AstraZeneca

Statistical Analysis Plan

Study Code	D4191C00004
Edition Number	3
Date	24 January 2018

A Phase III, Open-label, Randomised, Multi-centre, International Study of MEDI4736, Given as Monotherapy or in Combination with Tremelimumab, Determined by PD-L1 Expression, Versus Standard of Care in Patients with Locally Advanced or Metastatic Non-Small Cell Lung Cancer (Stage IIIB-IV) who Have Received at Least Two Prior Systemic Treatment Regimens Including One Platinum-based Chemotherapy Regimen and Do Not Have Known EGFR TK Activating Mutations or ALK Rearrangements (ARCTIC)

A Phase III, Open-label, Randomised, Multi-centre, International Study of MEDI4736, Given as Monotherapy or in Combination with Tremelimumab, Determined by PD-L1 Expression, Versus Standard of Care in Patients with Locally Advanced or Metastatic Non-Small Cell Lung Cancer (Stage IIIB-IV) who Have Received at Least Two Prior Systemic Treatment Regimens Including One Platinum-based Chemotherapy Regimen and Do Not Have Known EGFR TK Activating Mutations or ALK Rearrangements (ARCTIC)

Study Statistician

PPD

Jan 29, 2018
Date

A Phase III, Open-label, Randomised, Multi-centre, International Study of MEDI4736, Given as Monotherapy or in Combination with Tremelimumab, Determined by PD-L1 Expression, Versus Standard of Care in Patients with Locally Advanced or Metastatic Non-Small Cell Lung Cancer (Stage IIIB-IV) who Have Received at Least Two Prior Systemic Treatment Regimens Including One Platinum-based Chemotherapy Regimen and Do Not Have Known EGFR TK Activating Mutations or ALK Rearrangements (ARCTIC)

Global Product Statistician PPD

30 Jan 2018
Date

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

The following abbreviations and special terms are used in this study Statistical Analysis Plan.

Abbreviation or special term	Explanation
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse events of special interest
ALT	Alanine aminotransferase
APF6	Proportion of patients alive and progression free at 6 months from randomisation
APF12	Proportion of patients alive and progression free at 12 months from randomisation
AST	Aspartate aminotransferase
AZDD	AZ drug dictionary
Baseline	Refers to the most recent assessment of any variable prior to dosing with study treatment/randomisation (as appropriate)
BDRM	Blinded data review meeting
BoR	Best objective response
CI	Confidence interval
CR	Complete response
CRF / eCRF	Case Report Form (electronic)
CRO	Contract Research Organisation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed tomography
CTC / CTCAE	Common Terminology Criteria for Adverse Event
DAE	Discontinuation of investigational product due to adverse event
DBL	Database lock
DBP	Diastolic blood pressure
DCO	Data cut-off
DoR	Duration of response
ECG	Electrocardiogram
EGFR	Epidermal growth factor receptor
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire
EQ-5D	EuroQoL 5-dimension utility index
EQ-5D-5L	EuroQoL 5-dimension, 5-level health state utility index

Abbreviation or special term	Explanation
FAS	Full analysis set
HR	Hazard ratio
HRQoL	Health related quality of life
ICU	Intensive care unit
IDMC	Independent data monitoring committee
ITT	Intention to treat
iv	Intraveneous
IVRS	Interactive voice response system
LC13	Lung Cancer Module; 13-item self-administered questionnaire from the EORTC for lung cancer
LD	Lesion diameter
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milli-gram
MRI	Magnetic resonance imaging
NA	Not applicable
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
OS12	Proportion of patients alive at 12 months from randomisation
PD	Progressive disease
PD-L1	Programmed death ligand 1
PDx	Pharmacodynamic(s)
PFS	Progression free survival
PFS2	Time from randomisation to second progression or death
PGx	Pharmacogenetic(s)
PK	Pharmacokinetic(s)
PR	Partial response
PRO	Patient reported outcomes
Q12W	Every 12 weeks
Q2W	Every 2 weeks
Q4W	Every 4 weeks

Abbreviation or special term	Explanation
QTcF	QT interval (corrected for heart rate using Fridericia's correction)
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria In Solid Tumours
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SSA	Sub-Study A
SSB	Sub-Study B
SBP	Systolic blood pressure
TEAE	Treatment emergent adverse event
TL	Target lesions
ULN	Upper limit of normal
WHO	World Health Organisation

AMENDMENT HISTORY

Date	Brief description of change
5 Dec 2016	In line with the Clinical Study Protocol (CSP) Amendments: • Removal of BICR assessments including irRECIST • Clarification the contribution of components analysis and timing. Also, this will only be performed once so no multiplicity adjustment • Number of recruited and randomised subject updated. • Since Sub-study A is no-longer powered for hypothesis testing, analyses for Sub-study A are now descriptive. • The interim OS analysis for sub-study A has been removed • The interim OS analysis for sub-study B will be conducted at the same time as the final analysis of PFS, regardless of the exact number of deaths. • Clarification to the definition of the Full analysis set, Safety analysis set and PK analysis set • Removal of sensitivity analyses for PRO endpoints • Updated presentation of OS12, PFS6 and PFS12 • Study design section updated • Combination arm regimen change • IDMC first meeting change • Sample size section updated and consequently the statistical assumptions and trigger points for analysis changed accordingly • Change in requirements for re-treatment • Change in clinically meaningful difference for LC13 • Interim analysis for Overall Survival added • Multiplicity strategy • Change to CIs for Overall Survival analysis In line with project wide developments • Changes in definition of important protocol deviations • Removal of RDI2, PID2 and PID. RDI derivation changed • Added further details for definition of baseline • Change to method of subgroup analysis • Definition of TEAE added • Event rate per 100 patients years definition changed • Clarification of death summary Other Changes • Removal of time to first subsequent therapy or death, time to second subsequent therapy or death • Removal of sensitivity analysis modified for confirmation of progression for PFS and ORR • Removal of analysis of Expected Duration of Response • Added age at randomisation (<65, ≥65) 75, ≥75) as a subgroup, removed

Date	Brief description of change
	Standard of Care (gemcitabine versus vinorelbine versus erlotinib) • Clarified the definition of subsequent therapy for safety follow-up • Additional death summary • Textual edits to Derivation of RECIST Visit Responses to provide greater clarity • Small edits to overall visit responses table • Clarification of which CRF fields to use for any Overall Survival analysis where a survival sweep is not performed • Clarify that date of progression from the reviewer who read baseline first will be used if there is no adjudication • Add two missed visits definition for Time from randomisation to second progression • Removal of AESI categories as not mandated to include • Definitions of total and actual exposure for each treatment provided and Dose intensity simplified. Dose delay definition also provided • Health resource derivation, EQ-5D and treatment switching text changed to be in line with new Therapeutic area guidance • More detail about visit windowing added • Clarify PFS2 as time from randomisation to second progression or death and that it will only be analysed for sub-study B • Added details of thyrotoxicity tables • Updated details on Hy's law categories • Clarifications regarding waterfall plots
22 Jan 2018	Throughout: Corrections to minor typos and omitted words. List of Abbreviations: Added SSA (Sub-Study A) and SSB (Sub-Study B). Section 1.1.2 Corrected section references in footnote 'a' of the table outlining the secondary objectives. Section 1.1.3 Removed language indicating that exploratory analyses will be performed when in fact they may or may not be performed Sections 1.2, 4: Portions parallel to excerpts from CSP Amendment 7.0 or IDMC Charter v7.0 updated to reflect parallel changes or updates to those documents, mainly with respect to when and for how long re-treatment may be administered and timing of final analyses given that no interim analyses were performed. Section 2.2 updated information specifying and describing important protocol deviations as per clinicians. Section 3.1 'Example of scaling' corrected units (cm to mm). Section 3.2.5 Was corrected to correspond to the endpoint "Proportion of patients alive and progression free at 6 months"; previously, this section inadvertently repeated the endpoint described in Section 3.2.6, "Proportion of

Date	Brief description of change
	patients alive and progression free at 12 months". Section 3.4.7 Deleted "Note that visits up to 13 days after the last dosing date will be considered as being on treatment for the purposes of visit windowing and may be assigned to an on-treatment visit. Visits after this will be considered as follow-up and may be assigned accordingly." after it was confirmed to refer to a specific case depending on treatment assignment and not widely applicable as a general rule per se. Section 3.5 Added details of PD-L1 status clarifying the groups to be <1% and ≥1% to <25%. Sections 4, 4.2.1, 5.1, 5.2: Addition of language clarifying that final analyses of OS and PFS for SSA will occur at the same time as those for SSB if they have not already been conducted and it is reasonable to conclude that the required number of events have been reached as to be consistent with portions parallel to excerpts from CSP Amendment 7.0 or IDMC Charter v7.0. Also, added "…additionally, at this time the final analyses for OS and PFS for sub-study A will also be conducted if they have not yet been performed and it is reasonable to conclude that the required number of events have been reached." as per IDMC request. Section 4.1: Removed language noting SAS v9.2 or higher is used for all analyses. Section 4.2: Corrected table numbering. Section 4.2 Table 7: PD-L1 positive / negative designations here described the design of the sub-studies rather than the analyses described in the table. These could be misinterpreted and were therefore removed. Analyses specific to the PD-L1 subgroups are described elsewhere. Section 4.2.2.1 changed method of handling ties to be Breslow, rather than Efron to be consistent with the log-rank test. Section 4.2.2.1: Replaced subgroup Stage IIIB vs IV and with Metastatic vs locally advanced. Section 4.2.2.1 Clarified that "When used as stratification factors in stratified analyses, standard of care therapy type and histology will be based on the values entered into IVRS at randomisation, even if it is subsequently discovered that these values were incorrect." Section 4.2.2.1 Clarified that "When used as a stratification factor in a stratified analysis, histology should come from the IVRS; however, when used to define a subgroup or as a general covariate in an unstratified analysis, histology should come from the PATHGEN module of the eCRF." As well as: "Standard of care will come from the IVRS regardless of whether it is used as a stratification factor in a stratified analysis, used to define a subgroup, or as a general covariate in an unstratified analysis." Also, clarified that "Unless otherwise stated above (e.g., standard of care, which always comes from IVRS), note that in general data used to construct

Date	Brief description of change
	subgroups or as a general covariate in an unstratified analysis will come from the eCRF; however, note that when data are used in a stratified analysis, they should come from the IVRS (e.g., histology)." Section 4.2.2.1 and 4.2.2.2 added PD-L1 status (<1% or ≥1% to <25%) to Subgroup Analysis and Effect of Covariates sections, including that it should be included along with the other subgrouping covariates. Section 4.2.2.1 (Treatment Switching): Exploratory analyses described as 'will' be performed changed to 'may' be performed. Section 4.2.10 Health Resource Use: Analyses now noted to 'may' be performed as opposed to 'will' be performed to be consistent with CSP. Section 4.2.11 Adverse Events: Summaries described as CTCAE Grade 3 or Higher now changed to CTCAE Grades 3 or 4. Section 4.2.1.1 Clarified that "Fluctuations observed in CTCAE grades during study will be listed for all AEs." as to be parallel with PACIFIC and to reflect what is already being done. Section 4.2.3 Clarified that the logistic regression models are unstratified models although they adjust for the same stratification factors as in the analysis of the co-primary endpoints. Section 5.1 Added "…However, for practical considerations, if it happens that the analysis time points for the final PFS and OS analyses are closely aligned, based on the occurrences of the events, then one single analysis of OS will be conducted along with the PFS analysis In this case, the entire 0.04 alpha will be utilized for this OS analysis. Section 5.2 Clarified that "If the interim OS analysis is performed, the IDMC will review the efficacy data..." to create parallelism with the IDMC charter. Section 6: Noted the addition of the PD-L1 status grouped by (<1% or ≥1% to <25%) to Subgroup Analysis and Effect of Covariates as different from that which is described in the protocol.

1. STUDY DETAILS

1.1 Study objectives

1.1.1 Primary Objective

Primary Objective:	Outcome Measure:
Sub-study A (PD-L1-positive population) To assess the efficacy of MEDI4736 monotherapy compared with Standard of Care in terms of OS and PFS Sub-study B (PD-L1-negative population) To assess the efficacy of MEDI4736+tremelimumab treatment compared with Standard of Care in terms of OS and PFS	OS PFS using investigator site assessments according to RECIST 1.1 [a,b]

a In each of the Sub-studies A and B the assessment of PFS and OS will be considered co-primary objectives.
b The co-primary analysis of PFS will be based on programmatically derived PFS based upon investigator site assessment. For the analyses to be conducted, see Section 4.2.2.2.

OS Overall survival; PD-L1 Programmed death ligand 1; PFS Progression free survival; RECIST Response Evaluation Criteria In Solid Tumours.

1.1.2 Secondary Objectives

The following are the secondary objectives in Sub-study A (MEDI4736 monotherapy versus Standard of Care) and Sub-study B (MEDI4736+tremelimumab versus Standard of Care):

Secondary Objective:	Outcome Measure:
To further assess the efficacy in terms of: OS12, ORR, DoR, APF6, APF12 and PFS2	OS12 ORR using investigator site assessments according to RECIST 1.1[a] DoR using investigator site assessments according to RECIST 1.1 [a] APF6 and APF12 using investigator site assessments according to RECIST 1.1 [a] PFS2 as defined by local standard clinical practice
To assess the safety and tolerability profile	Adverse events, physical examinations, vital signs including blood pressure, pulse, electrocardiograms, and laboratory findings including clinical chemistry, haematology and urinalysis
To assess the PK of MEDI4736 and tremelimumab	Concentration of PK 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	ADA (confirmatory results: positive or negative; titres [ADA neutralising antibodies will also be assessed])

Secondary Objective:	Outcome Measure:
To assess symptoms and health-related QoL using EORTC QLQ-C30 v3 and LC13	EORTC QLQ-C30: Time to symptom deterioration (fatigue, pain, nausea/vomiting, dyspnoea, loss of appetite, insomnia, constipation, and diarrhoea). Time to QoL/function deterioration (physical function; role function; emotional function; cognitive function; social function and global health status/QoL) LC13: Time to symptom deterioration (dyspnoea, cough, haemoptysis, chest pain, arm/shoulder pain, other pain) Changes in World Health Organisation Performance Status will also be assessed
Sub-study B (PD-L1-negative population) To evaluate the efficacy of MEDI4736+tremelimumab treatment compared with a) MEDI4736 monotherapy and b) tremelimumab monotherapy	PFS[a], ORR[a], DoR[a], using investigator site assessments according to RECIST 1.1, and OS

a Analysis of ORR, DoR, APF6 and APF12 will be based upon investigator assessment. For the analyses to be conducted, see Sections 4.2.3, 4.2.4, 4.2.5, and 4.2.6.

ADA Anti-drug antibody; APF6 Proportion of patients alive and progression free at 6 months from randomisation; APF12 Proportion of patients alive and progression free at 12 months from randomisation; DoR Duration of response; EORTC QLQ-C30 European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire; LC13 Lung Cancer Module; ORR Objective response rate; OS12 Proportion of patients alive at 12 months from randomisation; PFS Progression free survival; PFS2 Time from randomisation to second progression; PK Pharmacokinetic(s); QoL Quality of Life; RECIST Response Evaluation Criteria In Solid Tumours.

In Sub-study B, PFS in MEDI4736+tremelimumab arm will be compared to each of the MEDI4736 and tremelimumab monotherapy arms as part of the contribution of components analysis. This analysis is planned to be performed when approximately 158 PFS events are observed in the MEDI4736+tremelimumab and tremelimumab monotherapy arms. However, for practical considerations, if this time is close to the time of the final PFS analysis, the contribution of components analysis will be conducted at the time of the PFS final analyses. This is considered a secondary objective of the study.

1.1.3 Exploratory Objectives

AE Adverse event; EQ-5D-5L EuroQoL 5-dimension, 5-level health state utility index; IFN Interferon; IL Interleukin; ORR Objective response rate; OS Overall survival; PD-L1 Programmed death ligand 1; PDx Pharmacodynamic(s); PFS Progression free survival; PK Pharmacokinetic(s).

With regards to PD-L1 expression determined by immunohistochemistry, this will be reported in the Clinical Study Report (CSR). CCI

1.2 Study design

This study is a Phase III, randomised, open label, multi-centre study assessing the efficacy and safety of MEDI4736 versus Standard of Care in NSCLC patients with PD-L1-positive tumours (where positive is defined as ≥25% of tumour cells with membrane staining [proprietary PD-L1 immunohistochemistry assay; Ventana Medical Systems, Inc]) and the combination of MEDI4736 plus tremelimumab (MEDI4736+tremelimumab) versus Standard of Care in NSCLC patients with PD-L1-negative tumours (where negative is defined as <25% of tumour cells with membrane staining [proprietary PD-L1 immunohistochemistry assay; Ventana Medical Systems, Inc]).

The original study design intended to recruit 250 patients to Sub-study A. However, due to low patient accrual, the recruitment to Sub-study A was closed in Q1 2016 at which time 126 patients had been randomised. As a result, the analysis plan for Sub-study A was updated and is presented in Protocol Amendment 6.0, dated 31 August 2016.

Approximately 1300 patients will be recruited, with 610 patients expected to be randomised (126 patients in Sub-study A and 480 patients in Sub-study B) at approximately 250 sites worldwide in a 1:1 ratio in Sub-study A and a 3:2:2:1 ratio (MEDI4736+tremelimumab:MEDI4736:Standard of Care:tremelimumab) in Sub-study B.

Sub-study A (patients with PD-L1-positive tumours):

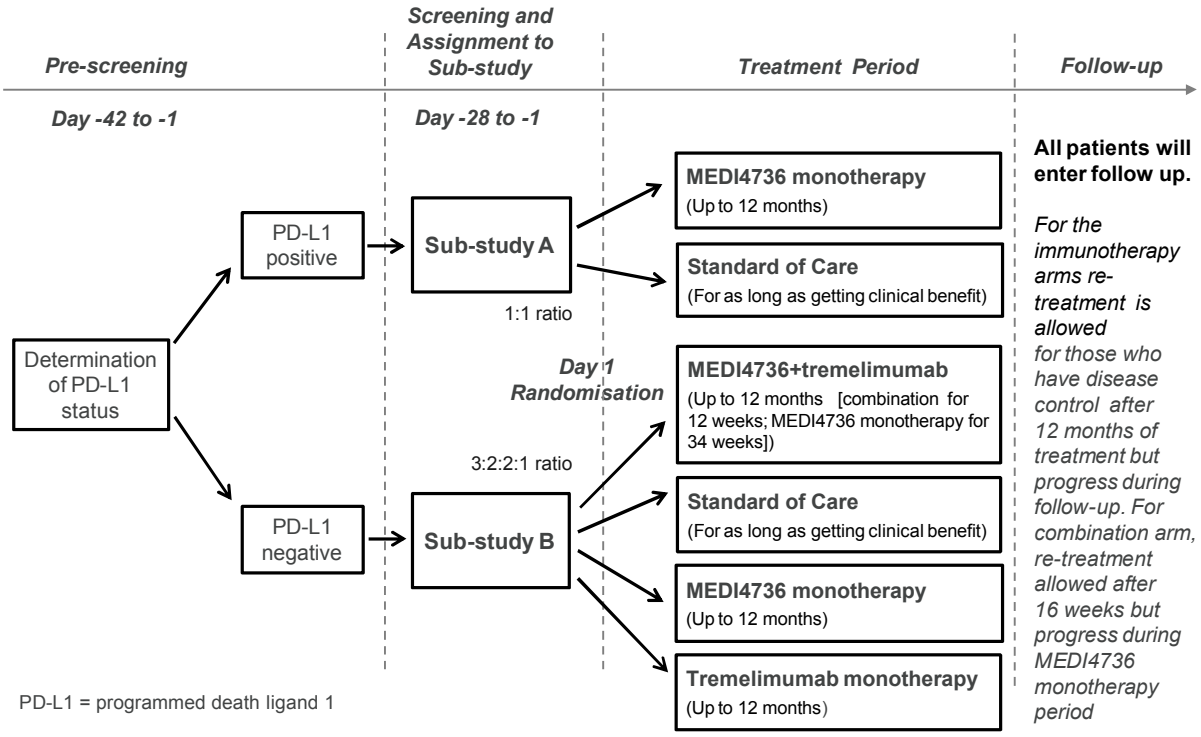
- MEDI4736 (10 mg/kg Q2W iv for up to 12 months) (approximately 60 patients)
- Standard of Care (restricted to the erlotinib, gemcitabine or vinorelbine) (approximately 60 patients). For each agent 4 weeks equates to 1 cycle of treatment.
  - Erlotinib: 150 mg once daily as a tablet for oral administration taken at least 1 hour before or 2 hours after the ingestion of food
  - Gemcitabine: 1000 mg/m^2 iv over 30 minutes on Days 1, 8, and 15 of a 28-day cycle
  - Vinorelbine: 30 mg/m^2 iv on Days 1, 8, 15 and 22 of a 28-day cycle.

Sub-study B (patients with PD-L1-negative tumours):

- MEDI4736+ tremelimumab (MEDI4736 20 mg/kg plus tremelimumab 1 mg/kg Q4W iv for up to 12 weeks [4 doses]) then MEDI4736 alone (10 mg/kg Q2W iv, starting at Week 16, for 34 weeks [18 doses]) (10 mg/kg Q4W iv for 24 weeks then Q12W for a further 24 weeks) (180 patients)
- Standard of Care (see under Sub-study A) (120 patients)
- MEDI4736 (10 mg/kg Q2W iv for up to 12 months) (120 patients)
- Tremelimumab (10 mg/kg Q4W iv for 24 weeks then Q12W for 24 weeks) (60 patients).

The sub-studies may not run concurrently with start and completion of recruitment potentially occurring at different time points. Assignment to the applicable sub-study will be preceded by the Pre-screening Period during which assessment of the patient's PD-L1 status, based on a tumour sample, will take place. After confirmation of PD-L1 status, patients will enter the main Screening Period within their assigned sub-study if it remains open for recruitment. The study design is shown in diagram form in Figure 1.

Figure 1 Study design schema



PD-L1 expression will be determined for all patients prior to randomisation. Patients will be assigned to Sub-study A and Sub-study B respectively based on PD-L1 tumour expression status (PD-L1 positive versus PD-L1 negative [based on an archival tumour sample or a recent tumour biopsy]). Patients will be stratified at randomisation based on the Standard of Care treatment that they would be administered (2 categories: gemcitabine/vinorelbine versus erlotinib) and histology (2 categories: squamous versus all other).

The primary objective of this study is to assess the efficacy of MEDI4736 monotherapy (patients with PD-L1-positive tumours) and MEDI4736+tremelimumab (patients with PD-L1-negative tumours) compared with Standard of Care in terms of OS and PFS (per RECIST 1.1 based on investigator data.

For Sub-study B a contribution of components analysis will be performed when approximately 158 PFS events are observed in the MEDI4736+tremelimumab and tremelimumab monotherapy arms. However, for practical considerations, if this time is close to the time of the final PFS analysis, the contribution of components analysis will be conducted at the time of the final PFS analysis. The purpose of the analysis is to compare the monotherapy and combination dosing regimens to determine if the combination therapy (i.e. MEDI4736+tremelimumab) is more efficacious than each of the monotherapy arms. As these treatment comparisons are secondary and serve a different purpose to that of the primary comparisons, they are not included in the multiple testing procedure for the primary comparisons, and an alpha of 0.05 will be used for either PFS or OS without multiplicity adjustment.

If the contribution of components analysis is performed prior to the final PFS analysis and the MEDI4736+tremelimumab arm is superior to either of the monotherapy arms then that monotherapy arm may be dropped from the study and all patients on that treatment will be discontinued. However, any patients on that treatment will have the opportunity to remain on treatment if they are gaining clinical benefit and if, after discussion with their treating physician, it is felt that this is the best treatment option for them.

Sub-study A: Treatment with MEDI4736 10 mg/kg will commence on Day 1 following randomisation after confirmation of eligibility and will continue on a Q2W schedule for up to 12 months. Treatment should be discontinued prior to 12 months if there is confirmed PD (unless the investigator considers the patient continues to receive benefit from treatment), initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or other reasons to discontinue treatment occur. Patients who have discontinued treatment due to toxicity, symptomatic deterioration or who have commenced subsequent anti-cancer therapy will be followed up until confirmed disease progression or death (whichever occurs first). Study drug should be discontinued if there is confirmed PD while on treatment following a previous response (PR or CR) to study drug while on treatment (ie, the response and progression events both occurred while receiving study drug during the same treatment period).

Sub-study B: Treatment with MEDI4736 or tremelimumab monotherapy, or the combination of MEDI4736+tremelimumab will commence on Day 1 following randomisation after confirmation of eligibility. MEDI4736 will continue on a 10 mg/kg Q2W schedule when given as monotherapy for up to 12 months and tremelimumab when given as a monotherapy, will continue on a 10 mg/kg Q4W schedule for 24 weeks then Q12W for a further 24 weeks. MEDI4736+tremelimumab in combination will be administered from Day 1 on a Q4W schedule up to 12 weeks (4 doses of MEDI4736 20 mg/kg and tremelimumab 1 mg/kg) after which MEDI4736 10 mg/kg alone will be administered on a Q2W schedule for 34 weeks starting at Week 16 (18 doses).

In the monotherapy arms in Sub-study B, treatment should be discontinued prior to 12 months (MEDI4736 monotherapy) or 48 weeks (tremelimumab monotherapy) if there is confirmed PD (unless the investigator considers the patient continues to receive benefit from treatment), initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or other reasons to discontinue treatment occur. Patients who have discontinued treatment due to toxicity, symptomatic deterioration or who have commenced subsequent anti-cancer therapy will be followed up until confirmed disease progression or death (whichever occurs first).

Study drug should be discontinued if there is confirmed PD while on treatment following a previous response (PR or CR) to study drug while on treatment (ie, the response and progression events both occurred while receiving study drug during the same treatment period).

In the MEDI4736+tremelimumab combination arm, if a patient experiences PD, retreatment with the combination regimen (MEDI4736+ tremelimumab) is allowed. Retreatment in the combination arm can only occur if PD, with or without confirmation, occurs during the

MEDI4736 monotherapy portion or after completion of 12 months of therapy. During the retreatment period, the patient would resume MEDI4736 dosing at 20 mg/kg Q4W as during the initial induction period, along with 1 mg/kg of tremelimumab Q4W for 4 doses. Monotherapy with MEDI4736 would then resume at 10 mg/kg Q2W 4 weeks after the last combination dose is administered. Retreatment will continue as long as the investigator judges the patient is deriving clinical benefit.

Both Sub-study A and B: Treatment in the Standard of Care arm will commence on Day 1 following randomisation after confirmation of eligibility and will continue on a 4-weekly schedule until PD (unless the investigator considers the patient continues to receive benefit from treatment), initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or other reasons to discontinue treatment occur.

Tumour assessments using computed tomography (CT)/magnetic resonance imaging (MRI) will be performed every 8 weeks for 48 weeks then 12-weekly thereafter.

Once a patient has had objective progression recorded and has discontinued study drug, the patient will be followed up for survival status every 2 months until death, withdrawal of consent or the final DCO. Patients will also be assessed every 12 weeks for a second progression defined according to local standard clinical practice and may involve any of: objective radiological, symptomatic progression or death.

At the time of the final DCO, the analysis portion of the clinical study will have been completed and all patients remaining in the study will be considered to have completed the analysis portion of the study.

Independent Data Monitoring Committee (IDMC)

An IDMC will be convened and will meet approximately 6 months after the study has started or 20 patients have been randomised to the combination arm on Sub-study B (whichever comes first) to review safety assessments and make recommendations to continue, amend, or stop the study based on safety findings. The committee will meet approximately every 6 months thereafter. The IDMC will also review:

- The interim analysis for OS on Sub-study B (if performed).
- The contribution of components analysis for the immunotherapy arms if this analysis occurs before the primary PFS analysis.

All patients who receive a dose of study treatment will be evaluated for safety and tolerability. Enrolment will continue unless there is an unexpected safety concern. The study may be adjusted or suspended depending on the IDMC review outcome.

Details on the IDMC are provided in Section 5.2 and full details of the IDMC procedures and processes can be found in the IDMC Charter.Number of subjects

The sample size for this study was selected to be consistent with the research hypotheses as described in the protocol Section 1.2.

The co-primary endpoints in sub-study are OS and PFS. To control for type I error, an alpha level of 0.04 will be used for analysis of OS (accounting for one formal interim analysis to assess efficacy) and an alpha level of 0.01 will be used for analysis of PFS. Sub-study B study will be considered positive (a success) if either the PFS analysis results and/or the OS analysis results are statistically significant. Sub-study A will be descriptive in nature with no statistical testing.

A total of approximately 1300 patients are expected to be recruited in the study to achieve 610 evaluable patients in the study (approximately 126 patients in Sub-study A and 480 patients in Sub-study B). The PD-L1-positive population is assumed to be approximately 30% of total population (AstraZeneca, unpublished data). If the prevalence assumption does not hold then it is likely that recruitment will need to be extended. The sample sizing for each sub-study assumes a delay in separation of the survival curves between each arm hence the use of average HRs.

OS analysis

The OS analysis in Sub-study A (PD-L1-positive population) will be performed when 82 deaths have occurred from 126 patients (65% maturity).

The OS analysis in Sub-study B (PD-L1-negative population) will be performed when approximately 205 deaths have occurred from the 300 patients (68% maturity) who have been randomised to the MEDI4736+tremelimumab and Standard of Care arms. Assuming the true average HR is 0.63 in PD-L1-negative population for OS (corresponding to a 4.4-month improvement in median OS from a control group median of 7 months), this analysis will have 90% power to demonstrate a statistically significant difference for OS, assuming an overall alpha level of 0.04 two-sided. The minimal difference in OS that could be deemed statistically significant in Sub-study B is an average HR of 0.75. A 15-month recruitment period and a minimum follow-up period of 12 months is assumed on Sub-study B for OS. Therefore it is anticipated that this OS analysis could be performed 27 months after the first patient has been recruited. Assuming that the survival curves of the two treatment arms do not separate for 2 months then the HR after that point would need to be 0.53 to produce an average HR of 0.63 over the follow-up period. This would be associated with 12-month survival rates of 30.5% for Standard of Care and 48.5% for MEDI4736+tremelimumab.

PFS analysis

The PFS analysis in Sub-study A (PD-L1-positive population) will be performed at the same time as the OS analysis on Sub-study A.

The PFS analysis in Sub-study B (PD-L1-negative population) will be performed when approximately 244 PFS events have occurred from 300 patients (81% maturity) who have been randomised to the MEDI4736+tremelimumab and Standard of Care arms. Assuming the true average HR is 0.63 in PD-L1-negative population for PFS (corresponding to a 1.3-month

improvement in median PFS from a control group median of 3 months), this analysis will have 84% power to demonstrate a statistically significant difference for PFS, assuming an alpha level of 0.01 two-sided. The minimal difference in PFS that would be deemed statistically significant in Sub-study B is an average HR of 0.72. Sub-study B is expected to have a recruitment period of 15 months and a follow-up period of 7 months for the PFS endpoint. Therefore it is anticipated that this PFS analysis could be performed at a minimum of 22 months after the first patient has been recruited. If the survival curves of the two treatment arms do not separate for 2 months then the HR after that point would need to be 0.43 to produce an average HR of 0.63 over the follow-up period. This would be associated with 12-month PFS rates of 6.3% for Standard of Care and 23.3% for MEDI4736+tremelimumab.

The estimates of median PFS (3 months) and median OS (7 months) in the control group used for the design of this study have been estimated following a literature review of available data in the NSCLC patient population (Hanna et al 2004).

2. ANALYSIS SETS

2.1 Definition of analysis sets

Three main analysis sets are defined for this study.

Full analysis set (FAS) (Intention to treat (ITT)):

The FAS will include all randomised patients with treatment groups assigned in accordance with the randomisation, regardless of the treatment actually received. Patients who were randomised but did not subsequently receive treatment are included in the FAS.

The analysis of data using the FAS therefore follows the principles of ITT. Therefore, all efficacy and HRQoL data will be summarised and analysed using the FAS on an ITT basis.

Safety analysis set

All patients who received at least one dose of randomised treatment (regardless of whether that was the randomised therapy intended or indeed whether, in rare cases, they received therapy without being randomised) in each of the sub-studies will be included in the safety population. Throughout the safety results sections, erroneously treated patients (eg, those randomised to Treatment A but actually given Treatment B) will be accounted for in the actual treatment arm. If a patient received treatment from more than one treatment arm then they would be accounted for based upon their initial treatment started.

When assessing safety and tolerability, summaries will be produced based on the safety analysis set.

PK analysis set

All patients who receive at least 1 dose of MEDI4736 (Sub-study A), or either MEDI4736+tremelimumab, MEDI4736 monotherapy or tremelimumab monotherapy (Sub-study B) per the protocol, for whom any post-dose data are available will be included in the PK analysis set. The population will be defined by the Study Team Physician, Pharmacokineticist and Statistician prior to any analyses being performed.

Table 1 gives a summary of outcome variables and analysis populations.

Table 1 Summary of Outcome Variables and Analysis Populations

Outcome variable	Populations
Efficacy Data	
OS, PFS	FAS(ITT)
OS12, ORR*, DoR*, APF6, APF12, PFS2, PRO endpoints*, World Health Organisation performance status	FAS(ITT)
Study Population/Demography Data	
Demography characteristics (e.g. age, sex etc.)	FAS(ITT)
Baseline and disease characteristics	FAS(ITT)
Important deviations	FAS(ITT)
Medical/Surgical history	FAS(ITT)
Previous anti-cancer therapy	FAS(ITT)
Concomitant medications/procedures	FAS(ITT)
Subsequent anti-cancer therapy	FAS(ITT)
PK Data	
PK data	PK
Immunogenicity	
Immunogenicity data	Safety
Safety Data	
Exposure	Safety
Adverse events	Safety
Laboratory measurements	Safety
Vital Signs	Safety
ECGs	Safety

APF6 Proportion of patients alive and progression free at 6 months from randomisation; APF12 Proportion of patients alive and progression free at 12 months from randomisation; DoR Duration of response; ITT Intent-to-Treat; ORR Objective response rate; OS Overall survival; OS12 Proportion of patients alive at 12 months from randomisation; PFS Progression free survival; PFS2 Time from randomisation to second progression; PK Pharmacokinetic; PRO Patient reported outcomes.
*Patients who are evaluable for the analysis of ORR are those with measurable disease at baseline. Patients who are evaluable for the analysis of DoR are those who responded in the ORR analysis. Patient evaluability for PRO endpoints are detailed in Section 3.3.

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. 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 may be performed excluding patients with deviations that may affect the efficacy of the trial therapy if > 10% of patients:

- did not have the intended disease or indication or
- did not receive 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:

- Patients randomised but who did not receive study treatment (Deviation 1).
- Patients who deviate from key entry criteria per the CSP Amendment 5 (Deviation 2). These are inclusion criteria 3, 4, 7 and exclusion criteria 5, 6, 10, 17 and, for sub-study B only, 31, 32, 33.
- Baseline RECIST scan > 42 days before date of randomisation (Deviation 3). Note that although the screening period for baseline RECIST assessment was 28 days, an additional 14-day window should be applied thus only baseline RECIST assessments of greater than 42 days will be deemed as constituting an important deviation.
- No baseline RECIST 1.1 assessment on or before date of randomisation (Deviation 4).
- Received prohibited concomitant systemic anti-cancer agents (Deviation 5). Please refer to the Clinical Study Protocol (CSP) section 5.6 for the systemic anti-cancer agents that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to database lock.
- Patients randomised who received treatment other than that to which they were randomised to (Deviation 6).

- Patients enrolled to the incorrect sub-study (ie patients who are PD-L1 positive who are enrolled onto sub-study B and patients who are PD-L1 negative who are enrolled onto sub-study A (Deviation 7).

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.

Errors in treatment dispensing, in addition to incorrect stratifications, will also be investigated more closely. This may occur when a patient is not randomised or treated according to the randomisation schedule on at least one occasion. It is envisaged that there will be 3 sub categories of this within the important deviations summary:

- 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 and the treatment differs from the randomised treatment.
- 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.

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

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST Visit Responses

For all patients, the RECIST version 1.1 (see further Appendix F of 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 in accordance with RECIST and also their best objective response to study treatment.

The baseline assessment should be performed no more than 28 days before the start of study treatment and ideally as close as possible to the start of randomised treatment. Efficacy for all

patients will be assessed by objective tumour assessments every 8 weeks for the first 48 weeks (relative to the date of randomisation per Table 3 in the CSP for MEDI4736 monotherapy, Table 4 in the CSP for tremelimumab monotherapy, Table 5 in the CSP for MEDI4736+tremelimumab, Table 8 in the CSP for gemcitabine and vinorelbine and Table 9 in the CSP for erlotinib) and then every 12 weeks thereafter until confirmed objective disease progression as defined by RECIST 1.1 (irrespective of the reason for stopping treatment and/or subsequent therapy). If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits. This schedule is to be followed in order to minimise any unintentional bias caused by some patients being assessed at a different frequency than other patients.

For patients who discontinue study drug due to toxicity or a reason other than confirmed PD, objective tumour assessments should be continued every 8 weeks for 48 weeks (relative to the date of randomisation) then every 12 weeks thereafter until confirmed progressive disease by RECIST 1.1 by investigational site review.

In the monotherapy arms and during the first 4 cycles of the combination arm, disease progression requires confirmation and the confirmatory scan should occur preferably at the next scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinically significant deterioration. Administration of study treatment will continue between the initial assessment of progression and confirmation for progression. For all patients who are treated through progression, the investigator should ensure patients do not have any significant, unacceptable or irreversible toxicities that indicate continuing treatment will not further benefit the patient. The patient must continue to meet those inclusion and exclusion criteria that are relevant to treatment through disease progression as specified in Section 4.3 of the CSP. Patients with rapid tumour progression or with symptomatic progression that requires urgent medical intervention (eg, central nervous system metastasis, respiratory failure due to tumour compression, spinal cord compression) will not be eligible to continue to receive study treatment.

Study drug should be discontinued if there is confirmed PD following a previous response (PR or CR) to study drug while on treatment (ie, the response and progression events both occurred while receiving study drug during the same treatment period).

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

Please refer to Section 3.1.3 for the definitions of CR, PR, SD and PD.

RECIST outcomes (ie PFS, ORR etc.) will be calculated programmatically for the site investigator data (see Section 3.2) from the overall visit responses.

3.1.1 Site Investigator Assessment Using RECIST 1.1: Target lesions (TLs)

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 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 (see Section 3.1.2 for further details). If a patient does not have measurable disease at baseline then the TL visit response will be not applicable (NA).

Table 2 TL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all target lesions. Any pathological lymph nodes selected as target lesions must have a reduction in short axis to <10mm.
Partial Response (PR)	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive Disease (PD)	A ≥ 20% increase in the sum of diameters of target lesions and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD

Table 2 TL Visit Responses

Visit Responses	Description
Not Evaluable (NE)	Only relevant in certain situations (i.e. if any of the target lesions were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a target lesion response
Not Applicable (NA)	No target lesions are recorded at baseline

Rounding of TL data

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

Missing TL data

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

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

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

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

Lymph nodes

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

TL visit responses subsequent to CR

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

- Step 1: If all lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is

also met i.e. if a lymph node lesion diameter (LD) increases by 20% but remains < 10 mm.

- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to NE irrespective of whether when referencing the sum of TL diameters the criteria for PD is also met.
- Step 3: If not all lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD
- Step 4: If after steps 1 - 3 a response can still not be determined the response will be set to remain as CR

TL too big to measure

If a TL becomes too big to measure this should be indicated in the database and a size ('x') above which it cannot be accurately measured should be recorded. If using a value of x in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team. 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 5 mm 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) and also biopsy lesions 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 scale up as described below, as long as there remain ≤ 1/3 of the TLs with missing measurements. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then if appropriate, a scaled sum of diameters will be calculated (as long as ≤ 1/3 of the TLs with interventions), 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 has a value of 0 (or <10mm for lymph nodes) recorded. If scaling up is not appropriate due to too few non-missing measurements then the visit response will be set as NE.

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

Scaling (applicable only for irradiated lesions/lesion intervention)

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

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

Example of scaling

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

Lesion 5 has had an intervention at the follow-up visit.

The sum of lesions 1-4 at the follow-up is 26 mm. The sum of the corresponding lesions at baseline visit is 26.8 mm.

Scale up as follows to give an estimated TL sum of 28.4 mm:

(26.0 mm / 26.8 mm) × 29.3 mm = 28.4 mm

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of TLs

CT and MRI are the only methods of assessment that can be used within the trial. 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 Site Investigator Assessment Using RECIST 1.1: Non-target lesions (NTLs) and new lesions

At each visit an overall assessment of the NTL response should be recorded by the investigator. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit.

NTL response will be derived based on the investigator's overall assessment of NTLs as follows:

Table 3 NTL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
Progressive Disease (PD)	Unequivocal progression of existing non-target lesions. 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 non-target lesions were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall non-target lesion assessment at this visit. Note: For patients without target lesions at baseline, this is relevant if any of the non-target lesions 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 Site Investigator Assessment Using RECIST 1.1: Overall visit response

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

Table 4 Overall Visit Responses

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

Table 4 Overall Visit Responses

Target Lesions	Non-target lesions	New Lesions	Overall Response
NA	NE	No (or NE)	NE
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NA	NA	No (or NE)	NED

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, NE = not evaluable, NED = no evidence of disease, NA = not applicable (only relevant if there were no TL/NTL at baseline).

3.2 Outcome Variables

All RECIST assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues investigational product. RECIST outcomes (ie PFS, ORR etc.) will be calculated programmatically for the site investigator data from the overall visit responses.

3.2.1 Co-primary endpoints

Within Sub-study A and Sub-study B, the co-primary endpoints are OS and PFS.

3.2.1.1 Overall survival

Overall survival is defined as the time from the date of randomisation 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, 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). The last date for each individual subject is defined as the latest among the following dates recorded on the CRFs:

- AE start and stop dates
- Admission and discharge dates of hospitalization
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anticancer treatment
- Date last known alive on survival status CRF
- End of study date

3.2.1.2 Progression free survival

PFS (per RECIST 1.1 as assessed by the investigator) will be defined as the time from the date of randomisation until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy prior to progression (ie, date of event or censoring - date of randomisation + 1). Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the two missed visits.

Given the scheduled visit assessment scheme (ie eight-weekly for the first 48 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 274 (ie week 39) then two missing visits will equate to 18 weeks since the previous RECIST assessment, allowing for early and late visits (ie, 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from eight-weekly to twelve-weekly this will equate to 22 weeks (ie, take the average of 8 and 12 weeks which gives 10 weeks and then apply same rationale hence 2 x 10 weeks + 1 week for an early assessment + 1 week for a late assessment = 22 weeks). The time period for the previous RECIST assessment will be from study days 274 to 344 (ie week 39 to week 49). From week 49 onwards (when the scheduling changes to twelve-weekly assessments), two missing visits will equate to 26 weeks (ie 2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks).

If the patient has no evaluable visits or does not have baseline data they will be censored at Day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

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

- RECIST 1.1 assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:
- For investigational assessments, the date of progression will be determined based on the earliest of the RECIST assessment/scan dates of the component that indicates progression
- When censoring a patient for PFS the patient will be censored at the latest of the dates contributing to a particular overall visit assessment.

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

In the absence of clinically significant deterioration the investigational site is advised to continue the patient on study treatment until progression has been confirmed.

3.2.2 Proportion of patients alive at 12 months (OS12)

The proportion of patients alive at 12 months will be defined as the Kaplan-Meier estimate of OS at 12 months.

3.2.3 Objective Response Rate

ORR (per RECIST 1.1 as assessed by the investigator) is defined as the number (%) of patients with at least 1 visit response of CR or PR and will be based on a subset of all randomised patients. If the investigator finds any patients do not have measurable disease at baseline then the analysis of ORR for the investigator data will exclude these patients, so that the denominator is a subset of the Intent-to-Treat (ITT) population who have measurable disease at baseline per investigator (note that although measurable disease per investigator is an inclusion criteria it is possible that the investigator may subsequently determine the patient did not have measurable disease).

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 anti-cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) and then respond will not be included as responders in the ORR.

3.2.4 Duration of Response

DoR (per RECIST 1.1 as assessed by the investigator) 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 (ie, 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 related analysis will be defined as described for ORR.

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

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

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

3.2.5 Proportion of patients alive and progression free at 6 months

The proportion of patients alive and progression free at 6 months (ie, APF6) will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed by the investigator) at 6 months.

3.2.6 Proportion of patients alive and progression free at 12 months

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

3.2.7 Time from randomisation to second progression or death (PFS2) (sub-study B)

PFS2 will be defined as the time from the date of randomisation to the earliest of the progression event (subsequent to that used for the PFS endpoint) 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 3.2.1.2 for details.) The date of second progression will be recorded by the investigator and defined according to local standard clinical practice and may involve any of: objective radiological, symptomatic progression or death. RECIST assessments will not be collected for assessment of PFS2. The date of the PFS2 assessment and investigator opinion of progression status (progressed or non-progressed) at each assessment will be recorded in the electronic case report form (eCRF). Second progression status will be reviewed every 12 weeks following the progression event used for the co-primary variable PFS (the first progression) and status recorded. Patients alive and for whom a second disease progression has not been observed should be censored at the last time known to be alive and without a second disease progression, ie, censored at the latest of the PFS or PFS2 assessment date if the patient has not had a second progression or death. However, if the patient experiences a second progression or dies after two or more missed

visits, the patient will be censored at the time of the last PFS2 assessment prior to the two missed visits.

3.2.8 Best objective response

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment. It is the best response a patient has had following randomisation but prior to starting any subsequent cancer therapy and prior to RECIST progression or the last evaluable assessment in the absence of RECIST progression and subsequent cancer therapy.

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

BoR will be determined programmatically based on RECIST from the overall visit response using all investigator assessment data up until the first progression event, the start of subsequent cancer therapy or the last evaluable assessment in the absence of progression/subsequent cancer therapy.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 8 weeks minus 1 week, i.e. at least 49 days (to allow for an early assessment within the assessment window), after randomisation (ie study day 50). For CR/PR, the initial overall visit assessment which showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

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

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

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

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

3.2.9 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 8, week 16 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 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 investigator data based upon RECIST assessments. Target lesions are measurable tumour lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to randomisation. 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 ± 1 week of the protocol scheduled visit will be used for that visit.

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

3.3 Patient Reported Outcome (PRO) Variables

PRO questionnaires will be assessed using the EORTC-QLQ-C30 with the LC-13 module (HRQoL with lung cancer specific additional concerns) and EQ-5D-5L. 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 Intent-to-Treat (ITT) study population, unless stated.

3.3.1 EORTC-QLQ-C30

The EORTC-QLQ-C30 consists of 30 questions which can be combined to produce 5 functional scales (physical, role, cognitive, emotional, social), 3 symptom scales (fatigue, pain, nausea/vomiting), 5 individual items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea) and a global measure of health status. The EORTC-QLQ-C30 will be scored according to the EORTC scoring manual (Fayers et al 1999). An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales/symptom items, the functional scales and the global health status scale in the EORTC-QLQ-C30 according to the EORTC-QLQ-C30 Scoring Manual. Higher scores on the global health status and functioning scales indicate better health status/function but higher scores on symptom scales/items represent greater symptom severity.

Baseline will be defined as the last non-missing assessment prior to randomisation 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 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 categorised as improvement, no change or deterioration as shown in Table 5.

Table 5 Mean change and visit response in health related quality of life

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

EORTC-QLQ-C30 European Organization for Research and Treatment of Cancer 30-item core quality of life questionnaire.

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 1999). 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 minimised.

For the visit 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 symptom deterioration

For each of the symptoms scales/items in the EORTC-QLQ-C30, time to symptom deterioration will be defined as the time from randomisation until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) or

death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration (ie date of symptom deterioration event or censoring - date of randomisation + 1). Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by EORTC-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 assessment visits or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated (prior to the two missed assessment visits). Given the scheduled visit assessment scheme (ie eight-weekly for the first 48 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change (ie 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks). If the previous PRO assessment is less than study day 274 (ie week 39) then two missing visits will equate to 18 weeks since the previous PRO assessment, allowing for early and late visits. If the two missed visits occur over the period when the scheduled frequency of PRO assessments changes from eight-weekly to twelve-weekly this will equate to 22 weeks (ie, take the average of 8 and 12 weeks which gives 10 weeks and then apply same rationale hence 2 x 10 weeks + 1 week for an early assessment + 1 week for a late assessment = 22 weeks). The time period for the previous PRO assessment will be from study days 274 to 344 (ie week 39 to week 49). From week 49 onwards (when the scheduling changes to twelve-weekly assessments),, two missing visits will equate to 26 weeks (ie 2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks). If the patient has no evaluable visits or does not have baseline data they will be censored at Day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

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

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

3.3.1.2 Time to HRQoL/Function deterioration

For HRQoL/function, time to deterioration will be defined as the time from the date of randomisation until the date of the first clinically meaningful deterioration (a decrease in the function scales or the global health status/HRQoL from baseline of ≥ 10) or 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). Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the HRQoL/function change could be evaluated.

Patients whose HRQoL/function (as measured by EORTC-QLQ-C30) has not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the HRQoL/function could be evaluated. Also, if HRQoL/function deteriorates after 2 or more missed PRO assessment visits (using the same definitions for two missed visits as used in the 'Time to Symptom deterioration' derivation above) or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated. If the patient has no evaluable visits or does not have baseline data they will be censored at Day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

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 analysis, RECIST 1.1 progression will not be considered as HRQoL/function deterioration and data will not be affected by RECIST progression.

3.3.1.3 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.

3.3.1.4 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 EORTC-QLQ-LC-13

The LC-13 is a lung cancer specific module from the EORTC comprising 13 questions to assess lung cancer symptoms (cough, haemoptysis, dyspnoea 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:

- Dyspnoea (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 dyspnoea 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 LC-13 is identical in principle to that for the symptom scales/single items of the EORTC-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 LC-13 (Osoba et al 1998). For example, a clinically meaningful deterioration or worsening in chest pain (as assessed by LC-13) 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 categorised as improvement, no change or deterioration as shown in Table 6.

Table 6 Visit Response for HRQoL and disease-related symptoms

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

HRQoL Health Related Quality of Life; LC13 Lung Cancer Module.

For the visit 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 symptoms scales/items in LC-13, time to symptom deterioration will be defined as the time from the date of randomisation until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration (ie date of symptom deterioration event or censoring - date of randomisation + 1). Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by LC-13) 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 progress after 2 or more missed PRO assessment visits (using the same definitions for two missed visits as used in the 'Time to symptom deterioration' derivation in Section 3.3.1) or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated. If the patient has no evaluable visits or does not have baseline data they will be censored at Day 1 unless they die within

2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

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

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

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 LC-13 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.

3.3.3 EQ-5D-5L

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 by100.

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

3.4 Safety

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

Data from the initial treatment period (ie, the initial 12 months of treatment) on the immunotherapy agents (MEDI4736, tremelimumab or MEDI4736+tremelimumab) will be compared against SOC in the main presentations of safety data and safety data from the re-treatment period 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, tremelimumab 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 certain safety outputs the period of time after the administration of subsequent anti-cancer therapy will not be considered 'on treatment' (see further Section 4.2.11).

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

3.4.1 Adverse events (AEs)

AEs and SAEs will be collected throughout the study, from the date of sub-study informed consent (AEs)/pre-screening informed consent (SAEs) and 90 days after the last dose of immunotherapy agents (ie, the last dose of MEDI4736, tremelimumab or MEDI4736+tremelimumab)/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, tremelimumab or MEDI4736+tremelimumab)/30 days after the last dose of the Standard of Care agents. For the MEDI4736+tremelimumab arm, 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.

Adverse events that have missing causality (after data querying) will be assumed to be related to study drug. Additionally, for the MEDI4736+tremelimumab arm a causality of related or missing for either component will be taken as related to study drug.

Other significant adverse events (OAE)

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

Examples of these are marked haematological 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

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. AESIs represent pre-specified risks which are considered to be of importance to a clinical

development program. These AESIs have been identified as a list of categories provided by the patient safety team.

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 MEDI4736 monotherapy, MEDI4736 on the MEDI4736+ tremelimumab combination arm, tremelimumab on the MEDI4736+ tremelimumab combination arm, tremelimumab monotherapy for the 12-month initial period of treatment and for the retreatment period as follows.

Total (or intended) exposure of MEDI4736 (monotherapy)

- 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 + 13 days" or death date or DCO.

  ie Total (or intended) exposure = min(last dose date where dose>0mg +13, date of death, date of DCO) - first dose date of study drug +1

Total (or intended) exposure of tremelimumab (monotherapy)

- 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 + (13 days or 83 days)" or death date or DCO. Thirteen days will be added in the above formulae if the subject stopped dosing before week 24 and 83 days will be added if the subject stopped dosing at week 24 or later.

  ie Total (or intended) exposure = min(last dose date where dose>0mg +(13 or 83 days), date of death, date of DCO) - first dose date of study drug +1

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 + (13 days or 27 days)" or death date or DCO or start of re-treatment (applies to initial treatment period only). Twenty-seven days will be added in the above formulae if the subject stopped dosing before week 16 and 13 days will be added if the subject stopped dosing at week 16 or later.

  ie Total (or intended) exposure = min(last dose date where dose>0mg +(13 or 27 days), date of death, date of DCO) - first dose date of study drug +1

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.

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

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 Erlotinib

- 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" or DCO.

  ie Total (or intended) exposure = min(last dose date where dose>0mg, date of death, date of DCO) - first dose date of study drug +1

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" or death date or DCO.

  ie Total (or intended) exposure = min(last dose date where dose>0mg + 6 days, date of death, date of DCO) - first dose date of study drug +1

Total (or intended) exposure of Vinorelbine

- 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" or death date or DCO.

  ie Total (or intended) exposure = min(last dose date where dose>0mg + 6 days, date of death, date of DCO) - first dose date of study drug +1

Dose reductions are not permitted per the CSP for the immunotherapy agents (MEDI4736, tremelimumab 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 SOC, the number of days in a cycle will be study specific, but generally a cycle corresponds to a period of 28 days. 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. Each immunotherapy agent will be measured in terms of number of doses given.

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

MEDI4736 (monotherapy):

- Since patients in the MEDI4736 monotherapy treatment group will receive 10 mg/kg MEDI4736 via IV infusion q2w for up to 12 months (up to 26 doses), the duration of dose delays will be calculated as follows:

  For all dosing dates:

  Total duration of dose delays= Sum of (Date of the dose - Date of previous dose - 14 days)

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

MEDI4736 (given in combination):

- Since Patients in the MEDI4736 + Treme treatment group will receive 20 mg/kg MEDI4736 via IV infusion q4w for 4 months and tremelimumab 1 mg/kg q4w for 4 doses followed by MEDI4736 monotherapy at a dose of 10 mg/kg q2w initiated 4 weeks after the last combination dose is administered for up to 18 additional doses, the duration of dose delays will be calculated as follows:

  For Cycle 1 to Cycle 4 (for Week 0 to Week 12) doses:

  Duration1= Sum of (Date of the dose - Date of previous dose - 28 days)

  For Cycle 5 to Cycle 13 (for Week 16 to Week 50) doses:

  Duration2= Sum of (Date of the dose - Date of previous dose - 14 days)

  Total duration of dose delays = Duration1 + Duration2

Tremelimumab (monotherapy):

- Since Patients in the Treme treatment group will receive 10 mg/kg tremelimumab via IV infusion q4w for 7 doses then q12w for 2 additional doses for up to 12 months (up to 9 doses in total), the duration of dose delays will be calculated as follows:

  For Cycle 1 to Cycle 7 (for Week 0 to Week 24) doses:

  Duration1= Sum of (Date of the dose - Date of previous dose - 28 days)

  For Cycle 8 to Cycle 9 (for Week 36 to Week 48) doses:

Duration2= Sum of (Date of the dose - Date of previous dose - 84 days)

Total duration of dose delays = Duration1 + Duration2

Tremelimumab (given in combination):

- Since Patients in the MEDI4736 + Treme treatment group will receive tremelimumab 1 mg/kg q4w for 4 doses only, the duration of dose delays will be calculated as follows:

  For Cycle 1 to Cycle 4 (for Week 0 to Week 12) doses:

  Total 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 immunotherapy agents. It will also be derived for the SOC agents. Relative dose intensity (RDI) is the percentage of the actual dose 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 for all treatments apart from erlotinib.

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 6.4.5 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 3.4.7 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 project-wide 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: Corrected Calcium will be derived during creation of the reporting database using the following formula:

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

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

3.4.5 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 analysis is necessary then QTcF (Fridericia) and QTcB (Bazzetts) will also be provided by the central vendor.

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

3.4.6 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 3.4.7 below will be used.

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

3.4.7 General considerations for safety and PRO assessments

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

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the time window should be based on the actual date and not the intended date of the visit.
- All unscheduled visit data should have the potential to be included in the summaries.
- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day.

  For example, the visit windows for vital signs data for MEDI4736 monotherapy (with 2 weeks between scheduled assessments) are:

  – Day 15, visit window 2 - 21
  – Day 29, visit window 22 - 35
  – Day 43, visit window 36 - 49
  – Day 57, visit window 50 - 63
  – Day 71, visit window 64 - 77
  – Day 85, visit window 78 - 91

  Note that due to the differing assessment schedules the visit windows will be different for the different study treatments.

  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. If there are two values recorded on the same day and the parameter is CTCAE gradeable then the record with the highest toxicity grade should be used. Alternatively, if there are two records recorded on the same day and the toxicity grade is the same (or is not calculated for the parameter) then the average of the two records should be used. The listings should highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all 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 immunotherapy agents 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 (in these cases the toxicity grade would be based upon this averaged 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, negative) is defined according to following criteria:

- Positive:- ≥ 25% tumour cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control.
- Negative:- < 25% tumour cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control.

In addition, for SSB (PD-L1 negative patients) PD-L1 will be further categorised as:

- PD-L1 <1% tumour cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control.
- PD-L1 ≥1% to < 25% tumour cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control.

3.6 Pharmacokinetic and Immunogenicity Variables

Analyses to evaluate the pharmacokinetics and immunogenicity of MEDI4736 and tremeluminab will be performed by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

3.6.1 PK non-compartmental analysis

The actual sampling times will be used in the PK calculations. Pharmacokinetic concentration data and summary statistics will be tabulated. Individual and mean blood MEDI4736 concentration-time profiles will be generated. PK parameters will be determined using standard non-compartmental methods. The following PK parameters will be determined after the first and steady state doses: peak and trough concentration (as data allow). Samples below the lower limit of quantification will be treated as missing in the analyses.

3.6.2 Population PK and exposure-response/safety analysis

A population PK model will be developed using a non-linear mixed-effects modelling approach in patients with NSCLC where possible. The impact of physiologically-relevant patient characteristics (covariates) and disease on PK will be evaluated. The relationship between PK exposure and the effect on safety and efficacy will be evaluated, if appropriate. The results of such an analysis will be reported in a separate report.

3.6.3 Immunogenicity analysis

Immunogenicity results will be analysed descriptively by summarizing the number and percentage of patients who develop detectable anti-MEDI4736 and/or anti-tremelimumab antibodies. The immunogenicity titre will be reported for samples confirmed positive for the presence of anti-MEDI4736 antibodies and/or anti-tremelimumab antibodies. Summaries will be based upon all patients from the safety population. The effect of immunogenicity on PK,

PDx, efficacy and safety will be evaluated, but such analyses, if applicable, will be reported in a separate report.

3.7 Health Resource Use

CCI

[redacted]

4. ANALYSIS METHODS

All analyses and reporting will be separated by sub-study.

Sub-study A has 1 treatment comparison of interest as follows:

- MEDI4736 10 mg/kg Q2W compared with Standard of Care

Sub-study B has 1 treatment comparison of interest that is considered primary as follows:

- MEDI4736 20 mg/kg Q4W plus tremelimumab 1 mg/kg Q4W for 12 weeks then MEDI4736 10 mg/kg Q2W for 34 weeks compared with Standard of Care

The co-primary endpoints in each of the sub-studies are OS and PFS using RECIST v1.1. The study was sized to assess PFS and OS endpoints in sub-study B for the treatment comparisons mentioned above. No hypothesis testing will be performed on OS and PFS in Sub-study A; the analyses will be descriptive.

The final analyses of OS will take place on sub-study B on a pre-specified date when approximately 205 deaths have occurred from 300 patients (68% maturity) who have been randomised to the MEDI4736+tremelimumab and Standard of Care arms in Sub-study B. Interim analyses of OS for the primary treatment comparison (Section 5) will be performed sub-study B at the same time as the primary PFS analysis. However, for practical

considerations, if it happens that the analysis time points for the final PFS and OS analyses are closely aligned based on the occurrences of the events, then 1 single analysis of OS will be conducted along with the PFS analysis. In this case, the entire 0.04 alpha will be utilized for this OS analysis; additionally, at this time the final analyses for OS and PFS for sub-study A will also be conducted if they have not yet been performed and it is reasonable to conclude that the required number of events have been reached.

The final analyses of PFS and OS interim will take place for sub-study B on a pre-specified date when it is predicted that approximately 244 PFS events from 300 patients (81% maturity) who have been randomised to the MEDI4736+tremelimumab and Standard of Care arms.

Sub-study B has the following treatment comparisons of interest that are considered secondary:

1. MEDI4736 20 mg/kg plus tremelimumab 1 mg/kg Q4W for 12 weeks then MEDI4736 10 mg/kg Q2W for 34 weeks compared with MEDI4736 10 mg/kg Q2W
2. MEDI4736 20 mg/kg plus tremelimumab 1 mg/kg Q4W for 12 weeks then MEDI4736 10 mg/kg Q2W for 34 weeks compared with tremelimumab 10 mg/kg Q4W for 24 weeks then Q12W
3. MEDI4736 10 mg/kg Q2W compared with Standard of Care
4. Tremelimumab 10 mg/kg Q4W for 24 weeks then Q12W compared with Standard of Care

A single contribution of components analysis will be performed for superiority analysing PFS and OS for both of the secondary treatment comparisons ‘1’ and ‘2’ above.

This analysis is planned to be performed when approximately 158 PFS events are observed in the MEDI4736+tremelimumab and tremelimumab monotherapy arms. However, for practical considerations, if this time is close to the time of the final PFS analysis, the contribution of components analysis will be conducted at the time of the final PFS analysis. As these treatment comparisons are secondary and serve a different purpose to that of the primary comparisons, they are not included in the multiple testing procedure for the primary comparisons, and an alpha of 0.05 will be used for either PFS or OS without multiplicity adjustment.

4.1 General principles

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

- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarised by the number of observations, mean, standard deviation, median, minimum, and maximum. For log transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median,

minimum and maximum. Categorical variables will be summarised by frequency counts and percentages for each category.

- Unless otherwise stated, percentages will be calculated out of the population total and for each treatment 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.

In general, for efficacy and HRQoL endpoints the last observed measurement prior to randomisation will be considered the baseline measurement. However, if an evaluable assessment is only available after randomisation but before the first dose of randomised treatment then this assessment will be used as baseline. For safety endpoints the last observation before the first dose of study treatment will be considered the baseline measurement unless otherwise specified.

For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose.

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

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

Efficacy and HRQoL data will be summarised and analysed on the FAS. Safety and treatment exposure data will be summarised based upon the safety analysis set. Study population and demography data will be summarised based upon the FAS.

Efficacy from the re-treatment period for the immunotherapy agents may be summarised separately for the site investigator data. For the site investigator data, any derivations relative to baseline (eg RECIST derivations) in the re-treatment period will be relative to the baseline scan prior to re-treatment.

Safety data will be summarised from the initial treatment period (ie the initial 12 months of treatment) only for the immunotherapy agents alongside the SOC agents. Safety data from the re-treatment period 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 analyses in Sub-study B will be presented using 95% CIs and 2-sided p-values, unless otherwise stated. In Sub-study A no p-values will be presented, but 95% CIs will be used throughout.

Table 7 details which endpoints are to be subject to formal statistical analysis, together with pre-planned sensitivity analyses making clear which analysis is regarded as primary for that endpoint.

Table 7 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoints Analysed	Notes
Overall Survival	Sub-study A Hazard ratio comparing MEDI4736 versus Standard of Care Sub-study B Stratified log rank test comparing MEDI4736+tremelimumab versus Standard of Care Sensitivity analysis using a Kaplan-Meier plot of time to censoring where the censoring indicator of the primary analysis is reversed - attrition bias Sub-study B Only The following secondary treatment comparisons will be performed: i) MEDI4736+tremelimumab versus MEDI4736 ii) MEDI4736+tremelimumab versus tremelimumab iii) MEDI4736 versus Standard of Care iv) Tremelimumab versus Standard of Care Supportive Analysis on Sub-study B only Subgroup analysis using Cox proportional hazards models Secondary analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate Secondary analysis using Cox proportional hazards models to determine the consistency of treatment effect between subgroups via the approach of Gail and Simon 1985.

Table 7 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoints Analysed	Notes
Progression Free Survival	Sub-study A Hazard ratio using site investigator data (RECIST 1.1) comparing MEDI4736 versus Standard of Care Sub-study B Stratified log rank test using site investigator data (RECIST 1.1) comparing MEDI4736+tremelimumab versus Standard of Care Sensitivity analyses using site investigator data (RECIST 1.1) 1) Interval censored analysis - evaluation time bias 2) Analysis using alternative censoring rules - attrition bias Sub-study B only The following secondary treatment comparisons will be performed: i) MEDI4736+tremelimumab versus MEDI4736 ii) MEDI4736+tremelimumab versus tremelimumab iii) MEDI4736 versus Standard of Care iv) Tremelimumab versus Standard of Care Supportive analysis on Sub-study B only Subgroup analysis using Cox proportional hazards models Secondary analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate Secondary analysis using Cox proportional hazards models to determine the consistency of treatment effect between subgroups via the approach of Gail and Simon 1985.
The analyses specified below will be performed in each of the Sub-studies A and B:	
Proportion of patients alive at 12 months	Kaplan-Meier estimates of survival at 12 months and p-value for comparison (sub-study B) (following the method described by Klein et al 2007)
Objective Response Rate	Logistic regression using site investigator data (RECIST 1.1)
Proportion of patients alive and progression free at 6 and 12 months	Kaplan Meier estimates of progression free survival at 6 and 12 months
Time from randomisation to second progression	Stratified log-rank test (sub-study B)
Symptom improvement rate (EORTC QLQ-C30 and LC13 endpoints)	Logistic regression
HRQoL/Function improvement rate (EORTC QLQ-C30 endpoints)	Logistic regression
Time to HRQoL/Function deterioration (EORTC QLQ-C30 endpoints)	Stratified log-rank test in Sub-study B. Hazard ratios only in Sub-study A

Table 7 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoints Analysed	Notes
Time to symptom deterioration (EORTC QLQ-C30 and LC13 endpoints)	Stratified log-rank test in Sub-study B. Hazard ratios only in Sub-study A

EORTC QLQ-C30 European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire; LC13 Lung Cancer Module; PD-L1 Programmed death ligand 1; HRQoL Health Related Quality of Life; RECIST Response Evaluation Criteria In Solid Tumours; Sub-study A MEDI4736 monotherapy v Standard of Care; Sub-study B MEDI4736+tremelimumab v Standard of Care.

All outputs will be summarised for each Sub-study by treatment arm for all randomised patients (ITT).

4.2.1 Multiplicity (Sub-study B only)

The multiple testing procedure will define which significance levels should be applied to the interpretation of the raw p-values for the 2 primary endpoints of PFS and OS and the key secondary endpoints of OS12 and ORR.

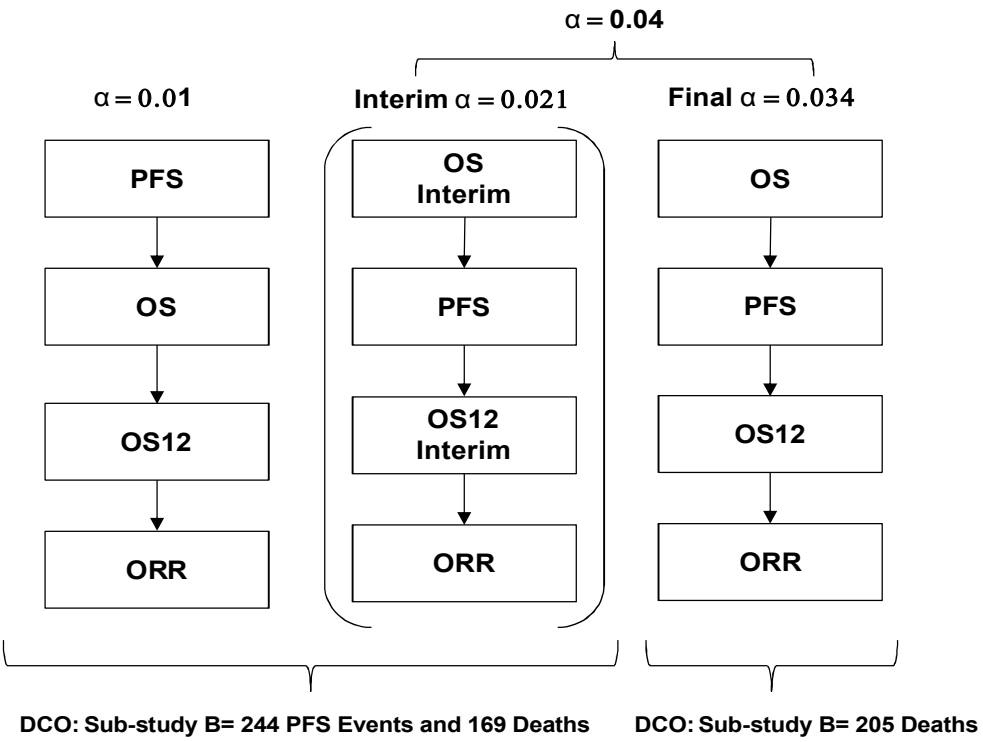
The overall type I error of 0.05 will be split between the co-primary endpoints OS and PFS. To control for type I error, an alpha of 0.04 will be used for the analysis of OS and an alpha of 0.01 will be used for the analysis of PFS. The study will be considered positive if the PFS analysis results and/or the OS analysis results are statistically significant. The 0.04 alpha level allocated to OS will be controlled at the interim and primary time point 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.

An interim OS analysis for superiority and the primary PFS analysis will occur at the same time and the primary OS analysis will be performed when it is expected 205 deaths have accumulated from patients who have been randomised to the MEDI4736+tremelimumab and Standard of Care arms. For example, if 82% of the deaths required at the time of the primary OS analysis are available at the time of the interim (ie, 169/205 deaths have occurred), the two-sided alpha level to be applied in the OS interim analysis would be 0.021 and the two-sided alpha level to be applied for the primary OS analysis would be 0.034.

At the time of the primary PFS, interim OS and primary OS analyses, the primary and key secondary hypotheses will be tested on the primary treatment comparisons only, using a multiple testing procedure with an alpha-exhaustive recycling strategy (Burman et al 2009). No adjustment will be made for the contribution of components analysis as it is not concerned with testing the primary treatment comparison. With this approach, hypotheses will be tested in a pre-defined order. At the time of the primary PFS analysis, the PFS endpoint will be tested first and at the time of the primary OS analysis, the OS endpoint will be tested first. The other hypotheses corresponding to secondary endpoints will then be tested in a pre-specified hierarchy following PFS and OS rejection. 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 0.05 (two-sided), amongst all key hypotheses. Figure 2 shows the multiple testing framework. Alpha will be recycled within Sub-study B.

Figure 2 Multiple testing procedures for controlling the type 1 error rate for sub-study B



DCO**:** Data cut-off

Upon achieving statistical significance on the PFS endpoint in sub-study B, the testing of the OS endpoint will be performed hierarchically as illustrated in Figure 2. Similarly the testing of the PFS endpoint will be done subsequent to achieving statistical significance on the interim/primary OS endpoint in sub-study B. If both of these endpoints are significant, the alpha level can be combined and passed down to lower levels in the hierarchy. Spending alpha between endpoints in this way will strongly control type I error (Glimm et al 2010).

It is currently anticipated that the cut-off for PFS co-primary analysis will be before the cut-off for OS co-primary analysis on sub-study B. Alpha will be recycled across the PFS and OS hierarchies at the time of the final analysis of the respective endpoints. If the PFS and OS analyses are closely aligned and performed at the same time, the same alpha split (0.01 vs. 0.04) will be applied to the PFS analysis and OS analysis, and the alpha will be recycled between PFS and OS if either of them is significant.

4.2.2 Co-primary endpoints

4.2.2.1 Overall survival

OS will be analysed in sub-study B using a stratified log-rank test adjusting for Standard of Care therapy type (gemcitabine/vinorelbine versus erlotinib) and histology (squamous versus all other) for generation of the p-value and using the Breslow approach for handling ties (Breslow, 1974). The effect of treatment will be estimated by the HR together with its corresponding ([1-adjusted alpha] x 100)% CI and p-value. Note that the alpha-adjusted CI p-value will only be generated on sub-study B. The HR and CI will be generated on sub-study A. The boundaries (ie, adjusted alpha levels) for the treatment comparison at the interim and final analysis for OS on sub-study B 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). In sub-study B, any of the secondary treatment comparisons at the time of final analysis will display the 95% CI.

When used as stratification factors in stratified analyses, standard of care therapy type and histology will be based on the values entered into IVRS at randomisation, even if it is subsequently discovered that these values were incorrect.

The HR and its CI will be estimated from a stratified Cox Proportional Hazards model (Cox 1972) (with ties = Breslow and the stratification variables included in the strata statement) and the CI calculated using a profile likelihood approach.

Kaplan-Meier plots of OS will be presented by treatment arm within each sub-study. 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.

OS12 will be summarised (using the Kaplan-Meier curve) and presented by treatment arm. For each treatment arm, the survival rate at 12 months based on Kaplan-Meier method will be presented, along with its 95% confidence interval. The computation of the confidence interval will be based on a log(-log(.)) transformation.

For the comparison between treatments (sub-study B only), the test will be based on the method described in Klein 2007 and p-value (Klein et al 2007). The test statistic and its variance estimate are as follows:

- test statistic = $\ln \frac{\ln \hat{S}_1(t)}{\ln \hat{S}_2(t)}$
- Variance estimate = $\frac{\hat{\sigma}_1(t)^2}{\ln^2 S_1(t)} + \frac{\hat{\sigma}_2(t)^2}{\ln^2 S_2(t)}$

where $\hat{\sigma}_i(t)^2 = \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}$ is the variance derived from Greenwood's formula $S(t)$ and can be estimated from standard software packages, where d_i and n_i refer to the number of deaths and patients at risk for each risk set.

The z-statistic is then calculated as: $\frac{\text{test statistic}}{\sqrt{\text{variance estimate}}}$

For stratified analysis, the test statistic and its variance estimate in each stratum will be combined by weighting inversely proportionately according to each within stratum variance (Whitehead and Whitehead 1991). A Z-test will be performed and the p-value from the test will be presented.

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

Sensitivity Analyses (Sub-study B only)

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

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

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

- In censored patients who are alive at DCO only: Time from randomisation to date of censoring (date last known to be alive) by treatment arm.

- In all patients: Time from randomisation to the date of death (ie overall survival) or to the date of censoring for censored patients regardless of treatment arm.

Subgroup analyses (Sub-study B only)

Subgroup analyses will be conducted in Sub-study B comparing OS between the treatments concerned in the primary treatment comparison in the following subgroups of the FAS:

- Sex (male versus female)
- Age at randomisation (<65 versus ≥65 years of age)
  - This will be determined from the date of birth (BIRTHDAT in the DEM module) and date of randomisation (RND_DAT in the CRIT1 module) on the eCRF at screening. Patients with a partial date of birth (ie for those countries where year of birth only is given) will have an assumed date of birth of 1st Jan [given year]). Patients with a missing age value will be included using the mean age (overall FAS) and categorised accordingly.
- Age at randomisation (<75 versus ≥75 years of age)
  - This will be determined from the date of birth (BIRTHDAT in the DEM module) and date of randomisation (RND_DAT in the CRIT1 module) on the eCRF at screening. Patients with a partial date of birth (ie for those countries where year of birth only is given) will have an assumed date of birth of 1st Jan [given year]). Patients with a missing age value will be included using the mean age (overall FAS) and categorised accordingly.
- Histology (squamous versus all other)
  - When used as a stratification factor in a stratified analysis, histology should come from the IVRS; however, when used to define a subgroup or as a general covariate in an unstratified analysis, histology should come from the PATHGEN module of the eCRF.
- Smoking (smoker [SUTRTNIC=1, 2, 3, 4, 13] versus non-smoker [never smoked])
  - This will be determined from the response to 'Nicotine Use Occurrence' (SU module) on the eCRF at screening. Patients with a missing smoking status will be included in the 'smoker' category.
- Standard of Care grouped (gemcitabine/vinorelbine versus erlotinib)
  - Standard of care will come from the IVRS regardless of whether it is used as a stratification factor in a stratified analysis, used to define a subgroup, or as a general covariate in an unstratified analysis.
- WHO performance status at baseline (normal activity [PSTAT=0] versus restricted activity [PSTAT=1]).
  - This will be determined from the response to "Performance status" (PSTAT module) on the eCRF at screening. Patients with a missing performance status will be included in the 'restricted activity' category.

- Region (Asia, Europe, South America versus North America).
  - This will be determined from the centre number (CENTRE). If there are less than 20 events in the "South America" category, these patients will be combined with those in North America.
- Race (White, Black/African-American, Asian, Other [Native Hawaiian/Pacific Islander or American Indian/Alaska Native or Others]).
  - This will be determined from the response to "Race" (DEM module) on the eCRF at screening.
- Metastatic versus locally advanced
- PD-L1 status <1% versus ≥ 1% to < 25%
- Line of therapy (3^{rd}, 4^{th}, $>4^{th}$)
  - This will be derived from the response to the "Number of Prior Anti-Cancer Therapy Regimens" (CAPRX module) on the eCRF at pre-screening.
- Site of Local/Metastatic Disease at Study Entry (Brain/CNS [DISSITES=1 or 16] and/or Liver [DISSITES=22 or 8], Other)
  - This will be determined from the response to the "Site of Local/Metastatic Disease at Study Entry" (DISEXT module) on the eCRF at screening.

Unless otherwise stated above (e.g., standard of care, which always comes from IVRS), note that in general data used to construct subgroups or as a general covariate in an unstratified analysis will come from the eCRF; however, note that when data are used in a stratified analysis, they should come from the IVRS (e.g., histology).

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

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

For each subgroup, the HR (MEDI4736+tremelimumab: Standard of Care in Sub-study B) and 95% CI will be calculated from an unstratified Cox proportional hazards with treatment as only covariate. The Cox models will be fitted using SAS® PROC PHREG with the Breslow method to control for ties, using the by statement to obtain HR and 95% CI for each subgroup level separately.

These hazard ratios and associated two-sided 95% CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis.

If there are too few events available for a meaningful analysis of a particular subgroup (it is not considered appropriate to present analyses where there are less than 20 events in a

subgroup), the relationship between that subgroup and OS will not be formally analysed. In this case, only descriptive summaries will be provided.

Effect of covariates on the HR estimate

Cox proportional hazards modelling will be employed to assess the effect of covariates on the HR estimate for the primary treatment comparison of Sub-study B. A model will be constructed, containing treatment and the stratification factors alone, to ensure any output from the Cox modelling is likely to be consistent with the results of the stratified log-rank test.

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

Additional covariates for this model will be sex, age at randomisation (<65 versus ≥65 years of age), smoking, WHO performance at baseline, region, race, stage, line of therapy, site of local/metastatic disease at study entry, and PD-L1 status (<1% versus ≥ 1% to < 25%).

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

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

Treatment switching / exploratory analysis of overall survival

CCI

[redacted]

4.2.2.2 Progression free survival

PFS based upon the programmatically derived RECIST outcome using the site investigator assessment data (using all scans regardless of whether they were scheduled or not) will be

analysed using stratified log-rank tests (on Sub-study B, no p-values will be generated for Sub-study A) using the same methodology as described for the OS analyses.
The effect of treatment will be estimated by the HR together with its corresponding 99% CI and p-value for sub-study B. The HR and 95% CI will be generated on Sub-study A. Kaplan-Meier plots of PFS will be presented by treatment arm for each sub-study. Summaries of the number and percentage of patients experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided along with median PFS for each treatment.

The assumption of proportionality will be assessed in the same way as for OS on Sub-study B. The analysis will be based on the PFS from the site investigator data.

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 (Sub-study B only)

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

Additionally, summary statistics will be given for the number of days from censoring to 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 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 (Sub-study B only)

The following sensitivity analyses will only be performed for the primary treatment comparison in Sub-study B.

- Evaluation-Time bias

A sensitivity analysis will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled time points. The midpoint between the time of progression and the previous evaluable RECIST assessment (using the final date of the assessment) will be analysed using a stratified log-rank test, as described for the co-primary analysis of PFS. For patients whose death was treated as PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). To support this analysis, the mean of subject-level average inter-assessment times will be tabulated for each treatment. This approach will use the site investigator RECIST data.

- Attrition bias

Attrition bias will be assessed by repeating the co-primary PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following two, or more, non-evaluable tumour assessments will be included. In addition, and within the same sensitivity analysis, patients who take subsequent therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) prior to their last evaluable RECIST assessment or progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier plot of the time to censoring from the primary analysis where the censoring indicator of the PFS analysis is reversed.

- Deviation bias

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

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of progression free survival on Sub-study B.

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

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

The effect of covariates upon the HR estimate and the consistency of treatment effect between subgroups will be analysed for PFS using the same methods as those described for OS.

4.2.3 Objective response rate

The ORR will be based on the programmatically derived RECIST outcome using the site investigator tumour data, and using all scans regardless of whether they were scheduled or not. The ORR will be compared between MEDI4736 versus Standard of Care in Sub-study A and MEDI4736+tremelimumab versus Standard of Care in Sub-study B using unstratified

logistic regression models adjusting for the same stratification factors as the co-primary endpoints. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour MEDI4736 monotherapy in Sub-study A/ MEDI4736+tremelimumab in Sub-study B) together with its associated profile likelihood 95% CI (e.g. using the option 'LRCI' in SAS procedure GENMOD). A p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model) will only be generated on Sub-study B.
If there are not enough responses for a meaningful analysis using logistic regression then a Fisher's exact test using mid p-values will be presented.

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

$$\text{Fisher's exact test mid p-value} = \text{Two sided p-value} - \frac{\text{Table probability}}{2}$$

Summaries will be produced that present the number and percentage of patients with a tumour response (CR/PR) based upon the number of patients with measurable disease at baseline per investigator (see Section 3.2.3). For each treatment arm, best objective response (BoR) will be summarised by n (%) for each category (CR, PR, SD, PD and NE). No formal statistical analyses are planned for BoR.

4.2.4 Duration of response

For both sub-studies descriptive data will be provided for the DoR in responding patients (ie median duration of response and 95% CIs) by treatment arm, including the associated Kaplan-Meier curves (without any formal comparison of treatment arms or p-value attached).

4.2.5 Proportion of patients alive and progression free at 6 months (APF6)

The proportion of patients alive and progression free at 6 months (ie at study day 184) will be summarised (using the Kaplan-Meier curve) and presented by treatment arm on both sub-studies. For each treatment arm, the APF6 based on Kaplan-Meier method will be presented, along with its 95% confidence interval. The computation of the confidence interval will be based on a log(-log(.)) transformation.

4.2.6 Proportion of patients alive and progression free at 12 months (APF12)

The proportion of patients alive and progression free at 12 months (ie at study day 366) along with 95% CIs will be summarised (using the Kaplan-Meier curve) and presented by treatment arm on both sub-studies. For each treatment arm, the APF12 based on Kaplan-Meier method will be presented, along with its 95% confidence interval. The computation of the confidence interval will be based on a log(-log(.)) transformation.

4.2.7 Time from randomisation to second progression (Sub-study B only)

Time from randomisation to second progression or death (PFS2) will be analysed using identical methods as outlined for the analysis of PFS and adjusting for the same set of covariates, but no subgroup analysis will be performed. The HR for the treatment effect

together with its 95% CI will be presented. Medians and Kaplan-Meier plots will be presented to support the analysis.

The number and percentage of subjects experiencing a PFS2 event and the type of progression (objective progression by RECIST, symptomatic progression, new or worsening of soft tissue/visceral or bone metastases or other) will also be summarised by treatment arm.

Time from randomisation to second progression will be summarised by treatment arm.

4.2.8 Change in tumour size

The absolute values, change in TL tumour size from baseline and percentage change in TL tumour size from baseline will be summarized using descriptive statistics and presented at each timepoint and by randomized 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 by randomised treatment group.

Tumour size will also be presented graphically using waterfall plots for each treatment arm, 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. A reference line at the -30% change in TL tumour size level will be added to the plots, which corresponds with the definition of 'partial' response. All progressions will be marked with a '●'. . The scale in these plots will be fixed to be from -100 to 100 to avoid presenting extreme values. Values that are capped as a result of this restriction to the scale are marked with '#'. Values are ordered in descending order with the imputations due to death appearing first followed by a gap followed by all other patients. On each of the waterfall plots the 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 on data based upon site investigator RECIST assessments.

4.2.9 Patient reported outcomes

The PRO endpoints that have been identified as primary are EORTC QLQ-C30 time to HRQoL deterioration for global health status and LC13 time to symptom deterioration for each of dyspnea, cough, hemoptysis, and chest pain. These are not part of the main multiple testing procedure and as supportive endpoints will need a Bonferroni adjustment to the significance level to aid interpretation. Therefore, these 5 endpoints will be tested at a 1% significance level and 99% CIs will be produced.

The other time to symptom deterioration endpoints will be tested at a 5% significance level and 95% CIs will be produced.

For sub-study A no p-values will be presented.

4.2.9.1 EORTC QLQ-C30

Time to symptom deterioration will be analysed for each of the 3 symptom scales (fatigue, pain, nausea/vomiting) and the 5 individual symptom items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea). Time to HRQoL/function deterioration will be analysed for the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/HRQoL. This will be achieved by comparing between treatment arms using a stratified log-rank test as described for the primary analysis of OS.

The HR and 95% CI for each scale/item will be presented graphically on a forest plot.

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 HRQoL/function improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/HRQoL will be produced. Symptom improvement rate and HRQoL/function improvement rate will be analysed by comparing between treatment arms using a logistic regression model as described for the analysis of ORR. The odds ratio and 95% CI for each scale/item will be presented graphically on a forest plot.

For the primary endpoint, global health status/HRQoL, and the endpoints appetite loss, fatigue and physical functioning time to deterioration will be presented using a Kaplan-Meier plot. Summaries of the number and percentage of patients experiencing a clinically meaningful deterioration or death, and the median time to deterioration will also be provided for each treatment arm.

Summaries of original and change from baseline values of each symptom scale/item, the global HRQoL score and each functional domain will be reported by visit for each treatment arm. Graphical presentations may also be produced as appropriate. Summaries of the number and percentage of patients in each response category at each visit for each ordinal item (in terms of the proportion of patients in the categories of improvement, no change, and deterioration as defined in Section 3.3.1) will also be produced for each treatment arm.

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

4.2.9.2 LC13

Time to symptom deterioration for each of the 6 individual symptoms (dyspnoea, cough, haemoptysis, chest pain, arm/shoulder pain, other pain) will be compared between treatment arms using a stratified log-rank test as described for the primary analysis of OS.

The HR and 95% CI for each scale/item will be presented graphically on a forest plot.

For the primary endpoints in LC13, time to deterioration in symptoms will be presented using a Kaplan-Meier plot. Summaries of the number and percentage of patients experiencing a clinically meaningful deterioration or death, and the median time to deterioration will also be provided for each treatment arm.

A summary of the symptom improvement rate for each of the 6 individual symptom items will be produced. The symptom improvement rate will be compared between treatment arms using a logistic regression model as described for ORR. The odds ratio and 95% CI for each symptom will be presented graphically on a forest plot.

Summaries of original and change from baseline values of each symptom (dyspnoea, cough, haemoptysis, chest pain, arm/shoulder pain, other pain) and each treatment-related side effect (sore mouth, dysphagia, peripheral neuropathy and alopecia) will be reported by visit for each treatment arm. Graphical presentations may also be produced as appropriate. Summaries of the number and percentage of patients in each response category at each visit for each ordinal symptom item (in terms of the proportion of patients in the categories of improvement, no change, and deterioration as defined in Section 3.3.2) will also be produced for each treatment arm.

4.2.9.3 EuroQol 5-Dimension 5-Level questionnaire

CCI [redacted]

4.2.10 Health Resource Use

CCI [redacted]

4.2.11 Safety data

Safety and tolerability data will be presented by actual treatment group in each sub-study using the safety population. Safety data will be summarised only. No formal statistical analyses will be performed on the safety data.

Any safety summaries examining retreatment with MEDI4736 monotherapy in Sub-study A or retreatment with MEDI4736+tremelimumab or MEDI4736 monotherapy or tremelimumab

monotherapy in Sub-study B will be produced separately. For safety summaries, if a patient starts re-treatment then this is considered a subsequent therapy.

The following sections describe the planned safety summaries for AEs, vital signs, laboratory parameters and ECG. However, additional safety tables (not specified in this SAP) may need to be produced 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 all 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 the summary tables.

AEs observed up until 90 days following discontinuation of the immunotherapy agents (ie, the last dose of MEDI4736, tremelimumab or MEDI4736+tremelimumab)/30 days following discontinuation of the Standard of Care agent or until the initiation of the first subsequent anti-cancer therapy (including radiotherapy, with the exception of palliative radiotherapy) following discontinuation of treatment (whichever occurs first) will be used for reporting of all of the AE summary tables. This will more accurately depict AEs attributable to study treatment only as a number of AEs up to 90 days following discontinuation of the immunotherapy agents/30 days following discontinuation of the Standard of Care agent are likely to be attributable to subsequent therapy.

However, to assess the longer term toxicity profile, 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 therapy into account).

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved) and investigator's assessment of severity and relationship to study drug. Frequencies and percentages of patients reporting each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from on any 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) (Sub-study B only)
- AEs with outcome of death
- AEs with outcome of death causally related to study medication (as determined by the reporting investigator)
- All SAEs
- All SAEs causally related to study medication (as determined by the reporting investigator)
- 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 dose delay of study medication
- Other significant AEs
- 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 higher, 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 each preferred term, the event rate is defined as the number of patients with that AE divided by the total drug exposure of patients and then multiplied by 365.25 x 100 to present in terms of per 100 patient years.

Summaries of the number and percentage of patients 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 all AEs.

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 over 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

This summary will be repeated, including all relevant rows, for all deaths within 90 days of last dose of study medication.

Adverse events of special interest

Preferred terms used to identify AESI 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 one AESI presented by outcome

- At least one AESI causally related to study medication (as determined by the reporting investigator) by CTCAE grade
- At least one AESI leading to discontinuation of study medication

A summary of total duration (days) of AESI will be provided for events which have an end date. 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.

Laboratory assessments

Data obtained up until the 90 days following discontinuation of immunotherapy agents (ie, the last dose of MEDI4736, tremelimumab or MEDI4736+tremelimumab)/30 days following discontinuation of the Standard of Care agent or until the initiation of the first subsequent anti-cancer-therapy (including radiotherapy, with the exception of palliative radiotherapy) following discontinuation of treatment (whichever occurs first) will be used for reporting. This will more accurately depict laboratory toxicities attributable to study treatment only as a number of toxicities up to 90 days following discontinuation of 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, some summaries of laboratory data may also be produced containing data collected up until 90 days following discontinuation of the immunotherapy agents/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/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/30 days last dose for Standard of Care agents will not be summarised.

Data summaries will be provided in preferred units.

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

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

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

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

- Haematology: Haemoglobin, 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

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

Liver Enzyme Elevations and Hy's law

The following summaries will include the number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study
  - ALT ≥ 3x -≤ 5x, > 5x - ≤8x, > 8x - ≤ 10x, >10x - ≤ 20x, and >20x Upper Limit of Normal (ULN) during the study
  - AST ≥ 3x- -≤ 5x, > 5x - ≤8x, > 8x - ≤ 10x, >10x - ≤ 20x, and >20x ULN during the study
  - Total bilirubin ≥2x-≤3x, >3x-≤5x, >5x ULN during the study
  - ALT or AST ≥3x - ≤5x, >5x - ≤8x, >8x - ≤ 10x, >10x - ≤ 20x, >20x ULN during the study
  - ALT or AST ≥3x ULN and Total bilirubin ≥2x ULN during the study (Potential Hy's law): The onset date of ALT or AST elevation should be 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 by cohort.

Shift tables showing baseline to maximum and baseline to minimum as well a summary of abnormal thyroid tests will also be produced for TSH, T3 and T4.

ECGs

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

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

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.

Physical examination

All individual physical examination data will not be summarised.

Other Safety Data

Data from positive pregnancy tests will not be summarised.

4.2.12 WHO performance status

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

4.2.13 PK data (MEDI4736 monotherapy, MEDI4736+tremelimumab and tremelimumab monotherapy arms only)

Pharmacokinetic concentration data will be listed for each patient and each dosing day, and a summary provided for all evaluable patients in each sub-study. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

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.

The effect of immunogenicity on PK, PDx, efficacy and safety will be evaluated if data allow. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

4.2.14 PK/PDx relationships (MEDI4736 monotherapy, MEDI4736+tremelimumab and tremelimumab monotherapy arms only)

If the data are suitable, the relationship between PK exposure and efficacy/safety parameters may be investigated graphically or using an appropriate data modelling approach. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

4.2.15 Biomarker data

CCI [redacted]

4.2.16 Demographic and baseline characteristics data

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

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations
- Inclusion in analysis populations

- Demographics (age, age group [<50, >=50-< 65, ≥ 65 - <75 years, and ≥ 75 years], sex, race and ethnicity)
- Patient characteristics at baseline (height, weight, weight 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 (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 at baseline
- Disease related medical history (past and current)
- Relevant surgical history
- 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 by randomised treatment group:

- Total exposure of each treatment group.
- Actual exposure of each treatment group (for immunotherapy agents only).
- Total number of cycles received (for SOC treatments only).
- Reasons for dose delays and infusion interruptions of MEDI4736 and tremelimumab and reasons for dose delays/infusion interruptions, dose reductions and dose modifications for the relevant Standard of Care agents (gemcitabine and

vinorelbine). Dose delays and infusion interruptions will be based on investigator initiated dosing decisions.

- Number of infusions received (for all treatments apart from Erlotinib).
- RDI of MEDI4736, tremelimumab and Standard of Care agents.

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

5. INTERIM ANALYSES

5.1 Analysis methods

An interim analysis will be performed to test OS on the primary treatment comparison in Sub-study B.

In Sub-study B the OS interim analysis will be performed when approximately 244 PFS events have occurred in patients who have been randomised to the MEDI4736+tremelimumab and Standard of Care arms. It is expected that approximately 169 (82% of the target 205) deaths will be observed at this time. However, for practical considerations, if it happens that the analysis time points for the final PFS and OS analyses are closely aligned, based on the occurrences of the events, then one single analysis of OS will be conducted along with the PFS analysis In this case, the entire 0.04 alpha will be utilized for this OS analysis.

All interim analyses will be assessed by an IDMC (further details are given in the IDMC charter). It is probable that recruitment will have completed prior to the results of the interim analysis being available.

5.1.1 Overall survival

No interim analysis will be performed on Sub-study A.

For sub-study B, the criterion for superiority is a statistically significant improvement in OS at the interim analysis. The Lan and DeMets approach that approximates the O'Brien Fleming spending function will be used to account for multiplicity introduced into the treatment comparison by including the interim analysis for superiority (Lan and DeMets 1983).

If 82% of the deaths required (assuming a HR of 0.63) at the time of the primary OS analysis (approximately 56% maturity) are available at the time of the interim (ie, 169/205 deaths have occurred), the two-sided alpha level to be applied in the OS interim analysis would be 0.021. This analysis would have 69% power to detect a HR of 0.65. The minimal difference in OS that would be deemed statistically significant is an average HR of 0.70.

It is estimated that this interim analysis will be performed approximately 22 months after the start of randomisation.

Under the above assumptions the two-sided alpha level to be applied for the primary OS analysis would be 0.034.

The OS interim will be analysed using a stratified log-rank test (see Section 4.2.2.1 for details). The hazard ratio will be estimated with corresponding CI and p-value. The size of the CI will be determined based on the actual number of events included in the interim analysis.

The survival status of patients at the time of the interim analysis will be summarised. A Kaplan-Meier plot of OS will be presented by randomised treatment group, along with median OS. The number of OS events will be presented by randomised treatment group.

If the PFS and/or OS results indicate superiority, then analyses of all other endpoints would be performed and the results of these analyses will form the basis for submissions for regulatory approval. Patients would continue to be followed for survival until the required number of patients have died on the respective sub-study, when an updated analysis would be performed.

If the PFS result is not statistically significant and the OS interim analysis result does not meet the criterion of stopping for superiority, then the extended study team at AstraZeneca/MedImmune will remain blinded and the patients will continue to be followed for survival.

The recommendations from the IDMC will not reveal the results of the analysis but will take the form of "Continue/Modify/Recommend Early Submission/Stop".

5.2 Independent Data monitoring committee

This study will use an external IDMC to assess ongoing safety analyses on both sub-studies as well as the interim analyses for superiority. The committee will meet approximately 6 months after the study has started or 20 patients have been randomised to the combination arm (whichever comes first) to review the safety data from the study. The IDMC will meet approximately 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.

In addition:

- If the contribution of components analysis is performed prior to the final PFS analysis, the IDMC will review the efficacy data on Sub-study B for the MEDI4736+tremelimumab versus tremelimumab monotherapy treatment comparison and the MEDI4736+tremelimumab versus MEDI4736 monotherapy treatment comparison.
- If the interim OS analysis is performed, the IDMC will review the efficacy data on Sub-study B when approximately 244 PFS events have occurred from patients randomised to the MEDI4736+tremelimumab and Standard of Care arms, at approximately 18 months post first randomisation, at the time of the primary analysis of PFS and interim analysis of OS.

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

Section of SAP Affected (If applicable)	**Change**	**Rationale**
4.2.3.1	Remove HR and added p-value to the presentation of OS12	p-value required for multiple testing procedure
4.2.6/7	Change presentation of proportion of patients alive and progression free at 6 and 12 months	To match TFL standards
4.2.2.1 and 4.2.2.2	Addition of PD-L1 status (with groups of <1% vs ≥1% to <25%) as a subgrouping factor and covariate to be analysed in a manner similar to covariates and subgroups previously established in the protocol	To align with emergent interest in this biomarker since the trial started
4.2.3.1 Overall Survival	Treatment Switching analysis references and choice of analysis changed	Changed to be in line with proposed Therapeutic area guidance
4.2.10.3 Euro-Qol 5-Dimension 5-Level questionnaire	Mixed model removed and more detail of descriptive statistics provided	Changed to be in line with proposed Therapeutic area guidance
4.2.10 Health Resource Use	More accurate description of summaries to be produced	Changed to be in line with proposed Therapeutic area guidance

7. REFERENCES

Berry et al 1991
Berry G, Kitchin RM, Mock PA. A comparison of two simple hazard ratio estimators based on the logrank test. Stat Med 1991;10(5):749-55.

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

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(5):739-61.

Glimm et al 2010
Glimm E, Maurer W and Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Statist Med 2010;29:219-28.

Collett 2003
Collett D. Modelling survival data in medical research: 2nd ed. Chapman and Hall/CRC. 2003. Available on request.

Cox 1972
Cox DR. Regression models and life-tables. J Royal Stat Society Series B 1972;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 1999
Fayers P, Aaronson NK, Bjordal K, Curran D, Groenvold M on behalf of the EORTC Quality of Life Study Group. EORTC QLQ-C30 Scoring Manual: 2nd Edition 1999. Available on request.

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.

Hanna et al 2004
Hanna N, Shepherd FA, Fossella FV, Pereira JR, De Marinis F, von Pawel J, et al. Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. J Clin Oncol 2004;22(9):1589-97.

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

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

Latimer 2014
Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL, Campbell MJ. Adjusting survival time estimates to account for treatment switching in randomised controlled trials - an economic evaluation context: Methods, limitations and recommendations. Med Decis Making 2014;34(3):387-402

Nishino et al 2013
Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH and Hodi FS. Developing a common language for tumour response to immunotherapy: Immune-related response criteria using unidimensional measurements. Clin Cancer Res 2013;19:3936-43

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

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

Pintilie M.
Pintilie M. Competing risks: A practical perspective. Wiley.

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.

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.

Sellke and Siegmund 1983
Sellke T, Siegmund D. Sequential analysis of the proportional hazards model. Biometrika 1983;70:315-26.

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. Statistics In Medicine 1991;10:1665-77.

Wolchok et al 2009
Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clin Cancer Res 2009;15(23):7412-20.