AstraZeneca

A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 in Combination with Tremelimumab Therapy or MEDI4736 Monotherapy Versus Standard of Care Platinum-Based Chemotherapy in First-Line Treatment of Patients with Advanced or Metastatic Non-Small-Cell Lung Cancer (NSCLC) (MYSTIC)

A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 in Combination with Tremelimumab Therapy or MEDI4736 Monotherapy Versus Standard of Care Platinum-Based Chemotherapy in First-Line Treatment of Patients with Advanced or Metastatic Non-Small-Cell Lung Cancer (NSCLC) (MYSTIC)

PPD

Study Statistician

22/06/2017

Date

A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 in Combination with Tremelimumab Therapy or MEDI4736 Monotherapy Versus Standard of Care Platinum-Based Chemotherapy in First-Line Treatment of Patients with Advanced or Metastatic Non-Small-Cell Lung Cancer (NSCLC) (MYSTIC)

PPD

Global Product Statistician

6/20/2017

Date

TABLE OF CONTENTS   PAGE

CCI ███

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse event
ALK	Anaplastic lymphoma kinase
APF12	Proportion of patients alive and progression free at 12 months from first dose
Baseline	Refers to the most recent assessment of any variable prior to dosing with study treatment/randomisation (as appropriate)
BICR	Blinded independent central review
BoR	Best objective response
CI	Confidence Interval
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)
DoR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EDoR	Expected Duration of Response
EGFR	Epidermal growth factor receptor
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
EQ-5D	EuroQoL 5-dimension utility index
EQ-5D-3L	EuroQoL 5-dimension, 3-level health state utility index
CCI	
FAS	Full analysis set
HR	Hazard ratio
HRQoL	Health-related Quality of Life

Abbreviation or special term	Explanation
IDMC	Independent data monitoring committee
IP	Investigational Product
ITT	Intention to Treat
IVRS	Interactive Voice Response System
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
MRI	Magnetic Resonance Imaging
MTP	Multiple testing procedure
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
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
PDx	Pharmacodynamic(s)
PFS	Progression-free survival
PFS2	Time from randomisation to second progression
PR	Partial Response
PRO	Patient Reported Outcome
QoL	Quality of Life
QLQ-LC13	Quality of Life Lung Cancer Module; 13 item self administered questionnaire from the EORTC for lung cancer
QTcF	QT interval (corrected for heart rate using Fredericia's correction)
RECIST 1.1	Response Evaluation Criteria In Solid Tumours, Version 1.1
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
TL	Target lesions

Abbreviation or special term	Explanation
WHO	World Health Organization

AMENDMENT HISTORY

Date	Brief description of change
20 June 2017	• Clarification on the interim efficacy boundary calculation for the OS comparison of MEDI4736 + tremelimumab versus SoC in PD-L1–positive$_{1\%}$ population
13 June 2017	• Updates to the definition of corresponding safety analysis set in PD-L1-positive$_{25\%}$ population and PD-L1-positive$_{1\%}$ population
	• Updates to Table 2 on the analysis of the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations
	• Updates to study population and demographic data will be summarized in FSA, and FAS for patients within the positive$_{25\%}$ and positive$_{1\%}$ populations, correspondingly.
	• Updates to safety analyses will be performed in Safety Analysis Set, and safety analysis set for patients within the positive$_{25\%}$ and positive$_{1\%}$ populations, correspondingly.
	• Updates to PFS/OS/ORR additional analyses
	• Clarification that PFS subgroup analysis will also be performed for the comparison between MEDI4736 monotherapy versus SoC
	• Updates to the PFS multivariate Cox model to evaluate the treatment effect adjusting for any potential imbalances in baseline prognostic factors
	• Clarification that OS subgroup analysis will also be performed for the comparison between MEDI4736 monotherapy versus SoC
	• Updates to additional analysis using Cox proportional hazards models to determine the effect of covariates on the OS HR estimate
	• Updates to additional analysis using Cox proportional hazards models to determine the consistency of OS treatment effect
	• Updates to ORR subgroup analysis

Date	Brief description of change
12 April 2017	In line with the Clinical Study Protocol (CSP) Amendments v6-7: - Clarifications in Objectives section to note inclusion of additional co-primary endpoint in OS and additional populations for the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations in the primary and secondary objectives. - Study design section updated to reflect the introduction of the additional populations for the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations in the revised primary and secondary objectives. - Updates to the power calculations appropriate to these updated objectives. Inclusion of modified rule for the timing of the primary PFS analysis CCI ███████████████ - Update to the definitions of the analysis sets due to introduction of the additional populations for the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations in the revised primary and secondary objectives - Updates to Table 2 to reflect the changes to the objectives and the introduction of the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations - Clarification in Secondary Endpoints section that ORR for investigator assessments is based on subjects with measurable disease at baseline - Clarification in Secondary Endpoints section of the details of the algorithm for determining time to second progression - Clarification in Secondary Endpoints section of the details of the algorithm for determining Best Objective Response - Clarification that PRO endpoints will be analysed using PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations as well as FAS - Clarifications in PRO endpoints section regarding specific endpoints for symptom deterioration. Clarification that PRO data is collected at assessment periods (via ePRO tool) and not collected at visits and specific details provided on how to determine 2 missed assessments. - CCI ████████████████ - Update of Analysis Methods section to reflect use of the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations for the revised co-primary and secondary objectives/endpoints and requisite updates to the multiple testing strategy - Update to analysis of co-primary endpoint PFS to reflect the primary and secondary endpoints based on PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations and the changes to relevant confidence intervals due to the revised multiple testing strategy

Date	Brief description of change
12 April 2017 (cont)	• Subgroup analyses CCI ███████████████ • Clarification of the analysis method for the subgroup analyses based on separate Cox models for each subgroup CCI ██████████ • Update to analysis of co-primary endpoint OS to reflect the primary and secondary endpoints based on PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations and the changes to relevant confidence intervals due to the revised multiple testing strategy • Update to analysis of all secondary endpoints to reflect the secondary objectives based on PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations • CCI ████████████████ • Analyses of safety data updated to include appropriate analyses based on the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations. • Interim analysis details updated to reflect the change in primary objectives now based on the PD-L1-positive$_{25\%}$ population as primary In line with project wide developments • Update to Adverse Event analyses to confirm sub-analyses of CTC of grade 3 or higher will only summaries grades 3 and 4 (and exclude grade 5)

Date	Brief description of change
19 April 2016	In line with the Clinical Study Protocol (CSP) Amendments v3-v5: - Clarifications in Objectives section to note inclusion of co-primary endpoints, use of BICR for PFS and updates to set of secondary objectives. - Study design section updated to reflect increase in sample size and allowing treatment to progression for both combination and monotherapy treatment groups - Changes in Primary and Secondary variable section to reflect use of co-primary endpoint and use of BICR. - Sample size section updated to reflect increase in sample size based on changes to the treatment comparisons of interest and inclusion of 2 formal interim analysis for OS endpoint. - Change in requirements for re-treatment only in the combination arm - Update of Analysis Methods section to reflect use of co-primary endpoints, the updates to secondary endpoints and requisite updates to the multiple testing strategy - Changes to Analysis Methods to note use of Cox proportional hazards for calculation of HR and CI and addition of additional subgroup for PD-L1 status (using 10% cut-off for positive/negative split) - Change to definition of key and supportive PRO endpoints. - Interim analysis for Overall Survival added In line with project wide developments - Changes in duration of Adverse Event and laboratory reporting period to include up to 90 days for immunotherapy agents and 30 days for SoC - Removal of PID, PID2 and RDI2 and increased clarification of the derivations of RDI - Clarification of summaries required for deaths Other Changes - Edits to category definitions for elevated liver enzymes - Clarification of definition of total and actual exposure for immunotherapy and SoC treatments and calculation of dose delays

1. STUDY DETAILS

1.1 Study objectives

All objectives will be evaluated for all patients, unless otherwise indicated, and where appropriate for patients with PD-L1 (programmed death ligand 1)-positive$_{25\%}$, PD-L1 positive$_{1\%}$ and/or PD-L1–low/negative NSCLC. The assessment of progression-free survival (PFS) and overall survival (OS)) in patients with PD-L1positive$_{25\%}$ will be considered co-primary objectives. Section 3.5 provides the definitions of these PD-L1 subsets:

- Positive$_{25\%}$:- $\geq$ 25% PD-L1 membrane-expression in tumoral tissue
- Positive$_{1\%}$:- $\geq$ 1% PD-L1 membrane-expression in tumoral tissue
- Low/Negative:- < 25% PD-L1 membrane-expression in tumoral tissue
- Negative:- < 1% PD-L1 membrane-expression in tumoral tissue

1.1.1 Primary objective

Primary Objective:	Outcome Measure:
To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of PFS in patients with PD-L1-positive$_{25\%}$ NSCLC	PFS using Blinded Independent Central Review (BICR) assessments according to RECIST 1.1
To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of OS in patients with PD-L1-positive$_{25\%}$ NSCLC	OS
To assess the efficacy of MEDI4736 monotherapy compared to SoC in terms of OS in patients with PD-L1 positive$_{25\%}$ NSCLC	OS

1.1.2   Secondary objectives

Secondary Objectives:	Outcome Measures:
To further assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of PFS, OS, objective response rate (ORR), duration of response (DoR), proportion of patients alive and progression free at 12 months from randomization (APF12) and time to second progression (PFS2)	OS in patients with PD-L1- $positive_{1\%}$ NSCLC and all patients PFS in patients with PD-L1-$positive_{1\%}$ NSCLC and all patients, using BICR assessments according to RECIST 1.1 ORR, DoR, and APF12 in patients with PD-L1-$positive_{25\%}$NSCLC, patients with PD-L1-$positive_{1\%}$ NSCLC and all patients using BICR assessments according to RECIST 1.1 PFS2 in patients with PD-L1-$positive_{25\%}$ NSCLC, patients with PD-L1-$positive_{1\%}$ NSCLC and all patients using local standard clinical practice[a]
To assess the efficacy of MEDI4736 monotherapy compared to SoC in terms of PFS, OS, ORR, DoR, APF12 and PFS2	PFS in patients with PD-L1-$positive_{25\%}$ NSCLC using BICR assessments according to RECIST 1.1 ORR, DoR, and APF12 in patients with PD-L1-$positive_{25\%}$NSCLC using BICR assessments according to RECIST 1.1 PFS2 in patients with PD-L1-$positive_{25\%}$ NSCLC using local standard clinical practice
To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to MEDI4736 monotherapy in terms of PFS, OS and ORR	PFS and ORR in patients with PD-L1-$positive_{25\%}$ and PD-L1-$positive_{1\%}$ NSCLC and all patients using BICR assessments according to RECIST 1.1 OS in patients with PD-L1-$positive_{25\%}$and $positive_{1\%}$ NSCLC and all patients
To assess disease-related symptoms and health related quality of life (HRQoL) in patients treated with MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy compared to SoC using the EORTC Q 30-item core quality of life questionnaire, Version 3 (LQ-C30 v3) and the 13-item lung cancer quality of life questionnaire (QLQ-LC13) module	EORTC QLQ-C30 EORTC QLQ-LC13 Changes in World Health Organization (WHO)/Eastern Cooperative Oncology Group (ECOG) performance status will also be assessed.
To assess the pharmacokinetics (PK) of MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy	Concentration of MEDI4736 and tremelimumab in blood and non-compartmental PK parameters, such as peak concentration and trough (as data allow; sparse sampling)
To investigate the immunogenicity of MEDI4736 and tremelimumab	Presence of anti-drug antibodies (ADAs) for MEDI4736 and tremelimumab

[a]   PFS2 will be 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 based on Investigator tumor assessments

16

1.1.3 Safety objective

Safety Objective:	Outcome Measures:
To assess the safety and tolerability profile of MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy compared to SoC in the first-line setting for treatment of advanced or metastatic patients with NSCLC	Adverse events (AE), physical examinations, laboratory findings, and vital signs

1.1.4 Exploratory objectives

Exploratory Objectives:	Outcome Measures:
CCI ███████████████████	███████████████████
███████████████████	███████████████████
███████████████████	███████████████████
███████████████████	███████████████████
███████████████████	███████████████████
███████████████████	███████████████████
███████████████████	███████████████████
███████████████████	███████████████████

17

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, global Phase III study to determine the efficacy and safety of MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy versus platinum-based SoC chemotherapy in the first-line treatment of patients with EGFR and ALK wild-type advanced or metastatic NSCLC. A schematic diagram of the overall study design is shown in Figure 1, and a detailed study flow chart is shown in Figure 2.

This study will enroll approximately 1850 patients at sites in North America, Asia, Australia, and Europe to randomize approximately 1092 patients to treatment (including approximately 160 patients with PD-L1-positive$_{25\%}$ NSCLC in each treatment group).

Patients will provide a tumor tissue sample at enrollment (newly acquired or archived sample <3 months old) to determine PD-L1 expression status (defined by the CCI in which ≥25% and ≥1% PD-L1–membrane expression in tumoral tissue are considered relevant positive subgroups, <25% is considered low/negative for PD-L1 expression and <1% is considered as negative for PD-L1 expression; these are referred to hereafter as patients with PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$, PD-L1-low/negative and PD-L1-negative tumors, respectively).

Patients will be randomized in a 1:1:1 ratio in a stratified manner according to PD-L1 tumor expression status (PD-L1-positive$_{25\%}$, versus PD-L1-low/negativetod) and histology (squamous versus non-squamous) to receive treatment with MEDI4736 + tremelimumab combination therapy, MEDI4736 monotherapy, or SoC therapy. Doses and treatment regimens are described in Section 7.2 of the CSP. Assessments will be conducted as indicated in Table 2, Table 3, and Table 4 in the CSP.

Figure 1 **Overall study design**

| Stratified randomization factors: |
| 1. PD-L1 tumor expression status (positive versus negative)[a] |
| 2. Histology (squamous versus non-squamous) |
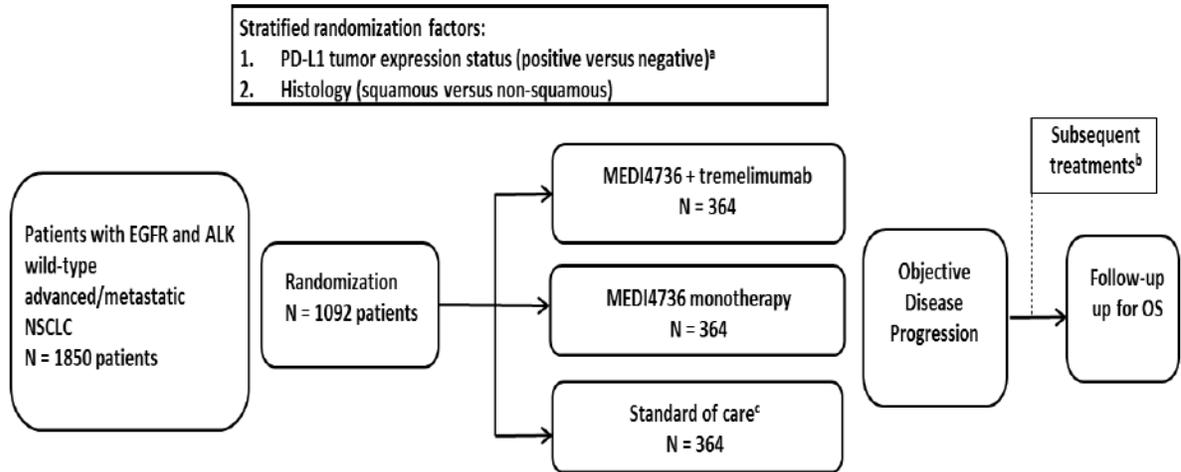
[a] Stratification by PD-L1 membrane-expression in tumoural tissue (≥25% , <25%). Sites will be supplied with PD-L1 status upon request at disease progression.

[b] Offer of standard chemotherapy per Investigator discretion.

[c] 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 prior to randomization.
b	Tumour assessments 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). (See Section 5.1 of CSP for more information.)

Independent Data Monitoring Committee (IDMC)

An IDMC comprised of 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, 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 at least every 6 months thereafter.

Details on the IDMC are provided in Section 5.1 and full details of the IDMC procedures and processes can be found in the IDMC Charter.

1.3　　Number of patients

The study will plan to enroll approximately 1850 patients in order to randomize 1092 eligible patients 1:1:1 to MEDI4736 + tremelimumab combination therapy, MEDI4736 monotherapy, or SoC. The 1092 patients will comprise approximately 160 patients who have PD-L1-positive$_{25\%}$ tumors in each treatment group).

The study is sized to characterize the PFS and OS benefit of MEDI4736 in combination with tremelimumab versus SoC in patients with EGFR and ALK wild-type advanced or metastatic NSCLC in patients with PD-L1-positive$_{25\%}$ tumours and OS benefit of MEDI4736 monotherapy versus SoC in patients with PD-L1-positive$_{25\%}$ tumors. CCI

Two interim analyses of OS will be performed; the first at the time of the primary PFS analysis and the second when 80% of the target OS events have occurred. The alpha will be split between the 3 OS 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.

The primary PFS analysis for superiority will be performed when both of the following conditions have been met:

- Approximately 231 PFS events have occurred across the MEDI4736 + tremelimumab and SoC treatment groups in patients with PD-L1-positive$_{25\%}$ tumours (72% maturity) AND

- CCI ████████████████████████████████████ to the

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

- Approximately 225 OS events have occurred across the MEDI4736 + tremelimumab and SoC treatment groups in patients with PD-L1-positive$_{25\%}$ tumours (70% maturity) AND

- Approximately 225 OS events have occurred across the MEDI4736 monotherapy and SoC treatment groups in patients with PD-L1-positive$_{25\%}$ tumours (70% maturity)

MEDI4736 + tremelimumab versus SoC (PFS in PD-L1-positive$_{25\%}$ population)

CCI

CCI

MEDI4736 + tremelimumab versus SoC (OS in PD-L1-positive$_{25\%}$ population)

CCI

MEDI4736 monotherapy versus SoC (OS in PD-L1-positive$_{25\%}$ population)

CCI

MEDI4736 + tremelimumab versus SoC (OS in PD-L1–positive$_{1\%}$ population)

CCI

CCI

MEDI4736 + tremelimumab versus SoC (OS, all patients)

CCI
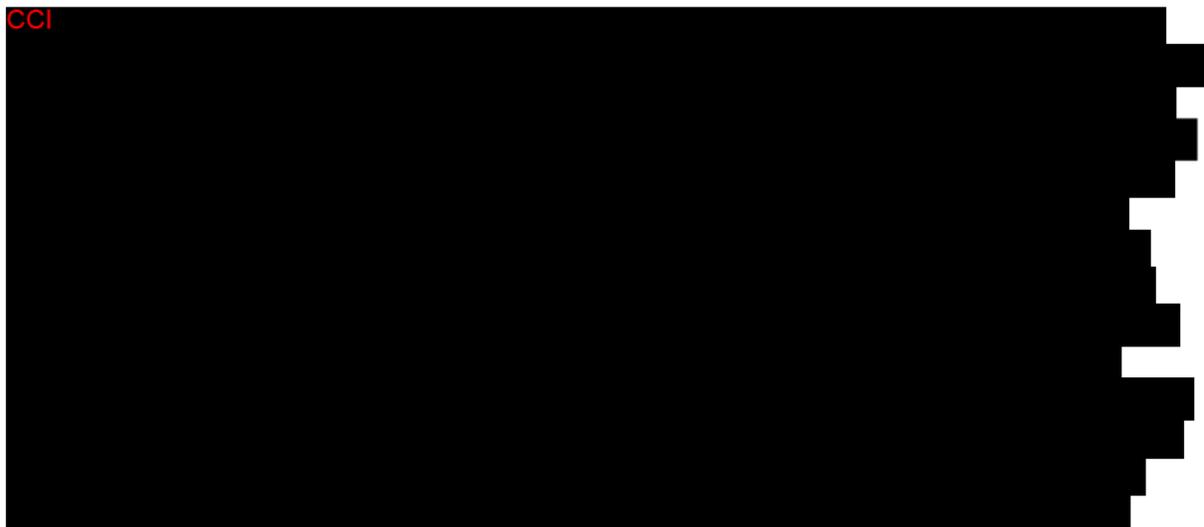
MEDI4736 monotherapy versus SoC (PFS in PD-L1-positive$_{25\%}$ population)

CCI

MEDI4736 + tremelimumab versus SoC (PFS in PD-L1-positive$_{1\%}$ population)

CCI
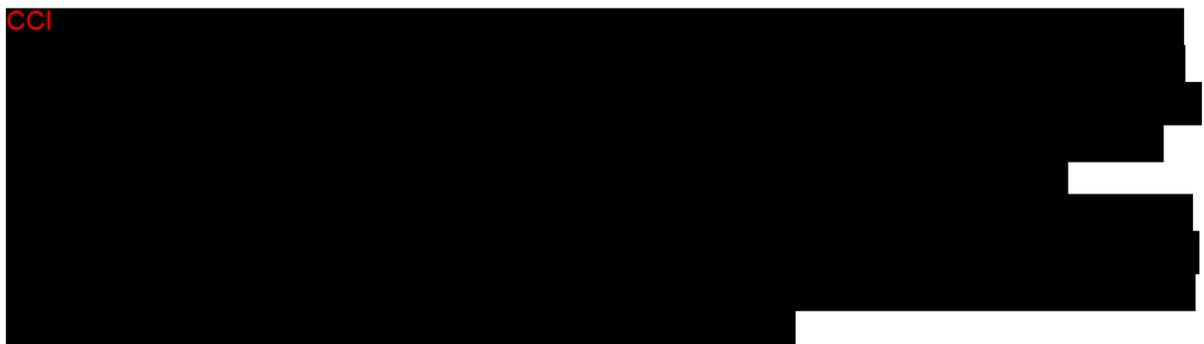
MEDI4736 + tremelimumab versus SoC (PFS, all patients)

CCI
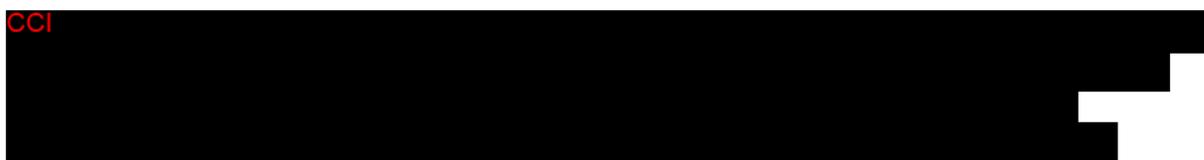
CCI

Table 1 provides a summary of the statistical assumptions.

CCI
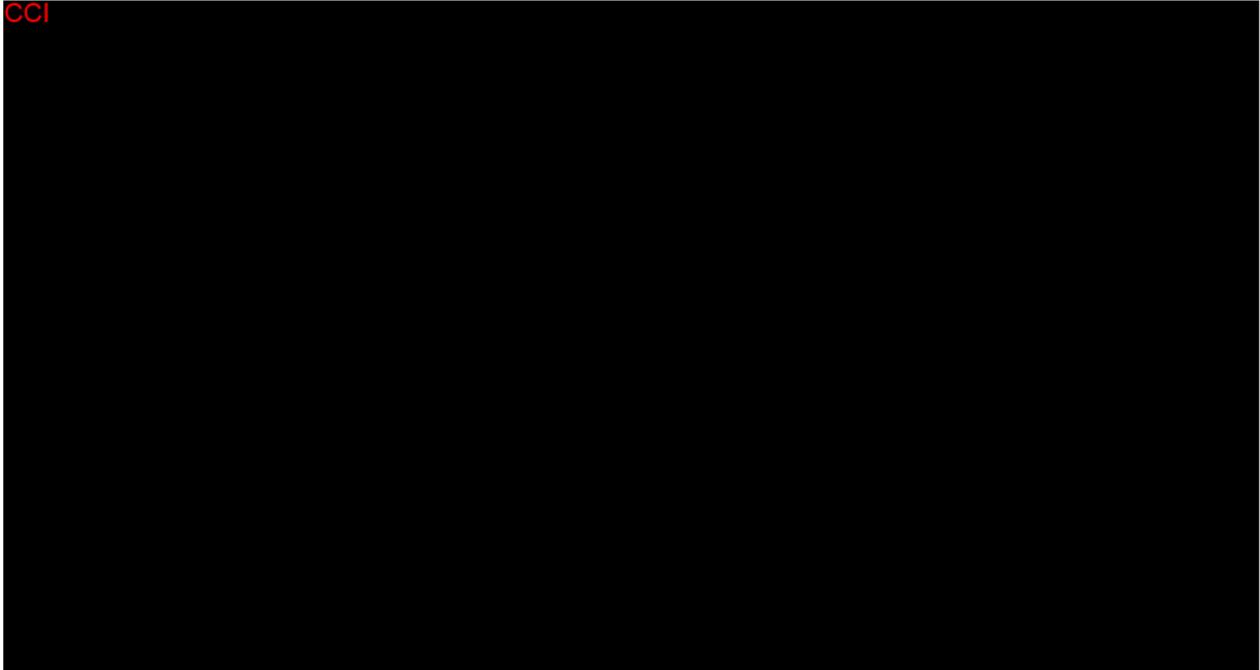
2. ANALYSIS SETS

2.1 Definition of analysis sets

Full analysis set (Intention to treat (ITT))

The full analysis set (FAS) will include all randomized patients. 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.

PD-L1-positive$_{25\%}$ analysis set

The PD-L1-positive$_{25\%}$ analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1 positive$_{25\%}$ as defined by the CCI (ie, $\geq 25\%$ PD-L1–membrane expression in tumoral tissue).

PD-L1- positive$_{1\%}$ analysis set

The PD-L1- positive$_{1\%}$ analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1 positive$_{1\%}$ as defined by the ███████████████████ (ie, $\geq 1\%$ PD-L1–membrane expression in tumoral tissue).

PD-L1-low/negative analysis set

The PD-L1-low/negative analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1 low/negative as defined by the ██████████████████ (ie, $<25\%$ PD-L1–membrane expression in tumoral tissue).

PD-L1-negative analysis set

The PD-L1-negative analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1 negative as defined by the ████████████████ (ie, $<1\%$ PD-L1–membrane expression in tumoral tissue).

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 SAS, 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.

The corresponding PD-L1-positive$_{25\%}$ safety analysis set will include the subset of patients in the SAS whose PD-L1 status is PD-L1 positive$_{25\%}$ as defined by the ███████████████ ██████ (ie, $\geq 25\%$ PD-L1–membrane expression in tumoral tissue).

The corresponding PD-L1-positive$_{1\%}$ safety analysis set will include the subset of patients in the SAS whose PD-L1 status is PD-L1 positive$_{1\%}$ as defined by the ████████████████ ██████ (ie, $\geq 1\%$ PD-L1–membrane expression in tumoral tissue).

Pharmacokinetic analysis set

All patients who received at least 1 dose of investigational product (IP) per the protocol for whom any post-dose data are available and who do not violate or deviate from the protocol in ways that would significantly affect the pharmacokinetic (PK) analyses will be included in the PK Analysis Set. The population will be defined by the Study Physician, Pharmacokineticist, and Statistician prior to any analyses being performed. Definitions of the analysis sets for each outcome variable are provided in Table 2.

Table 2   Summary of outcome variables and analysis populations

Outcome variable	Population
Efficacy data*	
PFS and OS*	Full analysis set (ITT population)
ORR, DoR, APF12, PFS2, PROs, and symptom endpoints*	Full analysis set (ITT population)
Demography*	Full analysis set (ITT population)
PK data	PK analysis Set
Safety Data*	
Exposure*	Safety analysis Set
AEs*	Safety analysis Set
Laboratory measurements*	Safety analysis Set
Vital signs*	Safety analysis Set
ECGs*	Safety analysis Set

Note: all the outcome variables with * will be repeated in the corresponding PD-L1-positive$_{25\%}$ analysis set and PD-L1-positive$_{1\%}$ analysis set.

2.2   Violations and deviations

The important protocol deviations will be listed and summarised by randomised treatment group. Deviation 1, below, will lead to exclusion from the Safety analysis set. Deviation 4, below, will lead to exclusion from the FAS analysis set. None of the other deviations will lead to patients being excluded from the analysis sets described in Section 2.1 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK). A per-protocol analysis excluding patients with significant 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 CCI of patients:

- did not have the intended disease or indication or

- did not receive any randomised therapy.

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

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:

- CCI ████████████████████████████████████

  ▮ ██
    ████████████████████████

  ▮ ██

  ▮ ██

  ▮ ██
    ██
    ██
    ████████████████████████████
    █████████████

  ▮ ██
    ████████████████

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.

Misrandomisations 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 misrandomisation is when a patient is not randomised or treated according to the randomisation schedule. It is envisaged that there will be 2 sub categories 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 randomisation code. However, the actual treatment may still match the randomised treatment. For example, a patient is given randomisation code 0001, which according to the

randomisation schedule is MEDI4736. However, at the randomisation visit they are given treatment pack 0003, but this still contains MEDI4736.

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 misrandomisations 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 CCI 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.

The baseline assessment should be performed no more than 28 days before randomization and ideally as close as possible to the start of the assigned IP. Efficacy for all patients will be assessed by objective tumor assessments every 6 weeks for the first 48 weeks (relative to the date of randomization; Table 2 in the CSP for MEDI4736 + tremelimumab or MEDI4736 monotherapy, Table 3 in the CSP for SoC, and Table 4 in the CSP for patients who have completed/discontinued randomized treatment) then every 8 weeks thereafter, until confirmed objective disease progression per 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.

For patients who discontinue IP (including SoC) due to toxicity in the absence of confirmed objective progression, objective tumor assessments per the scheduled assessments should be continued every 6 weeks for 48 weeks (relative to randomization) and then every 8 weeks until confirmed objective disease progression.

A confirmatory scan is required for all patients following the initial demonstration of PD. The confirmatory scan should occur no earlier than 4 weeks after the initial assessment of PD and preferably at the next scheduled visit in the absence of clinically significant deterioration. Treatment with MEDI4736 + tremelimumab, MEDI4736 monotherapy, or SoC may continue between the initial assessment of progression and confirmation of progression. Progression would be considered confirmed per RECIST 1.1 criteria available in CCI in the CSP using Investigator assessments.

If a patient discontinues treatment (and/or receives a subsequent anticancer therapy) prior to progression, then the patient should still continue to be followed until confirmed objective disease progression.

Categorization of objective tumor 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 tumor burden was at a minimum (ie, smallest sum of diameters previously recorded on study). In the absence of progression, tumor response (CR or PR) and SD will be calculated in comparison to the baseline tumor 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 timepoints indicated in Table 4 in the CSP.

Patients in the MEDI4736 + tremelimumab group who receive retreatment must have a baseline tumor assessment within 28 days of restarting treatment and additional scans every 6 weeks for the first 48 weeks relative to the date of randomization, and then every 8 weeks thereafter until disease progression. All assessments in Table 2 will be followed for patients who receive retreatment.

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 tumor 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).

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

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

Measurable disease (ie 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 3   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 $\geq$ 20% increase in the sum of diameters of TLs and an absolute increase of $\geq$ 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 / embolisation), 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 ≤ CCI 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 ≤ CCI 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)

CCI

CCI

CCI

CCI
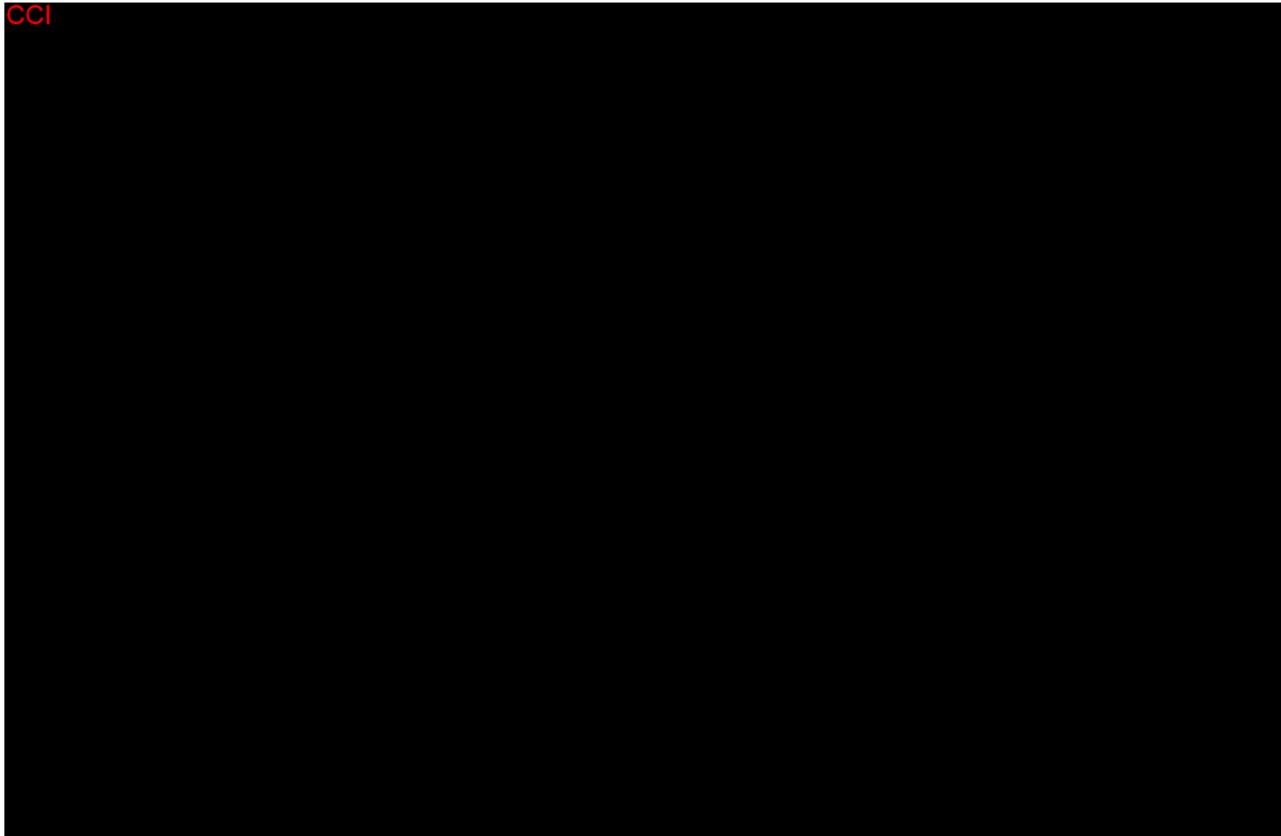
Lesions that split in two

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 4   **NTL Visit Responses**

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
Progressive Disease (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Non-CR/Non-PD	Persistence of one or more NTLs with no evidence of progression.
Not Evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit.
	Note: For 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 5 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 5   Overall visit responses

Target 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

3.1.4   Blinded Independent Central Review (BICR) of RECIST 1.1-based assessments

The BICR will be performed on all radiological scans of all patients for the evaluation of the co-primary PFS endpoint (and all secondary endpoints determined from the tumour assessments). All images will be collected centrally. The imaging scans will be reviewed by 2 independent radiologists using RECIST 1.1 and will be adjudicated, if required. For each patient, the BICR will define the overall visit response data (CR, PR, SD, PD, or NE) and the relevant scan dates for each time point (ie, for visits where response or progression is/is not identified). If a patient has had a tumor assessment that cannot be evaluated, then the patient

will be assigned a visit response of NE (unless there is evidence of progression, in which case the response will be assigned as PD). Endpoints (of PFS) will be derived from the overall visit response date and the scan dates.

The definitions of irCR, irPR, irSD, irPD, and irNED (ie responses according to irRC), as outlined by Wolchok et al 2009, will be outlined in the BICR charter, but a brief description of the methodology is given here. In this project irRC using a RECIST base will be implemented where the target lesions will be measured unidimensionally.

In irRC the presence of new lesions will not automatically trigger a declaration of Progressive Disease, but instead the new lesions will be measured and these measurements will be addedto the sum of diameters of the target lesions. Based on the sum of these measurements and % calculations thereof, the target lesion response assessment will be derived. The overall response assessment (irCR, irPR, irSD, irPD, irNE or irNED) will be obtained at the BICR and confirmation of irPD is required.

CCI

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

3.2 Outcome Variables

The analysis of the co-primary endpoint, PFS, and the analyses of the secondary endpoints, ORR, DoR, and APF12, will be based on BICR tumor assessments according to RECIST 1.1. OS will be evaluated as a co-primary endpoint from all-cause mortality. In addition, time to secondary progression (PFS2) will be defined by local clinical practice.

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 Co-Primary endpoints

PFS and OS are the co-primary endpoints.

3.2.1.1 Progression-free survival

Progression free survival (PFS) (per RECIST 1.1 using BICR 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 randomisation + 1). Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more missed visits, the patient will be censored at the time of the latest RECIST 1.1 assessment. If the patient has no evaluable visits or does not have baseline data, they will be

censored at study day 1 unless they die within 2 visits (2x6 weeks for tumor assessments + 2x7 days for visit window) of baseline.

A second (modified) PFS definition will also be applied to the co-primary PFS comparisons as a sensitivity analysis. This definition will be as described above, but for subjects who receive subsequent anticancer therapy prior to progression (or death), they will be censored at the time of the start date of taking the subsequent anticancer treatment.

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

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 for the adjudicated reviewer selecting PD or of either reviewer where both select PD as a time point response and there is no adjudication for central review (BICR).

- For investigational assessments, the date of progression will be determined based on the earliest of the RECIST 1.1 assessment/scan dates of the component that indicates progression.

- When censoring a patient for PFS, the patient will be censored at the latest of the 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.

Exploratory analyses

CCI
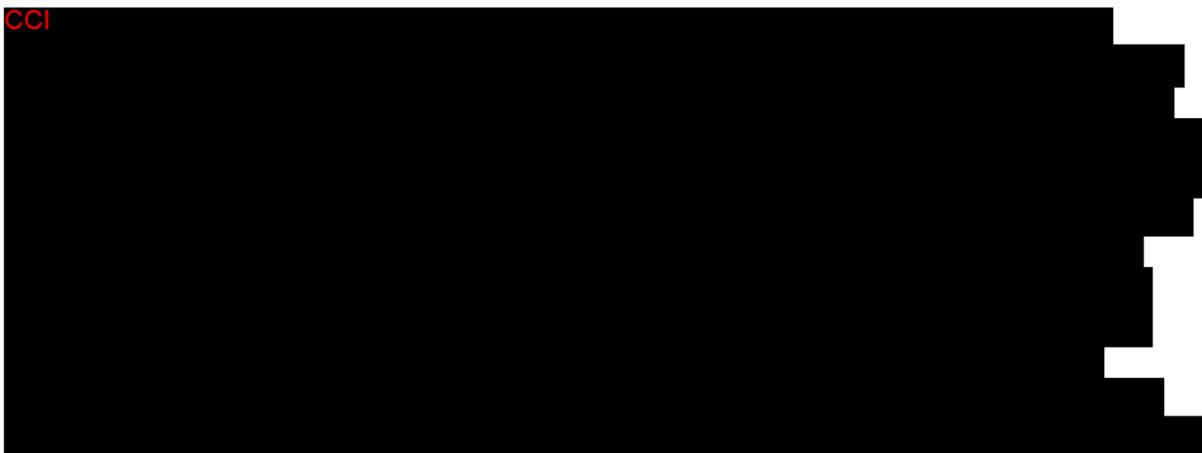
CCI

3.2.1.2   Overall survival

OS is defined as the time from the date of randomization until death due to any cause (ie, date of death or censoring – date of randomisation + 1). Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Note: Survival calls will be made in the week following the date of 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

3.2.2.1   Objective response rate

ORR (per RECIST 1.1 using BICR assessments) is defined as the number (%) of patients with at least 1 visit 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.

ORR will also be obtained using the algorithm described above for the RECIST 1.1 site investigator tumour data. The denominator for ORR will be all randomised patients with measurable disease at baseline per the site investigator (ie, the ITT population).

CCI

3.2.2.2   Duration of response

DoR (per RECIST 1.1 using BICR 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 denominator for DoR will be defined as described for ORR (see Section 3.2.2.1).

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.

3.2.2.3   Time from randomization to second progression (PFS2)

Time from randomisation to second progression (PFS2) is defined as the time from the date of randomisation to the earliest of the progression events (subsequent to that used for the primary variable PFS and excluding any confirmation of progression scans performed for first progression) or death (ie date of PFS2 event or censoring – date of randomisation + 1). The date of the first progression will be programmatically determined from investigator assessed data (See Section 3.2.1 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 nonprogressed) at each assessment will be recorded in the electronic case report form (eCRF).

Second progression status will be reviewed in line with scheduled follow-up (Table 4 CSP) following the progression event used for the co-primary variable PFS (the first progression) and status recorded.

The analysis of PFS2 should include all randomised patients. Patients who have a first PFS event who are alive but with no second event are censored at last known date alive. Patients who died as a first PFS event have their PFS2 event also at date of death. Patients who have a first PFS event and then die subsequently will have their PFS2 event at date of death. Patients alive without any first PFS event will be censored at their last known date alive.

3.2.2.4 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 BICR assessments) at 12 months.

3.2.2.5 Best objective response

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in CCI 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, as determined by BICR.

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

BoR will be determined programmatically based on RECIST 1.1 using all BICR 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 randomisation (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 ≤90 days (ie, 2*(6 weeks ±3 days)) 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 >90 days (ie, 2*(6 weeks ±3 days)) after the date of randomization then BoR will be assigned to the NE category.

Progression events that have been censored due to them being >90 days after the last evaluable assessment will not contribute to the BoR derivation.

3.2.2.6 Change in tumour size

For supportive purposes percentage change from baseline in tumour size will be derived at each scheduled tumour assessment visit (ie, 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 timepoint of interest. Tumour size is defined as the sum of the longest diameters of the target lesions for the BICR 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 starting treatment. 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.

Apply a window around the week X visit: Whenever tumour size data for the week X visit (Note: or visit at which progression was documented if before week X) is available then this should be used in the analysis. A windowing rule will be applied and will follow the protocol allowed visit window; therefore any RECIST scan performed within ± 7 days of the protocol scheduled visit will be used for that visit.

The above derivations will be programmed for the BICR 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), CCI All items/questionnaires will be scored according to published scoring guidelines or the developer's guidelines, if published guidelines are not available. All PRO analyses will be based on the Full Analysis Set (FAS; ITT population) and the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations..

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 symptom items (dyspnea, insomnia, appetite loss, constipation, and diarrhea), and a global measure of health status/QoL. 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/QoL scale in the EORTC QLQ-C30 according to the EORTC QLQ-C30 Scoring Manual. Higher scores on the global health status/QoL and functioning scales indicate better health status/function, but higher scores on symptom scales/items represent greater symptom severity.

Baseline will be defined as the last non-missing assessment prior to first dose for symptoms and summaries.

The change from baseline in HRQoL will be assessed using the EORTC-QLQ-C30 global QoL scale which includes 2 items from the QLQ-C30: "How would you rate your overall

42

health during the past week?" (Item 29) and "How would you rate your overall QoL during the past week?" (Item 30).

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 assessment period response in health-related quality of life**

Score	Change from baseline	Assessment period response
EORTC QLQ-C30 Global health status/QoL 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

For each subscale, if <50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers et al 2001). If at least 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing. The reason for any missing questionnaire will be identified and recorded. If there is evidence that the missing data are systematic, missing values will be handled to ensure that any possible bias is minimized.

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

3.3.1.1 Time to HRQoL/function deterioration

For the following HRQoL items of the EORTC QLQ-C30, time to deterioration will be analyzed:

The Global Health Status/ QoL scale consisting of items 29 and 30 of the EORTC QLQ C30.

- Item 29: How would you rate your overall health during the past week?

- Item 30: How would you rate your overall quality of life during the past week?

Patients are asked to rate their overall health and overall quality of life on a scale from 1 (very poor) to 7 (excellent).

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/QoL from baseline of ≥10) that is confirmed at a subsequent assessment period 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 deterioration. (ie date of HRQoL/function deterioration event or censoring-date of randomisation + 1). 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 assessment periods 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. CCI

The population for analysis of time to HRQoL/function deterioration will include a subset of the ITT population who have baseline scores ≥ 10. In the primary analysis, RECIST 1.1 progression will not be considered as HRQoL/function deterioration and data will not be affected by RECIST 1.1 progression. However a sensitivity analyses will be performed for time to HRQoL deterioration for global health status only where RECIST 1.1 progression is considered as an event of HRQoL deterioration. If a patient has both RECIST 1.1 progression and HRQoL deterioration, the earliest date of the two will be used.

3.3.1.2 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/items) in that symptom from baseline. The denominator will consist of a subset of the ITT population who have a baseline symptom score >10.

3.3.1.3 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 the ITT population who have a baseline HRQoL/function score ≤ 90.

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 LC-13 incorporates symptom scales including:

- Dyspnea (multi-item scale based on 3 questions: were you short of breath when you rested; walked; climbed stairs)

- Cough: 1 item (how much did you cough?)

- Haemoptysis: 1 item (did you cough up blood?)

- Pain: 3 individual items (Have you had pain in your chest; your arm or shoulder; other parts of your body?)

The dyspnea scale is only used if all 3 items have been scored; otherwise the items are treated as single-item measures. The scoring approach for the QLQ-LC-13 is identical in principle to that for the symptom scales/single items of QLQ-C30.

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 from baseline will be categorized as an improvement, no change or deterioration as shown in Table 7.

Table 7   **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

QLQ-LC13 Lung Cancer Module.

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

3.3.2.1   Time to symptom deterioration

For each of the following key symptom scales/items in the QLQ-LC13 and QLQ-C30, time to deterioration will be analyzed:

- Dyspnea (multi-item scale based on 3 questions: were you short of breath when you rested; walked; climbed stairs) from LC13

- Cough: 1 item (how much did you cough?) from LC13

- Pain (3 individual items): a) Have you had pain in your chest; b) your arm or shoulder; c) other parts of your body?) from LC-13

- Hemoptysis: 1 item (did you cough up blood?) from LC-13

- Appetite Loss (Have you lacked appetite) from C-30

- Fatigue (Have you felt weak, Did you need to rest, Were you tired) from C-30

The dyspnea scale is only used if all 3 items have been scored; otherwise the items are treated as single-item measures. The scoring approach for the QLQ-LC13 is identical in principle to that for the symptom scales/single items of the EORTC QLQ-C30.

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 assessment 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

46

symptom deterioration (ie date of symptom deterioration event or censoring – date of randomisation + 1). 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 assessment periods of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by QLQ-LC13 or QLQ-C30) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if symptoms deteriorate after 2 or more missed PRO assessmentsor the patient dies after 2 or more missed PRO assessments, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated. If the patient has no evaluable assessmentsor does not have baseline data they will be censored at 1 day unless they die within 2 assessments of baseline (126 days (i.e. 16 weeks x7 days) plus 2x7 days allowing for a late assessment within the assessment window, for a QLQ-C30 assessment, and 70 days (i.e. 8 weeks x7 days) plus 2x7 days allowing for a late assessment within the assessment window, for a QLQ-LC13 assessment).

Given the scheduled PRO assessment scheme (ie every 4 weeks for first 8 weeks and then every 8 weeks thereafter for QLQ-C30 and every 2 weeks for first 8 weeks and then every 4 weeks thereafter for QLQ-LC13) the definition of 2 missed assessments will change. CCI
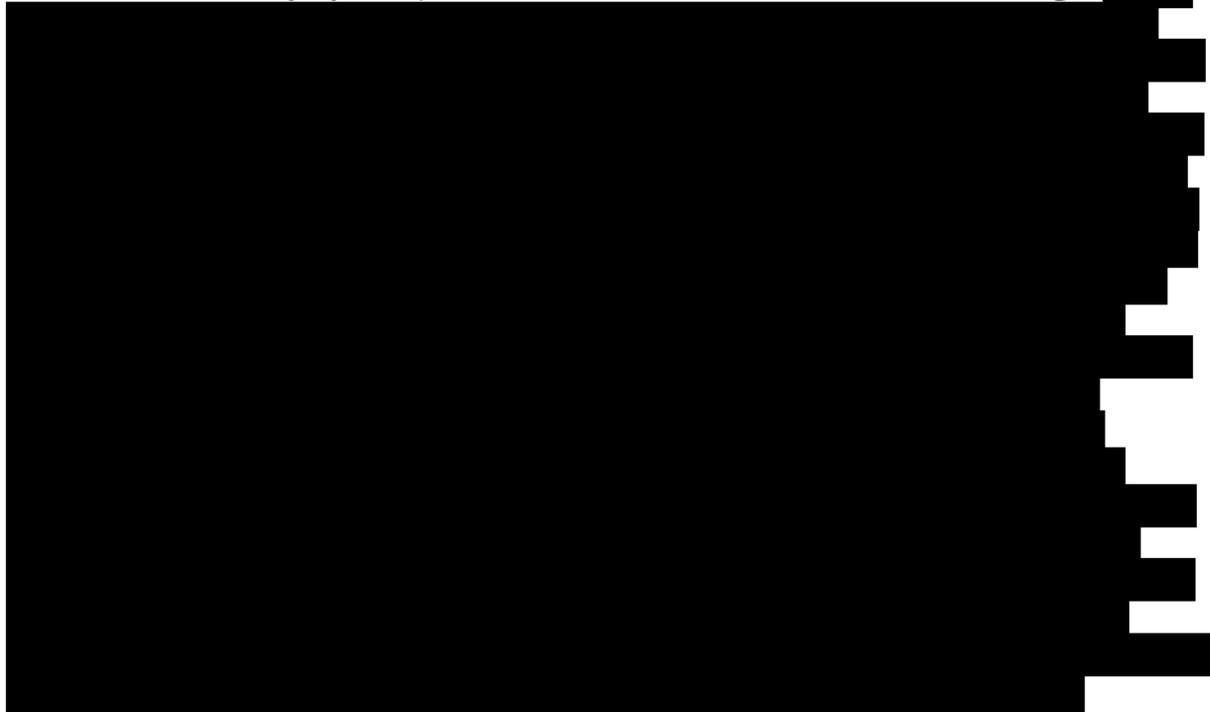
The population for analysis of time to symptom deterioration will include a subset of the ITT population who have baseline scores ≤ 90.

In the primary analysis, RECIST 1.1 progression will not be considered as symptom deterioration and data will not be affected by RECIST 1.1 progression. However a sensitivity analyses will be performed for QLQ-LC13 time to symptom deterioration for each of chest pain, arm/shoulder pain, and other pain where RECIST 1.1 progression is considered as an event of symptom deterioration.

If a patient has both RECIST 1.1 progression and symptom deterioration, the earliest date of the two will be used.

3.3.2.2   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 the ITT population who have a baseline symptom score >10.

CCI

3.3.4   PRO Compliance Rates

Summary measures of overall compliance and compliance over time will be derived for the EORTC-QLQ-C30, 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 HRQoL 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 randomised treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by 100.

Compliance over time will be calculated separately for each assessment period, 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 assessment period, 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 the immunotherapy agents alongside the SOC agents. Safety data from the re-treatment period for the MEDI4736 + tremelimumab group may also be summarised via a small set of headline summaries should there be sufficient number of patients re-treated to warrant this. Any safety summaries representing the re-treatment period will be based upon a subset of the safety analysis set representing patients who have had at least one dose of study treatment in the re-treatment period.

Data from the treatment period on the immunotherapy agents (MEDI4736 or MEDI4736+tremelimumab) will be compared against SoC in the main presentations of safety data and safety data from the retreatment period for the MEDI4736 + tremelimumab group

may also be summarised separately (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 (ie, the last dose of MEDI4736 or MEDI4736+tremelimumab) 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 4.2.10).

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

3.4.1   Adverse events (AEs)

AEs and SAEs will be collected throughout the study, from date of first dose and 90 days after the last dose of immunotherapy agents (ie, the last dose of MEDI4736 or MEDI4736+tremelimumab) and between date of first dose 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 immunotherapy agents (ie, the last dose of MEDI4736 or MEDI4736+tremelimumab) or 30 days after the last dose of the Standard of Care agents. For the MEDI4736+tremelimumab group, in the unlikely event of the components being administered separately then date of first dose/last dose will be considered as the earliest/latest dosing date of either component.

Other significant adverse events

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

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

AEs of special interest

An adverse event of special interest (AESI) is one of scientific and medical interest specific to understanding of the Investigational Product and may require close monitoring and rapid communication by the investigator to the sponsor. An AESI may be serious or non-serious.

The rapid reporting of AESIs allows ongoing surveillance of these events in order to characterize and understand them in association with the use of this investigational product.

AESIs for MEDI4736 include but are not limited to events with a potential inflammatory or immune-mediated mechanism and which may require more frequent monitoring and/or interventions such as steroids, immunosuppressants and/or hormone replacement therapy. These AESIs are being closely monitored in clinical studies with MEDI4736 monotherapy and combination therapy. An immune-related adverse event (irAE) is defined as an adverse event that is associated with drug exposure and is consistent with an immune-mediated mechanism of action and where there is no clear alternate aetiology. Serologic, immunologic, and histologic (biopsy) data, as appropriate, should be used to support an irAE diagnosis. Appropriate efforts should be made to rule out neoplastic, infectious, metabolic, toxin, or other etiologic causes of the irAE.

Some clinical concepts (including some selected individual preferred terms and higher level terms) have been considered "AEs of special interest" (AESI) to the MEDI4736 program. These AESIs have been identified as Pneumonitis, Colitis, Hepatitis, Hypothyroidism, Hyperthyroidism, Hypnophysitis, Adrenal Insufficiency, Dermatitis, Nephritis /Acute Renal Failure, Pancreatitis, Neuropathy, Infusion-related Reactions and Hypersitivity / Anaphylactic Reactions. 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. A further review will take place prior to Database lock (DBL) to ensure any further terms not already included are captured within the categories.

3.4.2   Treatment exposure

Exposure will be defined separately for the initial treatment period and for the re-treatment period for the MEDI4736 + tremelimumab group as follows:.

Total (or intended) exposure of MEDI4736 (combination)

- 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 or start of re-treatment (applies to initial treatment period only).

Total (or intended) exposure of tremelimumab (combination)

- 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 or start of re-treatment (applies to initial treatment period only).

Actual exposure of MEDI4736/tremelimumab

- 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. Actual exposure will not be calculated for SoC.

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

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

- 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.

Dose reductions are not permitted per Section 6.7 of the CSP for the immunotherapy agents (MEDI4736 or MEDI4736+tremelimumab). The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Exposure will also be measured by the number of cycles received. For all five choices of SoC regimen, a cycle corresponds to a period of 21 days, but for each 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), MEDI4736 and tremelimumab:

Duration of dose delays = Sum of (Date of the dose - Date of previous dose − 28 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 separately for the initial treatment period and the re-treatment period for the MEDI4736+tremelimumab group. It will also be derived for the SOC agents. 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 MEDI4736, tremelimumab and all Standard of Care therapy:

- 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. When accounting for the calculation of intended cumulative dose 3 days should be added to the date of last dose to reflect the protocol allowed window for dosing.

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

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)

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 need only have 1 post dose-value recorded.

3.4.5   Time to first subsequent therapy from discontinuation of study treatment

Time to subsequent therapy from date of last dose 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. For derivation of post baseline visit values considering visit window and to handle multiple records present in any visit window, derivation rules as described in Section 3.4.7 below will be used.

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 centrally via a digital read. This digital copy of all ECGs will be held centrally by a central ECG provider, and the data from this review will be stored for analysis if necessary at the end of the study. If it is necessary to analyse this data then QTcF (Fridericia) will be calculated programmatically using the reported ECG values (RR and QT).

$QTcF = QT/RR^{\wedge} (1/3)$ 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.

CCI

- 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 the re-treatment period for the MEDI4736 + tremelimumab group then baseline is similarly defined as the last non-missing measurement prior to the first dose on the re-treatment period. 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 (ie 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 (positive$_{25\%}$, and positive$_{1\%}$, low-negative and negative) is defined according to following criteria:

- Positive$_{25\%}$:- $\geq 25\%$ PD-L1 membrane-expression in tumoral tissue .

- Positive$_{1\%}$:- $\geq 1\%$ PD-L1 membrane-expression in tumoral tissue .

- Low/Negative:- $< 25\%$ PD-L1 membrane-expression in tumoral tissue.

- Negative:- $< 1\%$ PD-L1 membrane-expression in tumoral tissue

3.6   Pharmacokinetic and Immunogenicity variables

Pharmacokinetic analyses will be performed on the full PK analysis set and also on the subset of patients in the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ sub-populations as deemed necessary to support the submission.

3.6.1   Population pharmacokinetics and exposure-response/safety analysis

A population PK model will be developed using a non-linear mixed-effects modeling approach. The impact of physiologically-relevant patient characteristics (covariates) and disease on PK will be evaluated. CCI The results of such an analysis will be reported separately from the main CSR. The PK, pharmacodynamic (PDx), demographic, safety, and efficacy data collected in this study may also be combined with similar data from other studies and explored using population PK and/or PK-PDx methods.

3.6.2   Pharmacokinetic non-compartmental analysis

. The actual sampling times will be used in the PK calculations. PK concentration data and summary statistics will be tabulated. Individual and mean blood concentration-time profiles will be generated. PK parameters will be determined using standard non-compartmental methods. CCI Samples below the lower limit of quantification will be treated as missing in the analyses.

3.6.3   Immunogenicity analysis

Immunogenicity results will be analyzed descriptively by summarizing the number and percentage of patients who develop detectable ADAs against MEDI4736 and tremelimumab. The immunogenicity titer and presence of neutralizing ADAs will be reported for samples confirmed positive for the presence of ADAs. Summaries will be based upon all patients from the safety population. CCI

Additional summaries of immunogenicity will be produced based on patients in the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ as deemed necessary to support the submission.

CCI

- CCI ████████████████

█ ███████████████

█ █████████████████████

██
██
██
██
██████████████████████████████████

██
███████████████████████████████

██
█████████████████

4. ANALYSIS METHODS

4.1 General principles

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

- H0: No difference between MEDI4736 + tremelimumab and SoC

- H1: Difference between MEDI4736 + tremelimumab and SoC

The co-primary endpoints are PFS and OS in patients with PD-L1-positive$_{25\%}$ tumours (with PFS using BICR assessments per RECIST 1.1). The study has been sized to characterize the PFS and OS benefits of MEDI4736 + tremelimumab versus SoC.

The primary PFS analysis and the first interim OS analysis will be performed when both of the following conditions have been met:

- Approximately 231 PFS events have occurred across the MEDI4736+tremelimumab and SoC treatment groups in patients with PD-L1-positive$_{25\%}$ tumours (73% maturity) AND

- CCI ████████████████████████████████ to the

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

- Approximately 225 OS events have occurred across the MEDI4736+tremelimumab and SoC treatment groups in patients with PD-L1-positive$_{25\%}$ tumours (70% maturity) AND

- Approximately 225 OS events have occurred across the MEDI4736 monotherapy and SoC treatment groups in patients with PD-L1-positive$_{25\%}$ tumours (70% maturity)

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. For log transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), 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 9.2 will be used for all analyses.

Baseline will be the last assessment of the variable under consideration prior to the intake of the first dose of IP, except for efficacy variables. For efficacy variables, baseline is defined as the last visit prior to randomization.

All data collected will be listed. Study population, demography data, efficacy and PRO data will be summarized and analyzed based on the FAS and the PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations. PK data will be summarized and analyzed based on the PK Analysis Set. Safety and treatment exposure data will be summarized on the Safety Analysis Set and, for the subsets of the Safety Analysis Set for patients within the positive$_{25\%}$ and positive$_{1\%}$ populations.

All outputs will be summarized by treatment group for all randomized patients (ITT),all randomized patients within the positive$_{25\%}$ and positive$_{1\%}$ populations, and Safety Analysis Set, SAS for patients within the positive$_{25\%}$ and positive$_{1\%}$ populations, correspondingly.

Safety data will be summarised from the treatment period for the immunotherapy agents alongside the SOC agents. Safety data from the re-treatment period for the MEDI4736 + tremelimumab group may also be summarised via a small set of headline summaries should there be sufficient number of patients re-treated to warrant this. Any safety summaries representing the re-treatment period will be based upon a subset of the safety analysis set representing patients who have had at least one dose of study treatment in the re-treatment period.

4.2 Analysis methods

Results of all statistical analysis will be presented using a appropriately sized confidence intervals (CI) and 2-sided p-value, unless otherwise stated.

The following table (Table 8) 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 8   Pre-planned statistical and sensitivity analyses to be conducted

Endpoints analyzed	Notes

Progression-free survival	Stratified log-rank tests for:
	 • Co-Primary analysis using BICR RECIST 1.1 assessments:
	  – MEDI4736 + tremelimumab versus SoC for PD-L1-positive$_{25\%}$ population (stratified only for histology)
	 • Secondary analysis using BICR RECIST 1.1 assessments:
	  – MEDI4736 + tremelimumab versus SoC (ITT population)
	  – MEDI4736 + tremelimumab versus SoC for PD-L1-positive$_{1\%}$ population
	  – MEDI4736 monotherapy versus SoC for PD-L1-positive$_{25\%}$ population (stratified only for histology)
	  – MEDI4736 + tremelimumab versus MEDI4736 monotherapy (ITT population)
	  – MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{25\%}$ population (stratified only for histology)
	  – MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{1\%}$ population
	 • Sensitivity analyses using Investigator assessments (RECIST 1.1)
	 • Sensitivity analysis based on RECIST 1.1 modified for confirmation of progression using BICR assessments
	 • CCI

Endpoints analyzed	Notes
Overall survival	<u>Stratified log-rank test for:</u> • Co-Primary analysis 　– MEDI4736 + tremelimumab versus SoC for PD-L1-positive$_{25\%}$ population (stratified only for histology) 　– MEDI4736 monotherapy versus SoC for PD-L1-positive$_{25\%}$ population (stratified only for histology) • Secondary analysis: 　– MEDI4736 + tremelimumab versus SoC for PD-L1-positive$_{1\%}$ population 　– MEDI4736 + tremelimumab versus SoC (ITT population) 　– MEDI4736 + tremelimumab versus MEDI4736 monotherapy (ITT population) 　– MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{25\%}$ population (stratified only for histology) 　– MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{1\%}$ population • CCI ███████████████████
Objective response rate	<u>Logistic regression for:</u> • Secondary analysis using BICR RECIST 1.1 assessments 　– MEDI4736 + tremelimumab versus SoC for the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations 　– MEDI4736 monotherapy versus SoC for the PD-L1-positive$_{25\%}$ population 　– MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations • CCI ███████████████
Duration of response	<u>Analysis methods as described by Ellis et al 2008 for:</u> • Secondary analysis using BICR assessments (RECIST 1.1) 　– MEDI4736 + tremelimumab versus SoC for the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations 　– MEDI4736 monotherapy versus SoC for the PD-L1-positive$_{25\%}$ population

Endpoints analyzed	Notes

Proportion of patients alive and progression free at 12 months	Hazard ratio using the Kaplan Meier estimates of progression free survival at 12 months (following method described by Klein et al 2007)
	• Secondary analysis:
	   – MEDI4736 + tremelimumab versus SoC for the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations
	   – MEDI4736 monotherapy versus SoC for the PD-L1-positive$_{25\%}$ population

Time from randomization to second progression	Stratified log-rank test
	• Secondary analysis:
	   – MEDI4736 + tremelimumab versus SoC for the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations
	   – MEDI4736 monotherapy versus SoC for the PD-L1-positive$_{25\%}$ population

Time to deterioration (EORTC QLQ-C30 and QLQ-LC13 endpoints)	Stratified log-rank test

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 used across the co-primary endpoints (PFS, OS), analysis populations (ITT, PD-L1-positive$_{25\%}$, and PD-L1-positive$_{1\%}$ populations), and treatment regimens (MEDI4736 + tremelimumab, MEDI4736 monotherapy, and SoC). If the higher level hypothesis in the MTP is rejected for superiority, the following hypothesis will then be tested as shown in Figure 3.

Hypotheses will be tested using a multiple testing procedure with an alpha-exhaustive recycling strategy (Burman et al 2009). With this approach, hypotheses will be tested in a pre-defined order as outlined in Figure 3. According to alpha (test mass) splitting and alpha recycling, the test mass that becomes available after each rejected hypothesis is recycled to secondary hypotheses not yet rejected. Since OS is tested at multiple timepoints (ie, 2 interim analyses and final analysis), the OS tests that for the same comparison/population (ie, 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 next MTP level. This 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), among all key hypotheses. Figure 3 shows the multiple testing framework.

The details on how the alpha will be spent / controlled in all the possible scenarios are outlined below.

1. Test $H_{0,1}$, $H_{0,2}$, and $H_{0,3}$ at level 0.5%, 3% and 1.5% respectively.

(a)   If none of the 3 tests is significant, accept $H_{0,1}$, $H_{0,2}$, and $H_{0,3}$ and stop procedure.

(b)   For the PFS tests in the 1st column, if $H_{0,1}$ is statistically significant at 0.5% level, then reject $H_{0,1}$ (but accept $H_{0,2}$ and $H_{0,3}$) and continue testing $H_{0,4}$ at 0.5% level.

   (i)   If $H_{0,4}$ is not statistically significant at 0.5% level, accept $H_{0,4}$ and stop the procedure.

   (ii)   If $H_{0,4}$ is statistically significant at 0.5% level, then reject $H_{0,4}$ and continue testing $H_{0,6}$ at 0.5% level

   (iii)   If $H_{0,6}$ is not statistically significant at 0.5% level, accept $H_{0,6}$ and stop the procedure.

   (iv)   If $H_{0,6}$ is statistically significant at 0.5% level, then reject $H_{0,6}$ and continue testing $H_{0,8}$ at 0.5% level

   (v)   If $H_{0,8}$ is not statistically significant at 0.5% level, accept $H_{0,8}$ and stop the procedure.

   (vi)   If $H_{0,8}$ is statistically significant at 0.5% level, then reject $H_{0,8}$ and the 0.5% alpha will be recycled to $H_{0,3}$.

(c)   For the OS tests in the 2nd and 3rd columns of the MTP, if $H_{0,2}$ is statistically significant at 3% level, and $H_{0,3}$ is not statistically significant at 1.5% level, then reject $H_{0,2}$ and accept $H_{0,3}$, and continue testing $H_{0,5}$ at 3% level.

   (i)   If $H_{0,5}$ is not statistically significant at 3% level, accept $H_{0,5}$ and stop the procedure.

   (ii)   If $H_{0,5}$ is statistically significant at 3% level, then reject $H_{0,5}$ and continue testing $H_{0,7}$ at 3% level.

   (iii)   If $H_{0,7}$ is not statistically significant at 3% level, accept $H_{0,7}$ and stop the procedure.

   (iv)   If $H_{0,7}$ is statistically significant at 3% level, then reject $H_{0,7}$ and the 3% alpha will be recycled to $H_{0,3}$.

(d)   If $H_{0,3}$ is statistically significant at 1.5% level, and $H_{0,2}$ is not statistically significant at 3% level, then reject $H_{0,3}$ and accept $H_{0,2}$, and continue testing $H_{0,6}$ at 1.5% level.

   (v)   If $H_{0,5}$ is not statistically significant at 1.5% level, accept $H_{0,5}$ and stop the procedure.

(vi)   If $H_{0,5}$ is statistically significant at 1.5% level, then reject $H_{0,5}$ and continue testing $H_{0,7}$ at 1.5% level.

(vii)   If $H_{0,7}$ is not statistically significant at 1.5% level, accept $H_{0,7}$ and stop the procedure.

(viii) If $H_{0,7}$ is statistically significant at 1.5% level, then reject $H_{0,7}$ and the 1.5% alpha will be recycled to $H_{0,2}$.

(e)   If both the two hypotheses $H_{0,2}$ and $H_{0,3}$ are statistically significant at 3% and 1.5% levels, respectively, then reject both $H_{0,2}$ and $H_{0,3}$ and recycle 4.5% alpha from tests of $H_{0,2}$ and $H_{0,3}$ to test of $H_{0,5}$.

(ix)   If $H_{0,5}$ is not statistically significant at 4.5% level, accept $H_{0,5}$ and stop the procedure.

(x)   If $H_{0,5}$ is statistically significant at 4.5% level, reject $H_{0,5}$ and continue testing $H_{0,7}$ at 4.5% level.

(xi)   If $H_{0,7}$ is not statistically significant at 4.5% level, accept $H_{0,7}$ and stop the procedure..

(xii)   If $H_{0,7}$ is statistically significant at 4.5% level, reject $H_{0,7}$

The co-primary endpoint OS is tested at 2 interim and a final timepoint. The OS tests for the same comparison/population (e.g., $H_{0,2}$) will be considered as 1 test family. The alpha level allocated to one OS test family will be controlled at the interim and primary time points by using the Lan 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. The first OS interim analysis for superiority will occur at the primary PFS analysis, when it is expected that approximately 68% of the target death events may occur.

The second OS interim will subsequently be performed at approximately 80% of the target death events, with the primary OS analysis performed when 225 deaths have accumulated in the PD-L1-positive$_{25\%}$ population. If exactly 68% and 80% of the target events in the PD-L1-positive$_{25\%}$ population are available at the time of the first and second interims analyses, respectively (ie, 152/225 and 180/225 deaths have occurred), with an overall 2-sided alpha level 0.015 and 0.03 respectively for the comparisons of MEDI4736 + tremelimumab versus SoC and MEDI4736 monotherapy versus SoC, the 2-sided alpha to be applied for the interim and final analyses would be 0.0023, 0.0049 and 0.0132 for the comparisons of MEDI4736 + tremelimumab versus SoC and 0.0062, 0.0113 and 0.0258 for the comparisons of MEDI4736 monotherapy versus SoC.

If the interim or final analyses indicate superiority in OS for either monotherapy or combination therapy in PD-L1 positive$_{25\%}$, then subsequent analyses of secondary OS

endpoints will be performed in accordance with the hierarchical testing strategy. A separate Lan DeMets (O'Brien Fleming) spending function will be used to determine the alpha levels at the interim and final analyses for testing the PD-L1 positive$_{1\%}$ and the all-comers hypotheses.

The timing of analysis for the secondary endpoint of PD-L1-positive$_{1\%}$ population will be defined by meeting the target number of events for the primary endpoint in the PD-L1-positive$_{25\%}$ population. CCI

CCI

to calculate the sample size and subsequently the target number of events, it is possible that the actual prevalence of PD-L1-positive$_{1\%}$ population may be higher (or lower) than what has been assumed. CCI

CCI

If the interim results do not meet the criterion of stopping for superiority for a given hypothesis, then follow-up will continue until the final target number of OS events for that comparison has been observed, following which the hypothesis will be re-tested. If the hypothesis is then rejected, subsequent testing will continue hierarchically. The above testing procedure will ensure strong control of the family-wise error rate (Glimm et al, 2010).

Figure 3   Multiple testing procedures for controlling the type 1 error rate

$H_{0,1}$: Combo vs. SoC PFS, PDL1 ≥ 25% 0.5% α	$H_{0,2}$: Mono vs. SoC OS, PDL1 ≥ 25% 3% α	$H_{0,3}$: Combo vs. SoC OS, PDL1 ≥ 25% 1.5% α
$H_{0,4}$: Mono vs. SoC PFS, PDL1 ≥ 25% 0.5% α	$H_{0,5}$: Combo vs SoC OS, PDL1 ≥ 1% 3% α	$H_{0,5}$: Combo vs SoC OS, PDL1 ≥ 1% 1.5% α
$H_{0,6}$: Combo vs SoC PFS, PDL1 ≥ 1% 0.5% α	$H_{0,7}$: Combo vs SoC OS, all comers 3% α	$H_{0,7}$: Combo vs SoC OS, all comers 1.5% α
$H_{0,8}$: Combo vs SoC PFS, All Comers 0.5% α		
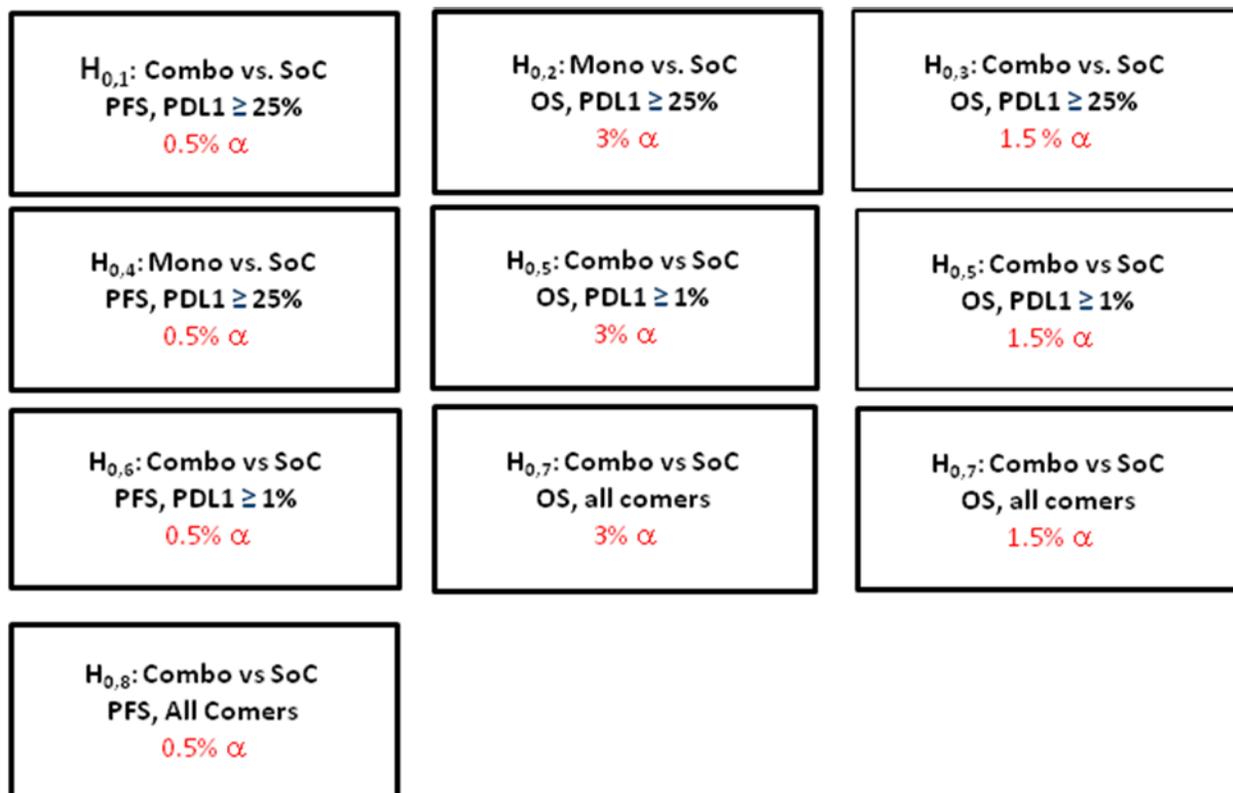
Combo MEDI4736 + tremelimumab combination therapy; Mono MEDI4736 monotherapy; SoC Standard of care.

Note: The alpha values shown in Figure 3 assume all the tests in the MTP higher levels are rejected successfully. The actual alpha for the lower level tests in the MTP depends on how many higher level tests are rejected. The general approach of alpha splitting shown in Figure 3 will be applied to the actual alpha after considering the higher level tests results.

4.2.2   Co-Primary endpoints

4.2.2.1   Progression-free survival

The co-primary PFS analyses will be based on the programmatically derived RECIST 1.1 using the BICR tumor assessments. The co-primary analysis will be performed in the PD-L1-positive$_{25\%}$ population using a stratified log-rank test adjusting for histology (squamous versus non-squamous) only. The effect of MEDI4736 + tremelimumab versus SoC treatment will be estimated by the HR together with its corresponding 99.5% CI and p-value.

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 HR and its CI can be estimated from the Cox proportional hazards model (Cox 1972):

Secondary analyses will be performed using the same methodology as for the primary analyses described above.

Kaplan-Meier plots of PFS will be presented by treatment group, and by treatment group and PD-L1 tumor status subgroup, 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.

CCI

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.

Additional supportive summaries/graphs

In addition, the number of patients prematurely censored will be summarised by treatment group together with baseline prognostic factors of the prematurely censored patients. CCI
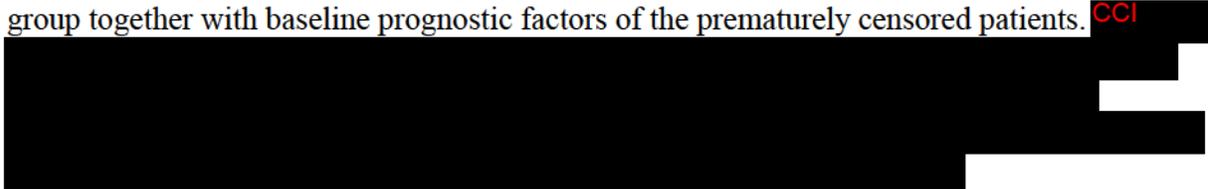
Additionally, summary statistics will be given for the number of days from censoring to data cut-off for all censored patients.

A summary of the duration of follow-up will be summarised using median time from randomisation 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 evaluable 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 randomised patients. In addition, a summary of new lesions (i.e., sites of new lesions) will be produced.

Sensitivity Analyses

The following sensitivity analyses will only be performed for the primary treatment comparison.

- CCI

Disagreements between investigator and central reviews of RECIST 1.1 progression will be presented for each treatment group. The summary will include the early discrepancy rate

which is the frequency of central review declared progressions before the investigator review as a proportion of all central review progressions and the late discrepancy rate which is the frequency of central review declared progressions after the investigator review as a proportion of all discrepancies.

CCI

Secondary Analysis

Secondary analyses of PFS based on the programmatically derived RECIST 1.1 using the BICR tumor assessments will be performed for the following treatment comparisons (in accordance with the multiple testing procedure):

- MEDI4736 + tremelimumab versus SoC (ITT population)
- MEDI4736 + tremelimumab versus SoC for PD-L1-positive$_{1\%}$ population
- MEDI4736 monotherapy versus SoC for PD-L1-positive$_{25\%}$ population (stratified only for histology)
- MEDI4736 + tremelimumab versus MEDI4736 monotherapy (ITT population)
- MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{25\%}$ population (stratified only for histology)
- MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{1\%}$ population

These analyses will be performed using a stratified log-rank test. The effect of treatment will be estimated by the HR together with its corresponding 95% CI and p-value, The HR and CI will be estimated using the same approach as specified above for the primary analysis of PFS.

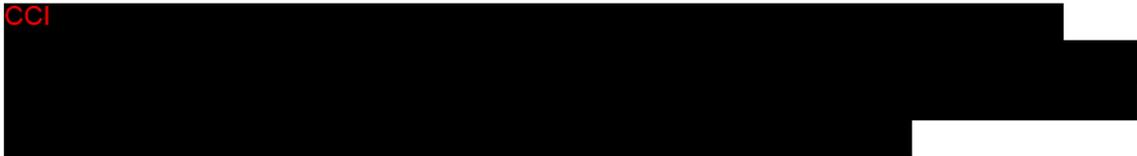## Exploratory Analyses

CCI

Subgroup Analyses

CCI

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

CCI

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 PFS will not be formally analysed. In this case, only descriptive summaries will be provided.

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

Effect of covariates on HR estimate

Cox proportional hazards modeling will be employed to assess the effect of covariates on the HR estimate. A 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.

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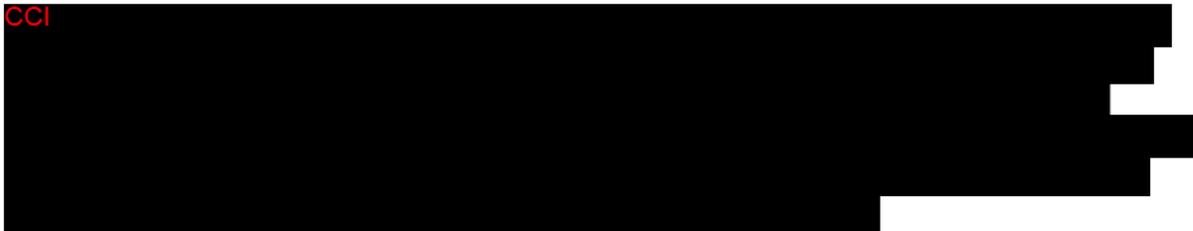
The model will include the effect regardless of whether the inclusion of effect significantly improves the fit of the model providing there is enough data to make them meaningful.

Consistency of treatment effect between subgroups

CCI

4.2.2.2   Overall survival

OS in the PD-L1-positive$_{25\%}$ population will be analyzed using a stratified log-rank tests, using the same methodology as described for the primary PFS endpoint. The effects of MEDI4736 + tremelimumab versus SoC and MEDI4736 monotherapy versus SoC will be

estimated by the HR together with its corresponding CI (98.5% for MEDI4736 + tremelimumab versus SoC and 97% for and MEDI4736 monotherapy versus SoC, adjusted for the interim analyses) and p-value. 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.

CCI

CCI

The alpha will be split between the final and the 2 interim analyses using the hierarchical testing strategy as already described in Section 4.2.1. The boundaries (ie, 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 5.1)

CCI

Secondary Analysis

Secondary analyses of OS will be performed for the following treatment comparisons (in accordance with the multiple testing procedure):

- MEDI4736 + tremelimumab versus SoC for PD-L1-positive$_{1\%}$ population
- MEDI4736 + tremelimumab versus SoC (ITT population)
- MEDI4736 + tremelimumab versus MEDI4736 monotherapy (ITT population)
- MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{25\%}$ population (stratified only for histology)
- MEDI4736 + tremelimumab versus MEDI4736 monotherapy for PD-L1-positive$_{25\%}$ population (stratified only for histology)

Subgroup Analyses

CCI

Effect of covariates on HR estimate

CCI

CCI

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

CCI

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

CCI

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

CCI

Subsequent therapies received after discontinuation of treatment will be summarised and listed by treatment group. Patients who subsequently received an immunotherapy agent or entered an immunotherapy trial will be summarised and listed by treatment group according to line of subsequent therapy, i.e. immediately after immunotherapy or as a later line.

4.2.3 Objective response rate

The ORR will be based on the programmatically derived RECIST 1.1 using the BICR tumor data. The ORR will be compared between MEDI4736 + tremelimumab versus SoC using logistic regression models adjusting for the same factors as the primary endpoint (PD-L1 tumor expression and histology). The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour MEDI4736+tremelimumab) together with

its associated profile likelihood CI CCI

██

██████████████████████████████ This analysis will be performed in the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations. The analysis of the PD-L1-positive$_{25\%}$ population will be performed using a logistic regression model adjusting for only histology.

Additional analyses of ORR comparing MEDI4736 monotherapy versus SoC in the PD-L1-positive$_{25\%}$ population and comparing MEDI4736 + tremelimumab versus MEDI4736 monotherapy versus SoC in the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations will be performed.

This analysis of ORR will be repeated using the results of the programmatically derived ORR using the site investigator tumour data based upon RECIST 1.1 as a sensitivity analysis to confirm the results of the primary analysis derived from the CRFs.

ORR by irRECIST 1.1 criteria using BICR assessments will also be reported in the ITT population.

If there are not enough responses for a meaningful analysis using logistic regression then a Fisher's exact test CCI ████████████ will be presented.

CCI

██

██

Summaries will be produced that present the number and percentage of patients with a tumor response (CR/PR). Overall visit response data will be listed for all patients (ie, 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 analyses are planned for BoR.

Subgroup Analyses

Subgroup analyses will be conducted comparing ORR between MEDI4736 + tremelimumab versus SoC and between MEDI4736 monotherapy versus SoC in the same subgroups as specified for the PFS subgroup analyses.

CCI

██

██

If there are too few responses available for a meaningful analysis of a particular subgroup, the relationship between that subgroup and ORR will not be formally analysed. In this case, only descriptive summaries will be provided.

4.2.4 Duration of response

In order to analyze the DoR between MEDI4736 + tremelimumab and SoC, the expected duration of response (EDoR) will be derived for each treatment group (Ellis et al 2008) using the BICR tumor data. The EDoR is the product of the proportion of patients responding to treatment and the mean DoR in responding patients and provides an estimate based on all randomized patients. CCI

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

This analysis will be performed on the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations.

Additional analyses of DoR comparing MEDI4736 monotherapy versus SoC in the PD-L1-positive$_{25\%}$ population will be performed.

4.2.5 Proportion of patients alive and progression free at 12 months

The APF12 (where 12 months equates to study day 366) will be summarized (using the Kaplan-Meier curve) and presented by treatment group. APF12 will be compared between MEDI4736 + tremelimumab and SoC by using the Kaplan-Meier estimator of PFS at 12 CCI

This analysis will be performed in the PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations.

Additional analysis of APF12 comparing MEDI4736 monotherapy versus SoC in the PD-L1-positive$_{25\%}$ population will be performed.

4.2.6 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. PFS2 in the ITT population will be analyzed using a stratified log-rank test, using the same methodology as described for the primary PFS endpoint. The effect of MEDI4736 + tremelimumab versus SoC will be estimated by the HR together with its corresponding 95% CI and p-value. 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 PD-L1-positive$_{25\%}$, PD-L1-positive$_{1\%}$ and ITT populations.

Additional analysis of PFS2 comparing MEDI4736 monotherapy versus SoC in the PD-L1-positive$_{25\%}$ population will be performed.

CCI

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

4.2.7 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 timepoint 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 timepoints may be produced if it is felt that these are warranted to provide greater clarity.

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

4.2.8 Patient reported outcomes

Health related quality of life 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, dyspnea, pain (in the chest, pain in other parts of the body) as well as fatigue and appetite loss. The assessments of cough,,dyspnea (breathlessness) as well as pain as assessed by the EORTC QLQ LC13 and fatigue and appetite loss from EORTC QLQ C30 will be used as secondary efficacy endpoints.

For these secondary efficacy endpoints, the overall type I error (5% 2 sided) will be controlled across the five primary PRO measures of cough, dyspnoea and pain as assessed by the EORTC QLQ-LC13 and fatigue and appetite loss as assessed by the EORTC QLQ-C30 using the Bonferroni-Holm procedure (Holm 1979).

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

4.2.8.1 EORTC QLQ-C30

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

The effect of MEDI4736 + tremelimumab versus SoC will be estimated by the HR together with its corresponding CI and p-value. 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 improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/QoL will be produced.

Summaries of original and change from baseline values of each symptom scale/item, the global health status/QoL score and each functional domain will be reported by assessment period for each treatment group. Graphical presentations may also be produced as appropriate. Summaries of the number and percentage of patients in each response category at each assessment 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.

4.2.8.2 EORTC QLQ-LC13

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

The effect of MEDI4736 + tremelimumab versus SoC will be estimated by the HR together with its corresponding CI and p-value. 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.

Summaries of original and change from baseline values of each symptom (dyspnea, cough, hemoptysis, chest pain, arm/shoulder pain, other pain) and each treatment-related side effect (sore mouth, dysphagia, peripheral neuropathy and alopecia) will be reported by assessment periodfor 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 assessmentfor each ordinal symptom 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.

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

CCI

Generic assessment time	Study Day

	Patient X	Patient Y
Baseline	Baseline	Baseline
1	29	28
2	57	50 (discontinuation)
3	85	85
4	113	113
5	130 (discontinuation)	141
6	169	169

CCI

These analyses will be performed in the FAS, and PD-L1-positive$_{25\%}$ and PD-L1-positive$_{1\%}$ populations as deemed appropriate.

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

CCI

CCI

CCI

CCI

CCI

4.2.10   Safety data

Safety and tolerability data will be presented by treatment group using the full safety population and the subset of patients in the PD-L1-positive$_{25\%}$ population and PD-L1-positive$_{1\%}$ population. 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. Any safety summaries examining re-treatment with MEDI4736+tremelimumab will be produced separately, if required.

Other safety data will be assessed in terms of physical examination, clinical chemistry, hematology, vital signs, and ECGs. Exposure to MEDI4736 + tremelimumab, MEDI4736 monotherapy, and SoC will be summarized. Time on study, MEDI4736 + tremelimumab, MEDI4736 monotherapy, and SoC, dose delays/interruptions and dose reductions will also be summarized. At the end of the study, appropriate summaries of all safety data will be produced.

The following sections describe the planned safety summaries. However, additional safety tables may be required to aid interpretation of the safety data. For example, if an imbalance is

seen in AEs or laboratory abnormalities that could be due to the differential follow-up periods (showing up more of the background/disease related AEs/abnormalities), additional summaries may be produced using a 30 day follow up period for both treatment arms to further explore / explain.

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. The majority of the AE summaries, unless stated otherwise, will be based on TEAEs. Any AE occurring before study treatment (i.e. before the administration of the first dose on Study Day 1) will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'. However, any AE occurring before the administration of the first dose on Study Day 1 that 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 (ie, the last dose of MEDI4736 or MEDI4736+tremelimumab) / 30 days following discontinuation of the Standard of Care agent or until the initiation of the first subsequent anti-cancer therapy 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 are likely to be attributable to subsequent therapy.

However, to assess the longer term toxicity profile, all of the AE summaries will 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 (ie 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 (ie 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 immunotherapy agents / 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 causally related to study medication (as determined by the reporting investigator)

- AEs with CTCAE grade 3 or 4

- AEs with CTCAE grade 3 or 4 , causally related to study medication (as determined by the reporting investigator)

- AEs with outcome of death

- AEs with outcome of death causally related to study medication (as determined by the reporting investigator)

- AEs by outcome

- All SAEs

- All SAEs causally related to study medication (as determined by the reporting investigator)

- SAEs leading to discontinuation of study medication

- AEs leading to discontinuation of study medication

- AEs leading to discontinuation of study medication, causally related to study medication (as determined by the reporting investigator)

- AEs leading to hospitalization

- AEs leading to dose delay of study medication

- Other significant AEs

- Other significant AEs causally related to study medication (as determined by the reporting investigator)

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

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

An overall summary of the number and percentage of patients in each category will be presented, as will an overall summary of the number of episodes in each category. In addition, a truncated AE table of most common AEs and another table showing most common AEs with CTCAE grade 3 or 4, showing all events that occur in at least 5% of patients overall will be summarised by preferred term, by decreasing frequency. This cut-off may be modified after review of the data. When applying a cut-off (ie, 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. The denominator is calculated as the total over each patient of days from first dose to the earlier of the date of onset of the event or the last day of study medication.

Summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, system organ class, preferred term and treatment group.

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

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 (regardless of the date of death)

- Death related to disease under investigation ONLY, as determined by investigator (regardless of the date of death)

- TEAE with outcome of death ONLY and onset date prior to initiation of subsequent anti-cancer therapy

- AE with outcome of death ONLY and onset date falling after 90 days following the date of last dose of immunotherapy /30 days following the date of last dose of SoC or initiation of subsequent anti-cancer therapy (whichever is earlier)

- Death related to disease under investigation, as determined by the investigator, and with TEAE with outcome of death and onset date prior to initiation of subsequent anti-cancer therapy

- Death related to disease under investigation, as determined by the investigator, and with AE with outcome of death and onset date falling after 90 days following the date of last dose of immunotherapy /30 days following the date of last dose of SoC or initiation of subsequent anti-cancer therapy (whichever is earlier)

- Death occurred 90 days after the date of last dose of immunotherapy/30 days following the date of last dose of SoC or after initiation of subsequent anti-cancer therapy (whichever is earlier), and unrelated to AE or disease under investigation

- Patients with unknown reason for death.

- Other deaths

These summaries will be produced twice; firstly accounting for subsequent therapy and, secondly, without taking subsequent therapy into consideration

Adverse events of special interest

Preferred terms used to identify adverse events of special 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 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:

- At least adverse event of special interest presented by outcome

- At least one adverse event of special interest causally related to study medication

- At least one adverse event of special interest leading to discontinuation of study medication

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

Additionally, there will be several summaries of AESIs requiring concomitant treatment, and particularly the relationship of AESIs to the use of immunosuppressive agents (ie, depicting which AESI triggered immunosuppressive use) and, separately, to the use of immunosuppressive agents at high doses.

Summary of long term tolerability

84

To assess long term tolerability, provided that there are a sufficient number of patients with events to warrant it, prevalence plots, life table plots and cumulative incidence plots will be presented for each of the AESI grouped terms and any other events considered important after review of the safety data, provided there are ≥ 10 events.

A prevalence plot provides information on the extent to which the events may be an ongoing burden to patients. The prevalence at time t after first dose of study treatment is calculated as the number of patients experiencing the event divided by the number of patients receiving study treatment or in safety follow-up at time t; generally, t is categorised by each day after dosing. The prevalence will be plotted over time and presented for each treatment group separately. Multiple occurrences of the same event are considered for each patient but a patient is only counted in the numerator whilst they are experiencing one of the occurrences of the event. These plots will only be produced for AESIs that have ≥10 events.

For each AE, median time to first onset of the AE from the date of first dose will be presented in patients in the safety analysis set by treatment group. Patients who did not experience the AE will be censored at the end of their safety follow-up. Summary tables of time to first onset for each AE will also be produced (e.g. 1-28 days, 29-56 days, 57-84 days, 85-112 days, >112 days). Median duration of the AE will be presented in patients who experienced each AE.

A life table plot can be used to describe the time to onset of the event and specifically when patients are at most risk of first experiencing the event. The hazard, or in other words, the probability of having an AE in a specified time period (e.g. 0-1 months, 1-3 months, 3-6 months, etc.) given that the patient reaches that time period without having an event is plotted for each time period. These plots will only be produced for AESIs that have ≥10 events.

A cumulative incidence plot is a plot of the raw cumulative incidence and cumulative incidence function over time with the treatment groups presented on separate plots. The raw cumulative incidence is the actual probability that a patient will have experienced their first occurrence of the event by a given time point. The cumulative incidence function estimates the cumulative incidence if the data cut-off had not been imposed and all patients had completed safety follow-up (Pintilie 2006). These plots will only be produced for AESIs that have ≥10 events.

Laboratory assessments

Data obtained up until the 90 days following discontinuation of immunotherapy agents (ie, the last dose of MEDI4736 or MEDI4736+tremelimumab) or 30 days following discontinuation of the Standard of Care agent or until the initiation of the first subsequent therapy 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 immunotherapy agents 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 will also be produced containing data collected up until 90 days following discontinuation of the immunotherapy agents or up until 30 days following discontinuation of the Standard of Care agent (ie, without taking subsequent therapy into account).

A small selection of summaries of laboratory data may also be produced containing data from initiation of the first subsequent therapy following discontinuation of study treatment until 90 days following discontinuation of immunotherapy agents or until 30 days following discontinuation of the Standard of Care agent (ie summarising the laboratory data collected on patients taking subsequent therapy during the safety collection follow-up window post discontinuation of study treatment). These outputs will only be produced if the number of laboratory toxicities observed warrant the inclusion of such outputs for interpretational purposes. Any data post 90 days last dose for immunotherapy agents or post 30 days 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 (ass 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.

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, Leukocytes, Lymphocytes, absolute count, Neutrophils, absolute count, Platelets

- Clinical chemistry: ALT, AST, ALP, Total bilirubin, Albumin, Magnesium – hypo and – hyper, Sodium – hypo and – hyper, Potassium – hypo and – hyper, Corrected calcium – hypo and – hyper, Glucose – hypo and – hyper, Creatinine

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

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

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 (ie ≥ 3x ULN), and elevated Total bilirubin (ie ≥ 2x ULN) (at any time) will be plotted. Individual patient data where ALT or AST (ie ≥ 3x ULN) plus Total bilirubin (ie ≥ 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, were evaluated at 30 days after last dose, 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, T3 and T4, as deemed necessary.

ECGs

ECG data obtained up until the safety follow-up will be included in the summary tables. Overall evaluation of ECG is collected at each visit in terms of normal or abnormal, and the relevance of the abnormality is termed as "clinically significant" or "not clinically significant". A shift table of baseline evaluation to worst evaluation will be produced.

Vital signs

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

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

Vital signs (systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate, temperature, respiratory rate and weight) will be summarised over time in terms of absolute values and 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. This summary is supportive of the Adverse Event and Laboratory data outputs.

Physical examination

All individual physical examination data will be listed only.

Other Safety Data

Data from positive pregnancy tests will be listed only.

4.2.11   WHO performance status

All WHO performance status will be summarised over time for the ITT population.

4.2.12   PK data (MEDI4736 monotherapy and MEDI4736+tremelimumab groups only)

CCI

88

4.2.13 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-MEDI4736 and anti-tremelimumab antibodies based on the safety population. The immunogenicity titre and neutralizing ADA data will be listed for samples confirmed positive for the presence of anti-MEDI4736 antibodies and/or anti-tremelimumab antibodies.

CCI

4.2.16 Demographic and baseline characteristics data

The following will be summarised for all patients in the FAS, PD-L1-positive$_{25\%}$ analysis set and PD-L1-positive$_{1\%}$ analysis set by treatment group:-

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

- Important protocol deviations

- Inclusion in analysis populations

- Demographics (age, age group[<50, $>=50-<65$, $\geq 65 - <75$ years and ≥ 75 years], sex, race and ethnicity)

- Patient characteristics at baseline (height, weight, weight group, body mass index (BMI) and body mass index group)

- Patient recruitment by country and centre

- Previous disease-related treatment modalities

- Number of regimens of previous chemotherapy at baseline

- Previous lung cancer therapy

- Disease characteristics at baseline / diagnosis (WHO performance status, primary tumour location, histology type, tumour grade and overall disease classification, best response to previous therapy)

- Extent of disease at baseline

- TNM classification

- Disease related medical history (past and current)

- Relevant surgical history

- Physical examination at baseline

- Time from most recent disease progression to start of study treatment

- 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 AZ drug dictionary (AZDD) will be used for concomitant medication coding.

Patient disposition data will also be summarised at the time of OS analysis.

4.2.17 Treatment exposure

The following summaries related to study treatment will be produced for the safety analysis set, PD-L1-positive$_{25\%}$ safety analysis set and PD-L1-positive$_{1\%}$ safety analysis set by randomised treatment group:

- Total exposure of each treatment group.

- Actual exposure of each treatment group.

- Total number of cycles received.

- Reasons for dose delays and infusion interruptions of MEDI4736 and tremelimumab 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 of MEDI4736. In addition, delays due to AEs and due to reasons other than AEs will be summarized separately.

- Number of infusions received.

- RDI (relative dose intensity) of MEDI4736, tremelimumab and Standard of Care agents.

- Exposure over time will be plotted.

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

4.2.18 Subsequent Therapy

Subsequent therapies received after discontinuation of study treatment will have summaries produced by treatment group, together with number of regimens received.
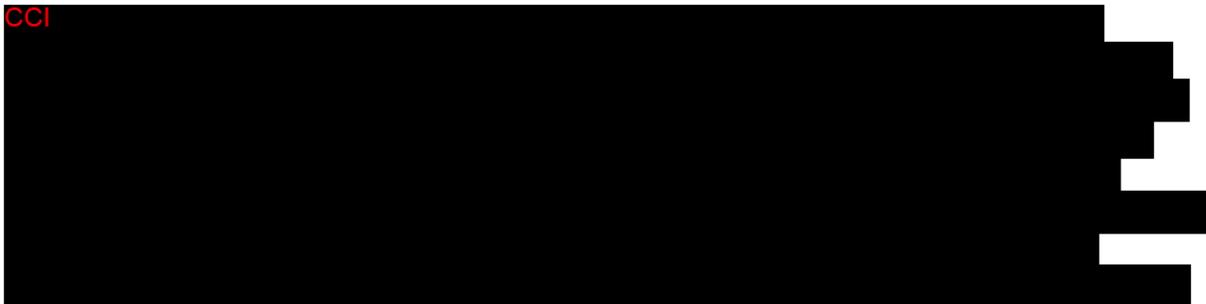
5. INTERIM ANALYSES

5.1 Analysis Methods

Interim safety monitoring will be conducted by an IDMC. Interim analyses will be performed for efficacy as described below:

Two OS interim analyses will be performed for superiority; the first one at the time of the primary PFS analysis and the second one when approximately 80% of the final number of deaths has been reached. These analyses 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 2 interim analyses for superiority (Lan and DeMets 1983).

CCI

CCI

███
███
████████████████████████████████

If the interim analyses indicate superiority in the PD-L1-positive$_{25\%}$ population, then subsequent analyses of the further secondary endpoints will be performed in accordance with the hierarchical multiple testing strategy.

CCI

███
███
███
███
██

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 analyses of OS. The committee will meet approximately 6 months after the study has started or after the first 30 patients have been randomized, whichever occurs first, to review the safety data from the study. The IDMC will meet at least every 6 months thereafter. 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

None from protocol version 7.

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.

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, randomised, 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

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

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

None.