baseline	Visit response
EORTC QLQ-C30 Global quality of life score	≥+10	Improvement
	≤-10	Deterioration
	Otherwise	No change
EORTC QLQ-C30 symptom scales/items	≥+10	Deterioration
	≤-10	Improvement
	Otherwise	No change
EORTC QLQ-C30 functional scales	≥+10	Improvement
	≤-10	Deterioration
	Otherwise	No change

Time to HRQoL/function/symptom deterioration

Time to deterioration will be analyzed for the all the EORTC QLQ-C30 scales and items (see Table 5). Dyspnea, fatigue and appetite loss from the QLQ-C30 will be key symptoms of interest.

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 function scales or the global health status/HRQoL from baseline of ≥10 or an increase in the symptom scores from baseline of ≥10) that is confirmed at a subsequent visit 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 HRQoL/function/symptom deterioration. Patients with a single deterioration and no further assessments will be treated as deteriorated in the analysis.

Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the HRQoL/function change could be evaluated.

Patients whose HRQoL (as measured by EORTC QLQ-C30) has 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 HRQoL/function could be evaluated. Also, if HRQoL deteriorates after 2 or more missed PRO assessment visits or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where HRQoL/function could be evaluated. If a patient has no evaluable visitsthey will be censored at 1 day unless they die within 2 visits of baseline.

The population for analysis of time to global health status/HRQoL or function deterioration will include a subset of patients who have baseline scores ≥10; the population for analysis of time to symptom deterioration will include a subset of patients who have baseline scores ≤90.

Symptom improvement rate

The symptom 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 (a decrease from baseline score ≥10 for EORTC QLQ-C30 symptom scales) in that symptom from baseline. The denominator will consist of a subset of patients who have a baseline symptom score ≥10. The symptom improvement rate will be derived for the 3 symptom scales and the 5 individual symptom items.

HRQoL/function improvement rate

The HRQoL/function 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 EORTC QLQ-C30 functional scales and global health status/HRQoL) in that scale from baseline. The denominator will consist of a subset of

patients who have a baseline HRQoL/functional scale score ≤90. The improvement rate will be derived for the global health status/ QoL and the 5 functional scales.

3.3.2 Lung cancer module (EORTC QLQ-LC13)

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. The scoring approach for the QLQ-LC13 is identical in principle to that for the symptom scales/items of the QLQ-C30 (see Section 3.3.1), using Table 7.

Table 7 Scoring the EORTC QLQ-LC13

Scale/item	Scale/item abbreviation	Number of items	Item range	Item numbers
Lung cancer symptoms				
Cough	LCCO	1	3	1
Hemoptysis	LCHA	1	3	2
Dyspnea	LCDY	3	3	3 to 5
Site-specific pain				
Pain in chest	LCPC	1	3	10
Pain in arm/shoulder	LCPA	1	3	11
Pain in other parts	LCPO	1	3	12
Treatment-related side-effects				
Sore mouth	LCSM	1	3	6
Dysphagia	LCDS	1	3	7
Peripheral neuropathy	LCPN	1	3	8
Alopecia	LCHR	1	3	9
Pain relief after medication *	LCPR	1	3	13

* Item must only be scored if the answer to the question “Did you take any medication for pain?” is “Yes”.

The dyspnea scale is only used if all 3 items have been scored; otherwise the items are treated as single-item measures.

Baseline will be defined as the last non-missing assessment prior to randomization.

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 QLQ-LC13 (Osoba et al 1998). For example, a clinically meaningful deterioration or worsening in chest pain (as assessed by the QLQ-LC13) is defined as an increase in the score from baseline of ≥10, whereas a clinically meaningful improvement is defined as a decrease in the score from baseline of ≥10. At each post-baseline assessment, the change in symptoms/items from baseline will be categorized as an improvement, no change or deterioration as shown in Table 8.

Table 8 Visit response for health-related quality of life (HRQoL) and disease-related symptoms

Score	Change from baseline	Visit response
QLQ-LC13 symptom scales/items	≥+10	Deterioration
	≤-10	Improvement
	Otherwise	No change

Time to symptom deterioration

Time to symptom deterioration will be defined as the time from the date of randomization until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) that is confirmed at a subsequent visit 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. Patients with a single deterioration and no further assessments will be treated as deteriorated in the analysis.

Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by 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. Also, if symptoms deteriorate after 2 or more missed PRO assessment visits or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated. If a patient has no evaluable visit they will be censored at day 1 unless they die within 2 visits of baseline.

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

Symptom improvement rate

The symptom 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

(a decrease from baseline score ≥10 for QLQ-LC13 symptom scales/items) in that symptom from baseline. The denominator will consist of a subset of patients who have a baseline symptom score ≥10.

3.3.3 CCI

[redacted]

3.3.4 PRO compliance rates

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

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

- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all 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 by 100.

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

3.4 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. Safety data will be summarised from the treatment period for durvalumab alongside the SoC agents.

Data from the treatment period on durvalumab will be compared against SoC in the main presentations of safety data (see Section 4.1). 'On treatment' will be defined as assessments between date of start dose and 90 days following last dose of the immunotherapy agents (i.e., the last dose of durvalumab) on each period of treatment and between date of start dose and 30 days following last dose of the Standard of Care agents. Note that for one version of the safety outputs the period of time after the administration of subsequent therapy will not be considered 'on treatment' (see further Section 4.2.11).

The Safety analysis set and the PD-L1 TC ≥ 25% and low risk of early mortality 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 time of the patient signing the informed consent and 90 days after the last dose of durvalumab and from time of the patient signing the informed consent and 30 days following last dose of the Standard of Care agents. 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 or a pre-existing AE worsening following the first dose of study treatment through to 90 days after the last dose of durvalumab or 30 days after the last dose of the Standard of Care agents.

AEs of special interest and AEs of possible interest

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

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

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

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

Immune-mediated Adverse Events (imAE)

imAE will be identified from both AEs of special interest (AESIs) and AEs of possible interest (AEPIs) based on programmatic rules that consider interventions involving systemic steroid therapy, immunosuppressant use, and/or endocrine therapy (which, in the case of AEPIs, occurs after first considering an Investigator’s causality assessment and/or an Investigator’s designation of an event as immune-mediated). Endocrine therapies include standard endocrine supplementation, as well as treatment of symptoms resulting from endocrine disorders (for example, therapies for hyperthyroidism include beta blockers [e.g., propranolol], calcium channel blockers [e.g., verapamil, diltiazem], methimazole, propylthiouracil, and sodium perchlorate). Further details are provided in the imAE charter.

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

3.4.2 Treatment exposure

Exposure will be defined as follows:

Total (or intended) exposure of durvalumab

- Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of "last dose date of study drug + 27 days" or death date or DCO.

Actual exposure of durvalumab

- Actual exposure is defined as above, but excluding total duration of dose delays

The total (or intended) exposure for each SoC treatment will be calculated using the same principle as above, according to the dose schedule required for each SoC. The total (or intended) exposure will also be summarised by combining the SoC treatments together (i.e, the maximum total exposure among all SoC agents will be considered).

The total (or intended) exposure for each SoC agent is defined as follows:

Total (or intended) exposure of Paclitaxel / Carboplatin / Cisplatin / Pemetrexed

- Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of "last dose date of study drug + 20 days" or death date or DCO.

Total (or intended) exposure of Gemcitabine

- Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of "last dose date of study drug + 6 days (if last dose is day 1 of cycle) or + 13 days (if last dose is day 8 of cycle)" or death date or DCO.

The actual exposure for SoC treatment will be summarised by combining the SoC agents together (i.e., the maximum actual exposure value among all SoC agents will be considered).

Actual exposure of each SoC agent

- Actual exposure is defined as the total treatment duration excluding total duration of dose delays.

Dose reductions are not permitted per Section 6.7 of the CSP for durvalumab, but are allowed per local prescribing information for the SoC agents. The actual exposure calculation makes no adjustment for any dose reductions that may have occurred with the SoC agents.

Exposure will also be measured by the number of cycles received. For all five choices of SoC regimen, a cycle corresponds to a period of 21 days, but for immunotherapy agent a cycle corresponds to one dose of treatment. 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), durvalumab:

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

Calculation of duration of dose delays (for actual exposure), each SoC agent:

Duration of dose delays = Sum of (Date of the dose - Date of previous dose – 21 days)

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.

3.4.3 Dose intensity

Dose intensity will be derived for durvalumab. It will be derived for the SoC agents as a maximum value using the minimum intended dose specified in the protocol. Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to treatment discontinuation.

Relative dose intensity (RDI) will be defined as follows for:

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule.

For subjects on Carboplatin or Gemcitabine or Cisplatin in the SoC arm, RDI will be presented with AUC5 as the intended dose for Carboplatin, the intended dose of 1000mg/m^2 for Gemcitabine and the intended dose of 75 mg/m^2 for Cisplatin.

3.4.4 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 haematology, clinical chemistry, and urinalysis will be collected as described in Section 5.2.1 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 3.4.8 below will be used.

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

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

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

Calculated creatinine clearance (CrCl) will be derived in the reporting database using the Cockcroft-Gault formula:

Creatinine clearance (mL/min) = ([140 – age at randomization] * weight (kg) [* 0.85 if subject is female]) / (72 * serum creatinine (mg/dL))

Plasma creatinine may be used as a substitute where no serum creatinine measurement is available.

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

3.4.5 Time to first subsequent therapy from discontinuation of study treatment

Time to subsequent therapy (excluding palliative radiotherapy) from date of treatment discontinuation is defined as the time from the date of discontinuation of study treatment to

the start date of the first subsequent therapy after discontinuation of treatment. Any patient not known to have had a first subsequent therapy will not have this calculation performed.

3.4.6 ECGs

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

At each time point the Investigator's assessment of the ECG will be collected locally. Heart rate, duration of QRS complex, RR, PR and QT intervals will be collected. If it is necessary to analyse this data then QTcF (Fridericia) will be calculated programmatically using the reported ECG values (RR and QT).

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

For triplicate ECGs, the mean of the three ECG assessments will be used to determine the value at that time point.

3.4.7 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 3.4.8 below will be used.

The denominator in vital signs data should include only those patients with recorded data.

3.4.8 General considerations for safety assessments

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

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the time window should be based on the actual date and not the intended date of the visit.

- All unscheduled visit data 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 for durvalumab (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

Day 141, visit window 127 – 154

Day 169, visit window 155 – 182

Day 197, visit window 183 – 210

Day 225, visit window 211 – 238

Day 253, visit window 239 – 266

Day 281, visit window 267 – 294

Day 309, visit window 295 – 322

Day 337, visit window 323 – 350

Note: Due to the differing assessment schedules the visit windows will be different for the different study treatments and endpoints.

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).
- Listings should display all values contributing to a time point for a patient.
- For visit based summaries:
  - If there is more than one value per patient within a time window then the closest value to the scheduled visit date should be summarised, or the earlier in the event the values are equidistant from the nominal visit date. The listings should highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all post baseline values collected are used including those collected at unscheduled visits regardless of whether or not the value is closest to the scheduled visit date.
  - To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group visit data should only be summarised if the number of observations is greater than the minimum of 20 and > 1/3 of patients dosed.

- For summaries at a patient level, all values should be included, regardless of whether they appear in a corresponding visit based summary, when deriving a patient level statistic such as a maximum.

- Baseline will be defined as the last non-missing measurement prior to dosing with study treatment. For laboratory data, any assessments made on day 1 will be considered pre-dose. Alternatively, if two visits are 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), the average can be taken as a baseline value. For non-numeric laboratory tests (i.e., some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline as this is the most conservative. In the scenario where there are two assessments on day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline. Where safety data are summarised over time, study day will be calculated in relation to date of first treatment.

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

3.5 Biomarker variables

PD-L1 expression status 25-49% (≥ 50%) is defined as 25-49% (≥ 50%) tumour cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control.

3.6 Immunogenicity variables

3.6.1 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 (for durvalumab arm only). 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.

The number and percentage of ADA-evaluable patients in the following ADA categories in the durvalumab arm will be determined. The number of ADA-evaluable patients in the treatment group will be used as the denominator for percentage calculation

- ADA positive at any visit (at baseline and/or post-baseline). The percentage of these patients in a population is known as ADA prevalence.
- 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
- Baseline ADA titer that was boosted by ≥4-fold following drug administration (treatment-boosted ADA)
- Treatment-emergent ADA positive (either treatment-induced ADA positive or treatment-boosted ADA positive). The percentage of these patients in a population is known as ADA incidence.
- Persistently positive ADA (having at least 2 post-baseline ADA positive measurements with >= 16 weeks between first and last positive, or an ADA positive result at the last available post baseline assessment)
- Transiently positive ADA (having at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive)
- nAb positive at any visit (at baseline and/or post-baseline)

3.7 CCI

[redacted]

4. ANALYSIS METHODS

4.1 General principles

The formal statistical analysis will be performed to test the main hypotheses:

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

The two primary endpoints are OS in patients with PD-L1-high expression tumours (PD-L1 TC ≥25%) and OS in patients with PD-L1 TC ≥25% and low risk of early mortality.

One interim analysis to assess the superiority of the durvalumab group (compared to SoC group) in terms of OS will be performed when all of the following conditions have been met

- approximately 85% of the target 521 OS events (approximately 68% maturity) have been achieved across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% population
  And
- approximately 85% of the target 414 OS events (approximately 64% maturity) have been achieved across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% and low risk of early mortality population
  And
- a minimum 12 months follow-up from last patient randomized to the study.

The final (primary) analysis of OS will be performed when all of the following conditions have been met

- approximately 521 OS events (approximately 80% maturity) have occurred across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% population
  And
- approximately 414 OS events (approximately 76% maturity) have occurred across the durvalumab and SoC treatment groups in the PD-L1 TC ≥ 25% and low risk of early mortality population.

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

- Descriptive statistics will be used for all variables, as appropriate, and will be presented by treatment group. Continuous variables will be summarized by the number of observations, mean, standard deviation, 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 for the corresponding 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.
- SAS® version 9.4 will be used for all analyses.

For efficacy variables, baseline is defined as the last visit prior to or on randomization. For ECOG performance status parameter, the last observed measurement prior to or on randomization will be considered the baseline measurement. However, if an evaluable assessment is only available after randomization but before or on 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.

Disposition, baseline characteristics, efficacy and PRO data will be summarized and analyzed based on the FAS and the PD-L1 TC ≥ 25% and low risk of early mortality analysis set. Selected efficacy data will also be summarized based on PD-L1 TC ≥ 50% analysis set and the PD-L1 TC ≥ 50% and low risk of early mortality analysis set. Safety and treatment exposure data will be summarized on the safety analysis set and the PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set. ADA data will be summarised based on the ADA evaluable analysis set and the ADA evaluable and low risk of early mortality analysis set.

Handling of missing/incomplete dates

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

Adverse events and medications

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

Concomitant medications: all medications will be considered as concomitant unless the opposite can be clearly stated.

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

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

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

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

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

- Missing day: impute the last day of the month unless month is same as month of last dose of study drug then impute last dose date;

- Missing day and month: impute 31st December unless year is the same as last dose date then impute last dose date;

- Completely missing:

    - AE: since there is no ongoing flag recorded in eCRF, then assume that AE is still present (i.e. do not impute a date);

    - Medication: if the ongoing flag is yes then the medication is still being taken (i.e. do not impute a date). If the medication has stopped and start date of medication is prior to first dose date then impute the first dose date. If the medication started on or after first dose date and before or on last dose date then impute a date that is after the last dose date. If the medication started after last dose date then impute a date that is after the start of medication.

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

Duration of AE/medication will not be derived using imputed dates.

Birth date

As described in Section 4.2.2, 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 January [given year]. Patients with a missing age value will be included using the mean age (overall FAS).

Imputation of partial death dates

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

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

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

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

Imputation of partial start dates of subsequent anti-cancer therapy

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

Other parameters

No other imputation will be made.

4.2 Analysis methods

Results of all statistical analysis will be presented using appropriate 95% confidence intervals (CIs) and 2-sided p-values, unless otherwise stated.

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

Table 9 Pre-planned statistical and sensitivity analyses to be conducted

Endpoints analyzed	Notes
Overall survival	Stratified log-rank test for: Primary analysis for the ITT population Primary analysis for the PD-L1 TC ≥ 25% and low risk of early mortality population Secondary analysis for the PD-L1 TC ≥ 50% population (stratified only for histology and smoking status) Secondary analysis for the PD-L1 TC ≥ 50% and low risk of early mortality population (stratified only for histology and smoking status) Sensitivity analysis: max-combo test in: • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population
Progression-free survival	Stratified log-rank test for: Secondary analysis using Investigator RECIST 1.1 assessments • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population • PD-L1 TC≥ 50% • PD-L1 TC≥ 50% and low risk of early mortality population

Endpoints analyzed	Notes
Objective response rate	Logistic regression for: Secondary analysis using Investigator assessments (RECIST 1.1) • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population • PD-L1 TC ≥ 50% • PD-L1 TC ≥ 50% and low risk of early mortality population
Duration of response	Kaplan-Meier estimates for Secondary analysis using Investigator assessments (RECIST 1.1) • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population • PD-L1 TC ≥ 50% • PD-L1 TC ≥ 50% and low risk of early mortality population
Proportion of patients alive at 18 months (OS18) and 24 months (OS24)	Kaplan Meier estimates of survival rate at 18 months and 24 months in • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population • PD-L1 TC ≥ 50% • PD-L1 TC ≥ 50% and low risk of early mortality population
Proportion of patients alive and progression free at 12 months	Kaplan Meier estimates of progression free survival at 12 months • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population • PD-L1 TC ≥ 50% • PD-L1 TC ≥ 50% and low risk of early mortality population
Time from randomization to second progression	Stratified log-rank test • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population • PD-L1 TC ≥ 50% • PD-L1 TC ≥ 50% and low risk of early mortality population
Change from baseline in PRO symptoms	Average change from baseline using a mixed model for repeated measures • ITT population • PD-L1 TC ⩾ 25% and low risk of early mortality population
Time to deterioration (EORTC QLQ-C30 and QLQ-LC13 endpoints)	Stratified log-rank test in • ITT population • PD-L1 TC ≥ 25% and low risk of early mortality population

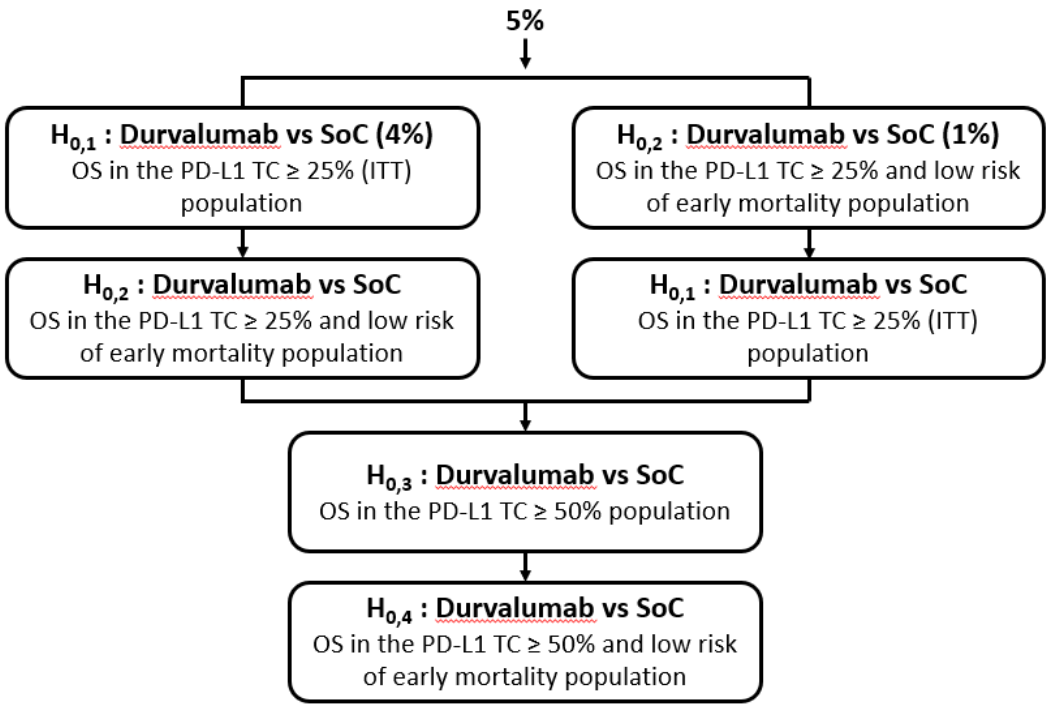
4.2.1 Multiple testing strategy

In order to strongly control the type I error at 5% (2-sided), a multiple testing procedure (MTP) with gatekeeping strategy will be employed across the 2 primary OS endpoints of OS in ITT population, OS in the PD-L1 TC ≥ 25% and low risk of early mortality population and selected secondary endpoints.

OS will be tested at 1 interim and a final time point. The OS tests for the same comparison (i.e., shown in 1 box in the MTP) will be considered as 1 test family. As long as one test in the family can be rejected, the family is rejected thus the assigned total alpha to the family can be recycled to the next MTP level. The testing procedure stops when the entire test mass is allocated to non-rejected hypotheses. Implementation of this pre-defined ordered testing procedure, including recycling, will strongly control type I error at 5% (2-sided) (Burman et al 2009).

The overall 5% type 1 error will be initially split between the 2 primary endpoints: an alpha level of 4% will be allocated to the analysis of OS in the ITT population and an alpha level of 1% will be allocated to the analysis of OS in the PD-L1 TC ≥ 25% and low risk of early mortality population. The MTP is shown in Figure 3. If either primary comparison is significant, the available alpha level will be recycled to the other primary comparison. If primary comparisons are both significant, the available alpha level will be recycled to secondary comparisons. Below is the detailed procedure.

Figure 3 Multiple testing procedure for controlling type I error



Note: alpha recycling between two dual primary OS comparisons. Two-sided alpha.

Test OS (ITT population) at 4% significance level and OS (PD-L1 TC ≥ 25% and low risk of early mortality population) at 1%, respectively.

A. If neither of the 2 tests is statistically significant, accept $H_{0,1}$ and $H_{0,2}$, and stop procedure
B. If $H_{0,1}$ is not statistically significant at 4%, but $H_{0,2}$ is statistically significant at 1% level, reject $H_{0,2}$ and recycle the 1% to $H_{0,1}$, and retest $H_{0,1}$ at 5% level.
   a) If $H_{0,1}$ is not statistically significant at 5% level, accept $H_{0,1}$ and stop.
   b) If $H_{0,1}$ is statistically significant at 5% level, reject $H_{0,1}$. Then test $H_{0,3}$ at 5%.
      i. If $H_{0,3}$ is not statistically significant at 5% level, accept $H_{0,3}$ and stop the procedure.
      ii. If $H_{0,3}$ is statistically significant at 5% level, reject $H_{0,3}$ and test $H_{0,4}$ at 5%.
         - If $H_{0,4}$ is statistically significant at 5% level, reject $H_{0,4}$ and stop the procedure.
         - Otherwise, accept $H_{0,4}$ and stop the procedure.
C. If $H_{0,2}$ is not statistically significant at 1%, but $H_{0,1}$ is statistically significant at 4% level, reject $H_{0,1}$ and recycle the 4% to $H_{0,2}$, and retest $H_{0,2}$ at 5% level.
   a) If $H_{0,2}$ is not statistically significant at 5% level, accept $H_{0,2}$ and stop.

b) If $H_{0,2}$ is statistically significant at 5% level, reject $H_{0,2}$. Then test $H_{0,3}$ at 5%. The rest is the same as in B(b) i. and ii.

D. If both tests are statistically significant, reject $H_{0,1}$ and $H_{0,2}$, and test $H_{0,3}$ at 5%. The rest is the same as in B(b) i. and ii.

The alpha level allocated to OS will be controlled at the interim and final time points by using the Lan and Demets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the alpha level applied at the interim depends upon the proportion of information available. Separate Lan DeMets spending functions will be used for 2 primary endpoints and secondary endpoints.

4.2.2 Analysis of two primary endpoints - overall survival

Primary analysis of OS in the ITT population will be performed using a stratified log-rank test, adjusting for level of PD-L1 expression (25-49% versus ≥50%) and histology and smoking status (squamous versus non-squamous + never smoker versus non-squamous + former/current smoker) for generation of the p-value. The effect of durvalumab versus SoC will be estimated by the HR from stratified Cox proportional hazards model together with its corresponding (1-adjusted alpha) × 100% CI (with adjustments both without and with alpha recycling – see Appendix A), 95% CI (Cox 1972), with ties are handled by Efron approach and the CI is calculated using a profile likelihood approach. Primary analysis of OS in the PD-L1 TC ≥ 25% and low risk of early mortality population will be conducted in the same manner.

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. Sensitivity analysis might be performed based on the values in the eCRF page if more than 10% randomized patients have discrepancies in stratification factors between IVRS and eCRF data.

A secondary analysis of OS will be performed using a stratified log-rank test, adjusting for only histology and smoking status using the PD-L1 TC ≥ 50% analysis set, and the PD- L1 TC ≥ 50% and low risk of early mortality analysis set. The corresponding HR and CI will be estimated using a stratified Cox model.

Kaplan-Meier plots 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 for each treatment.

The boundaries (i.e., adjusted alpha levels) for the treatment comparison at the interim and final analyses for OS will be derived based upon the exact number of OS events using the Lan and DeMets approach that approximates the O'Brien Fleming spending function (see Section 4.2.1).

The assumption of proportionality will be assessed, initially only with regards to the primary treatment comparisons. Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time-dependent covariate to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect will be described by presenting piecewise HR calculated over distinct time-periods. In such circumstances, the HR can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found, this may be a result of treatment-by-covariate interactions, which will be investigated. In addition, the KM curve along with landmark analyses (e.g., one year OS rate) will also help in understanding the treatment benefit.

Sensitivity analyses and additional supportive summaries

A sensitivity analysis for OS will examine the censoring patterns to rule out attrition bias, which is achieved by a Kaplan-Meier plot of time to censoring, where the censoring indicator of OS is reversed. This analysis will be conducted in both the ITT and the PD-L1 TC ≥ 25% and low risk of early mortality population.

The stratified max-combo test will be conducted as a sensitivity analysis on the OS data in both the ITT and the PD- L1 TC ≥ 25% and low risk of early mortality population, to test for treatment differences in the case of nonproportional hazards. The analysis will be based on adaptive procedure involving selection of best test statistics with log rank test ($G_{0,0}$) and the Fleming-Harrington (FH) test ($G_{0,1}$, $G_{1,0}$ and $G_{1,1}$) with alpha correction (Lin et al 2019).

The stratified piecewise Cox proportional hazards model will be fitted to generate piecewise HR over distinct time-periods (0-6 months, >6 months) and corresponding 95% CI for the ITT, PD- L1 TC ≥ 25% and low risk of early mortality, PD-L1 TC ⩾ 50%, and PD-L1 TC ⩾ 50% and low risk of early mortality populations. Additional time-periods might be considered as appropriate.

A sensitivity analysis may be conducted to assess for the potential impact of COVID-19 deaths on OS in both the PD-L1 TC ≥ 25% and the low risk of early mortality populations. This will be assessed by repeating the OS analysis where any patient who had a death with primary or secondary cause of COVID-19 infection, or a COVID-19 infection reported as a fatal AE, will be censored at their COVID-19 infection death date.

Supportive analyses might be conducted in the PD-L1 TC 25-49% population.

A sensitivity analysis based on unstratified log rank test and unstratified Cox proportional hazards model will be conducted for OS in the PD-L1 TC ≥ 25% and low risk of early mortality population. Supportive analyses might be conducted to characterize efficacy for patients that are identified as having high risk of early mortality (by a prognostic model developed by AstraZeneca) as exploratory assessment only, due to the expected limited number of patients identified.

To further assess impact of unknown prognostic category of EM on primary analysis (PD-L1 TC ≥ 25% and low risk of early mortality population), multiple imputation methods may be performed to estimate the expected treatment effect in the PD-L1 TC ≥ 25% and low risk of early mortality population and may be reported out of CSR.

Besides the secondary endpoints of OS 18 and OS 24, proportion of patients alive at 36 months (OS36) might also be summarized (using the Kaplan-Meier curve) if data permit and presented by treatment arm.

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

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

- Time from randomization to the date of death or to the date of censoring (data last known to be alive) by treatment group and in all patients
- Time from randomization to date of censoring (data last known to be alive) in censored patients only, presented by treatment group and in all patients.

Additional supportive analyses might also be considered as deemed appropriate.

Subgroup analyses

Subgroup analyses will be conducted comparing OS between durvalumab versus SoC in the following subgroups of the FAS and the PD-L1 TC ≥ 25% and low risk of early mortality population (but not limited to):

- Sex (male versus female)
- Age at randomization (<65 versus ≥65 years of age)

  This will be determined from the date of birth (BIRTHDAT in the DEM module) and date of randomization (RND_DAT in the CRIT1 module) on the eCRF at screening. 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 January [given year]). Patients with a missing age value will be included using the mean age (overall FAS) and categorised accordingly.
- Level of PD-L1 expression (25-49% vs ≥50%, from IVRS)
- Histology and smoking status (squamous versus non-squamous + never smoker versus non-squamous + former/current, from IVRS)
- Histology type (squamous versus non-squamous)

- Smoking status (Never smoker versus former/current smoker)
- Region (China versus other)
- Race (Asian versus Non-Asian)
- Baseline Liver Metastases (Yes versus No)
- Baseline brain/CNS Metastases (Yes versus No)
- Baseline ECOG (0 versus >=1)
- Subsets of FAS only: Baseline risk status of early mortality identified by a prognostic model developed by AstraZeneca (Low risk, High risk, Unknown)

The subgroup analyses for the stratification factors (level of PD-L1 expression, and histology and smoking status) will be based on the values entered into the IVRS, all other factors will be based on values recorded on the eCRF as indicated above.

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

Other baseline variables may also be assessed if there is clinical 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 and/or predictive factors. For each subgroup, the HR and 95% CI will be calculated from an unstratified Cox proportional hazards model that only contains a term for treatment. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties, using the by statement to obtain HR and 95% CI for each subgroup level separately.

These hazard ratios and associated two-sided 95% CIs will be summarised 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 in a subgroup), the relationship between that subgroup and OS will not be formally analysed. In this case, only descriptive summaries will be provided.

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

Effect of covariates on HR estimate

Cox proportional hazards modeling will be employed to assess the effect of covariates on the HR estimate. Before embarking on more detailed modeling, an initial model will be constructed, containing treatment and the stratification factors alone, to ensure that any output from the Cox modeling 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 age at randomization, sex, race, ECOG, liver metastases, and number of target lesions at baseline (1,2,>=3).

This multivariate Cox model will be performed in both the ITT and the PD- L1 TC ≥ 25% and low risk of early mortality population. Additional covariates might be considered as appropriate.

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 subgroups (i.e. covariates).

This is performed by comparing the fit of a Cox proportional hazards model including treatment, all covariates, and all covariate-by treatment interaction terms, with one that excludes the interaction terms and will be assessed at the 2-sided 10% significance level. If there are not more than 5 events per stratum (i.e., within each strata of the treatment*covariate interaction (2 treatments * 2 levels of the covariate = 4 stratum)) a pre-defined pooling strategy should be applied to the covariate. If the pooling strategy does not meet the event criteria then the covariate-by-treatment interaction term should be omitted from the model. Moreover, if the covariate does not have more than 5 events per level of covariate then the main effect of the covariate will also be excluded. If the fit of the model is not significantly improved then it will be concluded that overall the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and 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.

Any quantitative interactions between treatment and subgroups, identified using this procedure, will also be tested to rule out any qualitative interaction using the approach of Gail and Simon 1985.

Impact of switching to other immunotherapies (or other potentially active investigational agents) on OS analyses

4.2.3 Progression-free survival

The PFS analysis will be based on the programmatically derived RECIST 1.1 using the Investigator tumour assessments. The secondary analysis will be performed using a stratified log-rank test adjusting for level of PD-L1 expression 25-49% versus ≥50% and histology and smoking status (squamous versus non-squamous + never smoker versus non-squamous + former/current smoker) for the generation of p-value. The effect of durvalumab versus SoC will be estimated by the HR from stratified Cox proportional hazards model together with its corresponding 95% CI (Cox 1972), with ties are handled by Efron approach and the CI is calculated using a profile likelihood approach. This analysis will be performed in the analysis sets as mentioned in Table 1.

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

The following sections describe the planned supportive/sensitivity analyses for the ITT population, selected analyses will be conducted for the PD- L1 TC ≥ 25% and low risk of early mortality population.

Kaplan-Meier plots of PFS will be presented by treatment group, where appropriate. Summaries of the number and percentage of patients experiencing a PFS event and the type of event (RECIST 1.1 or death) will be provided along with median PFS for each treatment.

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

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

Sensitivity Analyses

Sensitivity analyses 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 will be analyzed using a log-rank test. For patients whose death was treated as 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 even in highly asymmetric assessment schedules (Sun and Chen 2010).

Attrition bias will be assessed by repeating the PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following 2 or more non-evaluable tumour assessments will be included. In addition, patients who take subsequent therapy prior to progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy. In addition, a Kaplan-Meier plot of time to censoring, where the censoring indicator of PFS is reversed will be presented.

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the secondary and sensitivity analyses of progression free survival.

Additional supportive summaries/graphs

In addition, 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 tumour assessment interval plus 2 weeks prior to the DCO date.

Additionally, summary statistics will be given for the number of days from censoring to data cut-off for all censored patients.

A summary of the duration of follow-up will be summarised 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 1.1 assessment prior to progression will be presented for each treatment group.

Summaries of the number and percentage of patients who miss two or more consecutive RECIST 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.

Additional supportive analyses might be considered as deemed appropriate.

Subgroup analyses and Effect of covariates on HR estimate

Subgroup analyses and a forest plot will be generated comparing PFS between treatments. Subgroup analyses will be conducted comparing PFS (per RECIST 1.1 using Investigator assessment) between durvalumab versus SoC using the same approach as primary OS.

The effect of covariates upon the HR estimate will be analysed for PFS using the same approach as primary OS.

4.2.4 Objective response rate

The ORR will be based on the programmatically derived RECIST 1.1 using the Investigator tumour data. ORR will be compared between durvalumab versus SoC using logistic regression models adjusting for the same factors as the primary endpoint (level of PD-L1 expression, histology and smoking status). The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour durvalumab) together with its associated profile likelihood CI (e.g., using the option 'LRCI' in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). This analysis will be performed in the analysis sets as mentioned in Table 1.

If there are not enough responses for a meaningful analysis using logistic regression then a Fisher's exact test using mid p-values will be presented.

The mid-p-value modification of the Fisher exact test amounts to subtracting half of the probability of the observed table from Fisher's p-value.

Fisher's exact test mid p-value = Two-sided p-value – (Table probability / 2)

Summaries will be produced that present the number and percentage of patients with a tumour response (CR/PR) with associated two-sided 95% confidence interval. Overall visit response data will be listed for all patients (i.e., the FAS). For each treatment group, best objective response (BoR) will be summarized by n (%) for each category (CR, PR, SD, PD, and NE). No formal statistical analysis is planned for BoR.

Analysis of confirmed ORR/BoR will also be conducted in the similar manner.

4.2.5 Duration of response

Descriptive data will be provided for the duration of response, on patients with documented response, (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). This analysis will be performed in the analysis sets as mentioned in Table 1.

DoR will also be assessed for the subgroup of patients who had a confirmed response.

4.2.6 Proportion of patients alive at 18 months and 24 months

The OS18 and OS24 (where 18 and 24 months equate to study day 548 and 731, respectively) will be summarized (using the Kaplan-Meier curve) and presented by treatment arm.

This analysis will be performed in the analysis sets as mentioned in Table 1.

4.2.7 Proportion of patients alive and progression free at 12 months

The APF12 (where 12 months equate to study day 366) will be summarized (using the Kaplan-Meier curve) and presented by treatment arm.

This analysis will be performed in the analysis sets as mentioned in Table 1.

4.2.8 Time from randomization to second progression

PFS2 is defined as the time from the date of randomization to the earliest of the progression event subsequent to that used for the PFS endpoint or death. It will be assessed using local practice guideline. PFS2 will be analyzed using a stratified log-rank test, using the same methodology as described for the PFS endpoint. The HR for the treatment effect together with its 95% CI will be presented. Kaplan-Meier plots will be presented by treatment group. Summaries of the number and percentage of patients who have an event as well as who were censored will be provided along with the medians for each treatment.

This analysis will be performed in the analysis sets as mentioned in Table 1. The following sections describe the planned supportive analyses for the ITT population, selected analyses will be conducted for the PD- L1 TC $\geq$ 25% and low risk of early mortality population.

For supportive purposes, the time to the start of first subsequent therapy or death will be analyzed using the same methodology and model. 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 first subsequent therapy or death will be presented by treatment group and the time between progression and starting subsequent therapy will be assessed. The time from first subsequent therapy to second progression or death might also be analyzed. No multiplicity adjustment will be applied as these are viewed as supportive endpoints.

A summary table of first subsequent therapies by treatment group will be provided.

4.2.9 Change in tumour size

The absolute values and percentage change in target lesion tumour size from baseline will be summarized using descriptive statistics and presented at each time point for each treatment group. The best change in target lesion tumour 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 summarised and presented for each treatment group.

Tumour size will also be presented graphically using waterfall plots for each treatment group, to present each subject's best percentage change in tumour 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 tumour size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively. On each of the waterfall plots the histology classification (Squamous versus All other) of each patient will be indicated.

Additional waterfall plots showing percentage change in tumour size at specific time points may be produced if it is felt that these are warranted to provide greater clarity.

The above outputs will be programmed for the Investigator RECIST 1.1 assessments.

4.2.10 Patient reported outcomes

Patient-reported outcomes will be assessed using the EORTC QLQ-C30, Lung cancer symptoms will be assessed with the QLQ-LC13 module.

Treatment efficacy will be evaluated primarily on what patients and clinicians consider the primary symptoms of lung cancer (NSCLC working group material, presentation at ASCO 2016): cough, dyspnoea, pain (in the chest, pain in other parts of the body) as well as fatigue and appetite loss. The assessments of cough, dyspnoea and chest pain as assessed by the EORTC QLQ LC13 and fatigue and appetite loss from EORTC QLQ C30 will be used as secondary efficacy endpoints.

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

PRO analyses will be performed in the analysis sets as mentioned in Table 1.

4.2.10.1 EORTC QLQ-C30

Time to deterioration will be analyzed using a stratified log-rank test, using the same methodology as described for the primary OS endpoint.

The effect of durvalumab versus SoC will be estimated by the HR together with its corresponding CI (i.e. 95% CI). Kaplan-Meier plots will be presented by treatment group. Summaries of the number and percentage of patients who have an event as well as who were censored will be provided along with the medians for each treatment.

A summary of the symptom improvement rate for each of the 3 symptom scales and the 5 individual symptom items will be produced. Similarly, a summary of QoL/function improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/QoL will be produced. Symptom improvement rates will be analyzed by comparing between treatment arms using a logistic regression model adjusting for the same factors as OS. The odds ratio (an odds ratio greater than 1 will favor durvalumab) together with its associated profile likelihood 95% CI (e.g. using the option 'LRCI' in SAS procedure GENMOD) for each scale/item will be presented graphically on a forest plot. If there are very few responses in 1 treatment arm, a Fisher's exact test will be considered.

Summaries of original and change from baseline values of each symptom scale/item, the global HRQoL score and each functional domain will be reported by visit 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 visit for each ordinal

item (in terms of the proportion of patients in the categories of improvement, no change, and deterioration) will also be produced for each treatment group.

4.2.10.2 EORTC QLQ-LC13

Time to deterioration will be analyzed using a stratified log-rank test, using the same methodology as described for the primary OS endpoint.

The effect of durvalumab versus SoC will be estimated by the HR together with its corresponding CI (i.e. 95% CI). Kaplan-Meier plots will be presented by treatment group. Summaries of the number and percentage of patients who have an event as well as who were censored will be provided along with the medians for each treatment.

A summary of the symptom improvement rate for each of the 6 individual symptom items will be produced. Symptom improvement rates will be analyzed by comparing treatment arms using a logistic regression model adjusting for the same factors as OS. The odds ratio (an odds ratio greater than 1 will favor durvalumab) together with its associated profile likelihood 95% CI (e.g. using the option 'LRCI' in SAS procedure GENMOD) for each scale/item will be presented graphically on a forest plot. If there are very few responses in 1 treatment arm, a Fisher's exact test will be considered.

Summaries of original and change from baseline values of each symptom (dyspnoea, cough, hemoptysis, chest pain, arm/shoulder pain, other pain) and each treatment-related side effect (sore mouth, dysphagia, peripheral neuropathy, alopecia and pain relief after medication) will be reported by visit 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 visit for each ordinal symptom item and side effect item (in terms of the proportion of patients in the categories of improvement, no change, and deterioration) will also be produced for each treatment group.

4.2.10.3 Mixed models repeated measures of change from baseline in PRO symptoms

In addition to the time to deterioration endpoints listed above the following longitudinal endpoints are of interest: change from baseline in EORTC QLQ-C30 global health status/QoL, physical functioning, fatigue and appetite loss; LC13 dyspnoea, cough and chest pain. These are not part of the main multiple testing procedure and are considered a separate set of PRO endpoints from the time to deterioration endpoints listed above. A Bonferroni adjustment to the significance level will be applied to control the overall Type I error at the 5% level to the five primary PRO measures of cough, dyspnoea and chest pain as assessed by the EORTC QLQ-LC13 and fatigue and appetite loss as assessed by the EORTC QLQ-C30.

Change from baseline in these pre-specified PRO symptom scores of dyspnea, cough, chest pain, fatigue and appetite loss, global health status and physical functioning will be analyzed using a mixed model for repeated measures (MMRM) analysis making use of all data from baseline up to 12 months or first disease progression (whichever occurs first). The analysis will be to compare the average treatment effect from the point of randomization until PD or 12 months (whichever is earlier) unless there is excessive missing data (defined as >75% missing

data). It is acknowledged that patients will discontinue treatment at different time points during the study and that this is an important time with regards to symptoms and HRQoL data collection. To account for this, and in order to include the discontinuation and follow up visits, a generic visit variable will be derived for each subject 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 in order to select only those visits occurring within the first 12 months of randomization or until PD.

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

Generic visit	Study Day	
	Patient X	Patient Y
Baseline	Baseline	Baseline
1	29	29
2	57	50 (discontinuation)
3	85	85
4	113	113
5	130 (discontinuation)	141
6	169	169

The MMRM model will include treatment, age at randomization (<65 vs ≥65 years of age), sex (male vs female), smoking history (smoker vs non-smoker), visit and the interaction between treatment and visit as fixed factors, baseline as a covariate and further adjusted for the interaction between baseline and visit. 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, (1-adjusted alpha) x 100% CI (for the 5 key symptoms only) and p-value. Graphical presentation of the estimated LSMEAN mean change from baseline at each timepoint may also be produced as appropriate.

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures

will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, autoregressive and compound symmetry.

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

An effect size estimate to interpret the magnitude of the effect and potential therapeutic benefit will be further specified in the payer analysis plan.

4.2.10.4 CCI

[redacted]

4.2.11 CCI

[redacted]

4.2.12 Safety data

Safety and tolerability data will be presented by actual treatment group based on safety analysis set and PD-L1 TC $\geq$ 25% and low risk of early mortality safety analysis set separately. Safety data will be summarized only. No formal statistical analyses will be performed on the safety data.

Data from all cycles of treatment will be combined in the presentation of safety data. AEs (both in terms of MedDRA preferred terms and CTCAE grade) will be listed individually by patient. The number of patients experiencing each AE will be summarized by treatment group and CTCAE grade. Additionally, data presentations of the rate of AEs per person-years at risk may be produced.

Other safety data will be assessed in terms of physical examination, clinical chemistry, hematology, vital signs, and ECGs. Exposure to durvalumab and SoC will be summarized. Time on study, dose delays/interruptions and dose reductions will also be summarized.

The following sections describe the planned safety summaries based on the safety analysis set. Selected summaries will be repeated for the PD-L1 TC $\geq$ 25% and low risk of early mortality safety analysis set. Additional safety tables may be required to aid interpretation of the safety data.

Adverse Events

All AEs, both in terms of current MedDRA preferred term and CTCAE grade, will be summarised descriptively by count (n) and percentage (%) for each treatment group. The current MedDRA dictionary will be used for coding. 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 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 immunotherapy agents (i.e., the last dose of durvalumab) / 30 days following discontinuation of the Standard of Care agent or until the initiation of the first subsequent anti-cancer therapy (excluding palliative radiotherapy) following discontinuation of treatment (whichever occurs first) will be used for reporting of all of the AE summary tables. This will more accurately depict AEs attributable to study treatment only as a number of AEs up to 90 days following discontinuation of the immunotherapy agents / 30 days following discontinuation of the Standard of Care agent in patients who have gone onto receive subsequent therapy during this follow up period maybe more likely to be attributable to the subsequent therapy.

However, to assess the longer term toxicity profile, some of the AE summaries might also be produced containing AEs observed up until 90 days following discontinuation of the immunotherapy agents / 30 days following discontinuation of the Standard of Care agent (i.e., without taking subsequent anti-cancer therapy into account).

A selection of AE summaries may also be produced containing AEs (by system organ class and preferred term) observed from the initiation of the first subsequent anti-cancer therapy following discontinuation of study treatment until 90 days following discontinuation of immunotherapy agents / 30 days following discontinuation of the Standard of Care agent (i.e., summarising those AEs experienced by patients taking subsequent therapy during the AE collection follow-up window post discontinuation of study treatment). These outputs will only be produced if the number of AEs observed warrant the inclusion of such outputs for interpretational purposes. Any data post 90 days last dose for durvalumab / 30 days following discontinuation of Standard of Care agents will be presented in a separate summary that presents any events that occur prior to dosing or starting more than 90/30 days after discontinuing treatment.

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 any on 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 treatment (as determined by the reporting investigator)
- AEs by SOC, PT and maximum CTCAE grade
- AEs with maximum CTCAE grade 3 or 4
- AEs possibly related to study treatment with maximum CTCAE grade 3 or 4
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, possibly related to study treatment (as determined by the reporting investigator)
- AEs with outcome of death
- AEs with outcome of death possibly related to study treatment (as determined by the reporting investigator)
- AEs by outcome
- All SAEs
- All SAEs possibly related to study treatment (as determined by the reporting investigator)
- SAEs leading to discontinuation of study treatment
- AEs leading to discontinuation of study treatment
- AEs leading to discontinuation of study treatment, possibly related to study treatment (as determined by the reporting investigator)
- AEs leading to hospitalization
- AEs leading to dose delay of study treatment
- AEs of special interest/AEs of possible interest
- AEs of special interest/AEs of possible interest possibly related to study treatment (as determined by the reporting investigator)

- Immune mediated AEs (as determined by the reporting investigator)
- Infusion reaction AEs (as determined by the reporting investigator)

Summaries of other significant AEs may be produced. For example, a summary and listing of COVID-19 infections may be produced.

An overall summary of the number and percentage of patients in each category will be presented, as well as an overall summary of the number of episodes in each category. In addition, a truncated AE table of most common AEs showing all events that occur in at least 5% of patients in any treatment group will be summarised by preferred term, by decreasing frequency. Another table showing most common AEs with CTCAE grade 3 or higher will be produced (at least 1% of patients in any treatment group). Similarly, a table showing most common AEs with maximum CTCAE grade 3 or 4 will be produced (at least 1% of patients in any treatment group). This cut-off may be modified after review of the data. When applying a cut-off (i.e., x %), the raw percentage should be compared to the cut-off, no rounding should be applied first (i.e., an AE with frequency of 4.9% will not appear if a cut-off is 5%). Summary statistics showing the time to onset and the duration of the first AE may also be presented as appropriate.

Each AE event rate (per 100 patient years) will also be summarised by preferred term within each system organ class for the output summarising all AEs. For each preferred term, the event rate is defined as the number of patients with that AE divided by the total drug exposure of patients at risk of AE, multiplied by 100. The denominator is calculated as the total over each patient of days from first dose to the last day of study medication.

Fluctuations observed in CTCAE grades during study will be listed for those AEs.

In addition, all AEs will be listed.

Deaths

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

- Total number of deaths
- Death before starting study treatment
    - Related to disease under investigation only
    - AE with outcome of death only
    - Unknown reason for death
- Death after starting study treatment

- Related to disease under investigation only
- Related to disease under investigation and a TEAE with outcome of death
  - TEAE onset on or prior to subsequent therapy or no subsequent therapy
  - TEAE onset after start of subsequent therapy
- TEAE with outcome of death only
  - TEAE onset on or prior to subsequent therapy or no subsequent therapy
  - TEAE onset after start of subsequent therapy
- Death after end of safety follow up period (i.e., 90 days following the date of last dose of durvalumab or 30 days following after the date of last dose of SoC treatment)
  - Related to disease under investigation and an AE with outcome of death
  - AE with outcome of death only
  - Not related to disease under investigation
  - Unknown relationship to disease under investigation
- Unknown reason for death
- Other deaths

Separate summaries of deaths on-treatment or during safety follow-up period will be produced.

Adverse events of special interest and possible interest

Preferred terms used to identify adverse events of special interest and possible interest (as defined in Section 3.4.1) will be listed before DBL and documented in the Study Master File. Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI/AEPI grouping. Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

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

- Any ASEI/AEPI

- Any AESI/AEPI by SOC, PT and maximum CTCAE grade
- Any AESI/AEPI of maximum CTCAE grade 3 or 4
- Any serious AESI/AEPI
- Any AESI/AEPI with outcome of death
- Any AESI/AEPI possibly related to study treatment
- Any AESI/AEPI leading to concomitant medication use (steroids)
- Any AESI/AEPI leading to concomitant medication use (high dose steroids)
- Any AESI/AEPI leading to concomitant medication use (endocrine therapy)
- Any AESI/AEPI leading to concomitant medication use (other immunosuppressants)
- At least one AESI/AEPI leading to discontinuation of study treatment

An overall AESI/AEPI summary will be presented, including number and percentage of patients in each of these categories. Any AESI/AEPI presented by outcome will also be provided.

Infection AEs

Infection AEs will be summarized by pooled terms and PTs in two ways: (1) using MedDRA HLGT/HLT pooled terms (2) Custom pooled terms (pneumonia, sepsis and urinary tract infections). The following summaries will be reported for both HLGT/HLT pooled terms and custom pooled terms and PTs:

- Infection AEs (including event rate)
- Infection AEs by maximum reported CTCAE grade
- Serious Infection AEs
- Infection AEs presented by outcome
- Infection AEs of maximum CTCAE grade 3 or 4
- Infection AEs with outcome of death
- Infection AEs leading to discontinuation of any study treatment
- Infection AEs leading to dose delay/interruption of any study treatment

An overall infection AE summary will be presented, including the number and percentage of patients in each of these categories, and additionally possibly related infection AEs, possibly related serious infection AEs, possibly related infection AEs of maximum CTCAE grade 3 or 4, possibly related infection AEs with outcome of death, possibly related infection AEs leading to discontinuation of any study treatment.

Immune-mediated Adverse events (imAEs)

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

Laboratory assessments

Data obtained up until the 90 days following discontinuation of durvalumab (i.e., the last dose) or 30 days following discontinuation of the Standard of Care agent or until the initiation of the first subsequent anti-cancer therapy (excluding palliative radiotherapy) following discontinuation of 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 durvalumab or 30 days following discontinuation of the Standard of Care agent 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 immunotherapy agent or up until 30 days following discontinuation of the Standard of Care agent (i.e., without taking subsequent therapy into account).

Any data post 90 days after last dose for durvalumab or post 30 days after last dose for Standard of Care agents will not be summarised.

Data summaries will be provided in International System (SI) of 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.

Box-plots of absolute values by week, and box-plots of change from baseline by week, may be presented for certain parameters, if warranted after data review.

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

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

- Haematology: Haemoglobin (low), Leukocytes (low), Lymphocytes (low), absolute Neutrophils count (low), absolute Platelets count

- Clinical chemistry: ALT (high), AST (high), ALP (high), Total bilirubin (high), Albumin (low), Magnesium – hypo and – hyper, Sodium – hypo and – hyper, Potassium – hypo and – hyper, Corrected calcium – hypo and – hyper, Glucose – hypo and – hyper, Creatinine (high), GGT (high), Amylase (high), Lipase (high)

For the parameters with no CTCAE grading that are listed in the CSP, shift tables from baseline to worst value on-treatment will be provided. Urinalysis data will be listed.

Additional analyses of six lab parameters (NLR, neutrophils, albumin, lactate dehydrogenase, gamma glutamyltransferase and aspartate aminotransferase) may be generated as appropriate.

Reversibility of creatinine clearance (CrCl) will be tabulated. Shift table in creatinine clearance from baseline to minimum value on treatment will be provided.

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, and >20x ULN during the study

  - ALT or AST ≥3x ULN and Total bilirubin ≥2x ULN during the study (Potential Hy's law): The onset date of ALT or AST elevation should be prior to or on the date of Total Bilirubin elevation

- 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 at any visit.

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.

Assessment of Thyrotoxicity

After the discontinuation of the study medication, the thyroid function tests, TSH, T3 and T4, are evaluated up to 30 days after last dose of study medication, hence, the analysis of thyroid function tests will be based on data up to 30 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 summarised using descriptive statistics at each scheduled assessment time.

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

ECGs

ECG data will be listed. Overall evaluation of ECG is collected in terms of normal or abnormal, and the relevance of the abnormality is termed as “clinically significant” or “not clinically significant”.

Vital signs

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

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

Time to Subsequent Therapy from discontinuation of study treatment

A descriptive summary will be produced for time to subsequent therapy from discontinuation of study treatment.

Other Safety Data

Data from positive pregnancy tests will be listed only.

Overdose listing will also be presented.

4.2.13 WHO performance status

All WHO performance status will be summarised over time for both the ITT and PD-L1 TC ≥ 25% and low risk of early mortality population.

4.2.14 Immunogenicity analysis

Immunogenicity results will be listed by patient and a summary will be provided of the number and percentage of patients who develop detectable anti-durvalumab antibodies based on the ADA-evaluable population and PD-L1 TC ≥ 25% and low risk of early mortality ADA-evaluable population. The immunogenicity titre and neutralizing ADA data will be listed for samples confirmed positive for the presence of anti-durvalumab antibodies. AEs in ADA positive patients by ADA positive category will be listed and summarized.

The effect of immunogenicity on efficacy and safety may be evaluated if it provides further support for benefit risk assessment and if data allow.

4.2.15 CCI [redacted]

This will be assessed using similar summary and graphical representations to those that are outlined for the efficacy outputs.

PD-L1 expression determined by immunohistochemistry will be reported in the CSR. Sensitivity analyses for the PD-L1 dispenser issues are detailed in Appendix B. CCI [redacted]

4.2.16 Demographic and baseline characteristics data

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

- Patient disposition (including screening failures and reason for screening failure)

- Important protocol deviations
- For FAS only: Inclusion in analysis populations
- Demographics (age, age group [<50, >=50-< 65, ≥ 65 - <75 years and ≥ 75 years], sex, race and ethnicity)
- Patient characteristics at baseline (height, weight, weight group [<70, 70 to 90, >90], body mass index (BMI) and body mass index group [<18.5, >=18.5-<25, >=25-<30, >=30])
- Patient recruitment by region, country and centre
- Previous disease-related treatment modalities
- Number of regimens of previous chemotherapy at baseline
- Previous anti-cancer therapy
- Disease characteristics at baseline (WHO performance status, primary tumour location, overall disease classification, best response to previous therapy, liver metastases, and PD-L1 subgroups)
- Disease characteristics at diagnosis (primary tumour location, histology and AJCC staging)
- Extent of disease at baseline
- TNM classification at diagnosis and at baseline
- Relevant medical history (past and current)
- Relevant surgical history
- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer therapy
- Nicotine use, categorised (never, current, former)
- Stratification factors as per IVRS and eCRF data

The WHO drug dictionary will be used for the coding of concomitant medications and prior and subsequent systemic anti-cancer treatments. Majority demographic and baseline characteristics tables will be repeated for the PD-L1 TC ≥ 25% and low risk of early mortality

analysis set. Selected demographic and baseline characteristics tables might be repeated for the PD-L1 TC ≥ 50%, and PD-L1 TC ≥ 50% and low risk of early mortality analysis sets.

4.2.17 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.
- Number and reasons for dose delays and infusion interruptions of durvalumab and number and reasons for dose delays/interruptions, dose reductions and dose modifications for the Standard of Care agents. Dose 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.
- For SoC arm only: number of patients that switched treatment (switch from one regimen to another regimen) and number of patients that switched between Cisplatin and Carboplatin and vice versa
- RDI (relative dose intensity) of durvalumab and Standard of Care agents.
- Exposure over time will be plotted.

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

For SoC patients that received pemetrexed + cisplatin or pemetrexed + carboplatin, patients who have not progressed after 4 cycles will be eligible for pemetrexed maintenance therapy. A summary table will be produced by actual treatment group, and additionally in the SoC arm, by patients who received the pemetrexed maintenance and who didn't receive the pemetrexed maintenance.

Majority summaries will be repeated for the PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set.

4.2.18 Subsequent therapy

Subsequent therapies received after discontinuation of study treatment will have summaries produced by treatment group.

4.2.19 China sub-group analysis

China subgroup analysis will be performed using the same methodology as for the entire study population. The primary OS analyses will be conducted on China FAS and China PD-L1 TC ≥ 25% and low risk of early mortality analysis set using same method outlined in Section 4.2.2. Efficacy analyses of the secondary OS endpoints will be performed in the China PD-L1 TC ≥ 50% analysis set, and China PD-L1 TC ≥ 50% and low risk of early mortality analysis set. Efficacy analyses of the other secondary endpoints PFS, ORR, DoR, APF12 and PFS2 using Investigator assessments according to RECIST 1.1, OS18 and OS24 will be performed for China FAS, China PD-L1 TC ≥ 25% and low risk of early mortality analysis set, China PD-L1 TC ≥ 50% analysis set, and China PD-L1 TC ≥ 50% and low risk of early mortality analysis set. Key demographic and baseline characteristics, PROs and immunogenicity analysis will be repeated for China subgroup as well.

All statistical analyses for China subgroup will be considered exploratory. No adjustment for multiplicity will be made and so the multiple testing procedure and splitting the alpha detailed in Section 4.2.1 will not be followed.

Majority safety and tolerability analyses outlined in Sections 4.2.12 and 4.2.17 will be repeated for China safety analysis set and China PD-L1 TC ≥ 25% and low risk of early mortality safety analysis set.

4.2.20 Coronavirus Disease 2019 (COVID-19)

Depending on the extent of any impact, summaries and listings of data relating to subjects diagnosed with COVID-19, and impact of COVID-19 on study conduct (in particular missed visits, delayed or discontinued IP, and other protocol deviations) may be generated.

5. INTERIM ANALYSES

5.1 Analysis methods

Interim safety monitoring will be conducted by an IDMC. Details of the plan and communication process will be provided in an IDMC Charter.

One OS interim analysis is planned for durvalumab versus SoC in the ITT population as well as in the PD- L1 TC ≥ 25% and low risk of early mortality population. The interim analysis of OS will be conducted when approximately 444 OS events have occurred (approximately 85% of the target 521 OS events, i.e., approximately 68% maturity) in the ITT population; approximately 350 OS events have occurred (approximately 85% of the target 414 OS events, i.e., approximately 64% maturity) in the PD- L1 TC ≥ 25% and low risk of early mortality population and a minimum 12 months follow-up from last patient randomized to the study (whichever occurs later). This analysis will be performed by an IDMC.

The Lan DeMets spending function that approximates an O'Brien Fleming approach will be used to account for multiplicity introduced by including the 1 interim analysis for superiority

(Lan and DeMets 1983). A detailed calculation of alpha allocation at IA and FA is provided in Appendix A.

The criterion for superiority is a statistically significant improvement in OS at the interim analysis.

If the interim analyses indicate superiority, then subsequent analyses of the further secondary endpoints will be performed.

If the interim results do not meet the criterion of stopping for superiority, then follow-up will continue. OS will be retested at the final analysis.

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."

Details of the IDMC plan and communication process is provided in the IDMC Charter.

5.2 Independent Data Monitoring Committee

This study will use an external IDMC to assess ongoing safety analyses, and to perform the formal interim analysis of OS. The committee will meet approximately 6 months after the study has started or after the first 30 patients have been randomized and received at least 2 cycles of treatment, whichever occurs first, to review the safety data from the study. The IDMC will meet at least every 6 months thereafter until a decision to unblind the study is made. In addition, the IDMC will review the unblinded interim analysis summaries of efficacy data. Following each meeting, the IDMC will report to the sponsor and may recommend changes in the conduct of the study.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AstraZeneca/MedImmune and do not have any major conflict of interest.

Following the reviews, the IDMC will recommend whether the study should continue unchanged, be stopped, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca/MedImmune. The report will include the recommendation and any potential protocol amendments, and will not contain any unblinding information.

The final decision to modify or stop the study will sit with the sponsor. The sponsor or IDMC may call additional meetings if at any time there is concern about the safety of the study.

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

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

6. CHANGES OF ANALYSIS FROM PROTOCOL

There are no changes made to the analysis specified in the protocol.

7. REFERENCES

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

Chen, et al 2014
Chen et al, The Epidemiology of Lung Cancer in China, Lung Cancer China, 2014

Ciuleanu et al 2009
Ciuleanu T, Brodowicz T, Zielinski C, Kim JH, Krzakowski M, Laack E, et al. Maintenance pemetrexed plus best supportive care versus placebo plus best supportive care for non-small-cell lung cancer: a randomized, double-blind, phase 3 study. Lancet 2009;374(9699):1432-40.

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

Ellis et al 2008
Ellis S, Carroll KJ, Pemberton K. Analysis of duration of response in oncology trials. Contemp Clin Trials 2008;29(4):456-65.

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.

Fehrenbacher et al 2016
Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, Phase 2, randomized, controlled trial. Lancet 2016; published online 09 March 2016.

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;30(29):3373-86

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

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

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

Lan and DeMets 1983
Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983; 70 (3):659-663

Lin et al 2019
Ray S. Lin, Ji Lin, Satrajit Roychoudhury, Keaven M. Anderson, Tianle Hu, Bo Huang, et al. Alternative Analysis Methods for Time to Event Endpoints Under Nonproportional Hazards: A Comparative Analysis. Statistics in Biopharmaceutical Research. DOI: 10.1080/19466315.2019.1697738.

CCI [redacted]

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;16(1):139-44.

Paz-Ares et al 2013
Paz-Ares LG, de Marinis F, Dediu M, Thomas M, Pujol JL, Bidoli P, et al. PARAMOUNT: Final overall survival results of the phase III study of maintenance pemetrexed versus placebo immediately after induction treatment with pemetrexed plus cisplatin for advanced nonsquamous non-small-cell lung cancer. J Clin Oncol 2013;31(23):2895-902.

Pintilie 2006.
Competing risks: A practical perspective. Wiley, 2006.

Robins 1993
Robins JM. Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers. Proceedings of the Biopharmaceutical Section, American Statistical Association 1993; 24-33.

Robins and Tsiatis 1991
Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics-Theory and Methods 1991; 20(8):2609-31.

Scagliotti et al 2008
Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol 2008;26(21)3543-51.

Sun and Chen 2010
Sun X, Chen C. Comparison of Finkelstein's Method with the conventional approach for interval-censored data analysis. Stat Biopharm Res 2010;2(1):97-108.

Whitehead and Whitehead 1991
Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 1991;10(11):1665-77.

8. APPENDIX

Appendix A**: Alpha Level Calculation**

The alpha allocated at IA and FA will depend on the actual number of death events observed at these timepoints. Hence, the alphas in the tables below are examples of different possible scenarios under the assumption that 85% of the target events occur at IA.

Comparison	# of Events targeted at FA	# of Events at IA (information fraction)	Without consideration of alpha recycling			With consideration of alpha recycling		
			Overall alpha	alpha for IA	alpha at FA	Overall alpha	alpha for IA	alpha at FA
Durva vs SoC in ITT population	521	444 (85%)	4%	2.325%	3.334%	5%	3.010%	4.144%
Durva vs SoC in PD-L1 TC ≥ 25% and low risk of early mortality population	414	350 (85%)	1%	0.466%	0.862%	5%	3.010%	4.144%

Appendix B: CCI [redacted]

CCI [redacted]

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: d419ac00002-sap-ed-3		
Document Title:	Statistical Analysis Plan Edition 3	
Document ID:	Doc ID-003560997	
Version Label:	4.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
16-Aug-2021 08:15 UTC	PPD	Content Approval
16-Aug-2021 06:35 UTC	PPD	Author Approval

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