



Statistical Analysis Plan

Study Code D419AC00003

Edition Number 3.0

Date 17 May 2019

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

	PAGE
TITLE PAGE.....	1
TABLE OF CONTENTS.....	2
LIST OF ABBREVIATIONS.....	6
AMENDMENT HISTORY.....	9
1. STUDY DETAILS.....	10
1.1 Study objectives.....	10
1.2 Study design.....	12
1.3 Number of patients.....	15
2. ANALYSIS SETS.....	17
2.1 Definition of analysis sets.....	17
2.1.1 Full analysis set (Intention to Treat (ITT)).....	17
2.1.2 bTMB \geq 20 mut/Mb analysis set.....	17
2.1.3 bTMB \geq 16 mut/Mb analysis set.....	17
2.1.4 bTMB \geq 12 mut/Mb analysis set.....	17
2.1.5 bTMB non-evaluable analysis set.....	17
2.1.6 bTMB evaluable analysis set.....	17
2.1.7 bTMB no-call analysis set.....	17
2.1.8 bTMB $<$ 20 mut/Mb analysis set.....	18
2.1.9 PD-L1-negative analysis set.....	18
2.1.10 PD-L1-based analysis sets.....	18
2.1.11 tTMB - based analysis sets.....	18
2.1.12 Safety analysis set.....	18
2.1.13 bTMB \geq 20 mut/Mb safety analysis set.....	18
2.1.14 bTMB \geq 16 mut/Mb safety analysis set.....	18
2.1.15 bTMB \geq 12 mut/Mb safety analysis set.....	19
2.1.16 PD-L1-negative safety analysis set.....	19
2.1.17 bTMB-evaluable safety analysis set.....	19
2.1.18 Pharmacokinetic analysis set.....	19
2.1.19 ADA evaluable set.....	19
2.2 Violations and Deviations.....	22
3. PRIMARY AND SECONDARY VARIABLES.....	23
3.1 Derivation of RECIST Visit Responses.....	23
3.1.1 Investigator RECIST 1.1-based assessments: Target lesions.....	24
Rounding of TL data.....	26
Missing TL data.....	26

	Lymph nodes	26
	TL visit responses subsequent to CR	27
	TL too big to measure	27
	TL too small to measure	27
	Irradiated lesions/lesion intervention	27
	Scaling (applicable only for irradiated lesions/lesion intervention)	28
	Lesions that split in two	29
	Lesions that merge	29
	Change in method of assessment of TLs.....	29
3.1.2	Investigator RECIST 1.1-based assessments: Non-target lesions and new lesions.....	29
3.1.3	Investigator RECIST 1.1-based assessments: Overall visit response	31
3.2	Outcome Variables.....	31
3.2.1	Primary endpoint - overall survival	32
3.2.2	Secondary endpoints	32
3.2.2.1	Progression-free survival.....	32
3.2.2.2	Objective response rate	33
3.2.2.3	Duration of response	33
3.2.2.4	Time from randomization to second progression	34
3.2.2.5	Best objective response	34
3.2.2.6	Change in tumor size.....	35
3.3	Safety.....	35
3.3.1	Adverse events (AEs).....	36
	AEs of special interest.....	36
3.3.2	Treatment exposure.....	37
	Patients who permanently discontinue during a dose delay.....	38
3.3.3	Dose intensity	38
3.3.4	Laboratory data	39
3.3.5	ECGs	40
3.3.6	Vital signs.....	40
3.3.7	General considerations for safety assessments.....	40
3.4	Other Biomarker Variables.....	43
3.5	Pharmacokinetic and Immunogenicity variables	43
3.5.1	Population pharmacokinetics and exposure-response/safety analysis	43
3.5.2	Pharmacokinetic non-compartmental analysis	43
3.5.3	Immunogenicity analysis.....	43
3.6	Health Resource Use.....	44
4.	ANALYSIS METHODS.....	44
4.1	General principles	44
4.2	Analysis methods.....	46
4.2.1	Multiple testing strategy.....	50
4.2.2	Analysis of primary endpoint - overall survival	51

4.2.3	Progression-free survival.....	54
4.2.4	Objective response rate	56
4.2.5	Duration of response	57
4.2.6	Survival rate at 12 months, 18 months, and 24 months	57
4.2.7	Proportion of patients alive and progression free at 12 months	57
4.2.8	Time from randomization to second progression	57
4.2.9	Subsequent Therapy.....	57
4.2.10	Change in tumor size in the bTMB \geq 20 mut/Mb analysis set.....	58
4.2.11	Healthcare resource use.....	58
4.2.12	Safety data	58
	Adverse Events	59
	Deaths.....	62
	Adverse events of special interest.....	63
	Summary of long term tolerability.....	64
	Laboratory assessments.....	65
	Liver Enzyme Elevations and Hy's law	66
	Thyroid function tests	67
	ECGs	67
	Vital signs.....	67
	Time to Subsequent Therapy from discontinuation of study treatment.....	67
	Other Safety Data.....	68
4.2.13	ECOG performance status.....	68
4.2.14	PK data (MEDI4736+tremelimumab arm only).....	68
	Immunogenicity analysis.....	68
4.2.15	PK/PDx relationships (MEDI4736+tremelimumab)	68
4.2.16	Biomarker data.....	68
4.2.17	Demographic and baseline characteristics data	69
4.2.18	Treatment exposure.....	70
4.2.19	Subsequent Therapy.....	71
5.	INTERIM ANALYSES	71
5.1	Independent Data Monitoring Committee.....	71
6.	CHINA COHORT	72
7.	CHANGES OF ANALYSIS FROM PROTOCOL	72
8.	REFERENCES	72
9.	APPENDIX	74

LIST OF TABLES

Table 1	Summary of statistical assumptions.....	16
Table 2	Summary of outcome variables and analysis populations.....	19
Table 3	TL visit responses	25
Table 4	NTL Visit Responses	30
Table 5	Overall visit responses	31
Table 6	Overview of Main Pre-planned statistical and sensitivity analysis to be conducted.....	46

LIST OF FIGURES

Figure 1	Overall study design.....	13
Figure 2	Study flow chart.....	14
Figure 3	Multiple testing procedure for controlling the Type I error rate	50

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
≥	Greater than or equal to
<	Less than
ADA	Anti-drug antibody
AE	Adverse event
AJCC	American Joint Committee on Cancer
ALK	Anaplastic lymphoma kinase
ALT	Alanine aminotransferase
APF12	Proportion of patients alive and progression free at 12 months from first dose
AST	Aspartate aminotransferase
Baseline	Refers to the most recent assessment of any variable prior to dosing with study treatment, except for efficacy
BDR	Blinded data review
BDRM	Blinded data review meeting
BoR	Best objective response
CI	Confidence Interval
CR	Complete Response
CRF/eCRF	Case Report Form (electronic)
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed tomography
CTC/CTCAE	Common Terminology Criteria for Adverse Event (National Institutes of Health, National Cancer Institute)
DBP	Diastolic blood pressure
DCO	Data cut off
DoR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal growth factor receptor
FAS	Full analysis set
FDA	Food and Drug Administration

Abbreviation or special term	Explanation
HR	Hazard ratio
IDMC	Independent data monitoring committee
IC	Immune Cells
IP	Investigational Product (includes MEDI4736, tremelimumab and Standard of Care therapies as specified in the protocol)
ITT	Intention to Treat
IVRS	Interactive Voice Response System
Mb	Megabase
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
mut	Mutations
MRI	Magnetic Resonance Imaging
MTP	Multiple testing procedure
nAb	Neutralizing antibody
NE	Not evaluable
NSCLC	Non-small cell lung cancer
NTL	Non-target lesions
OAE	Other Significant Adverse Event
ORR	Objective Response Rate
OS	Overall Survival
OS12	Overall Survival, proportion of patients alive at 12 months
OS18	Overall Survival, proportion of patients alive at 18 months
OS24	Overall Survival, proportion of patients alive at 24 months
PAP	Payer Analysis Plan
PD	Progressive disease
PD-L1	Programmed cell death ligand 1
PD-L1-negative analysis set	See section 2.1.7
PD-L1 positive _{e25%}	See section 2.1.8
PD-L1 positive _{e50%}	See section 2.1.8
PDx	Pharmacodynamic(s)
PFS	Progression-free survival
PFS2	Time from randomisation to second progression or death

Abbreviation or special term	Explanation
PK	Pharmacokinetic
PR	Partial Response
QT/QTcF	QT interval/QT interval (corrected for heart rate using Fredericia's correction)
RDI	Relative dose intensity
RR	Response Rate
RECIST 1.1	Response Evaluation Criteria In Solid Tumors, Version 1.1
SAE	Serious adverse event
SAS	Safety analysis set
SBP	Systolic blood pressure
SD	Stable disease
SoC	Standard of Care
TL	Target lesions
TMB	Tumor mutational burden
bTMB	Blood tumor mutational burden
tTMB	Tissue tumor mutational burden
ULN	Upper limit of normal
WHO	World Health Organization

AMENDMENT HISTORY

Date	Brief description of change
17 May 2019	Document updated to reflect changes made within clinical protocol amendment version 8. <ul style="list-style-type: none">- Changed primary and secondary objectives- Changed the Multiple Testing Procedure (MTP)- Included statistical analyses for all new endpoints/analysis sets- Introduced a new sensitivity analyses for the primary endpoint- Maturity of the final analysis has been increased
13 March 2018	Document updated to reflect changes made within clinical protocol amendment version 7. For example, the interim analysis planned to occur after approximately 80% of the target events has been removed. In addition, the maturity of the final analysis has been increased.

1. STUDY DETAILS

This statistical analysis plan (SAP) contains a more detailed description of the analyses in the clinical study protocol (CSP). This SAP is based on version 8 of the CSP dated 3 April 2019.

1.1 Study objectives

All objectives will be evaluated for all patients, unless otherwise indicated. The primary objective is assessment of overall survival (OS) in patients in the analysis set bTMB \geq 20mut/Mb (Section 2.1.1)

Primary objectives:	Outcome measures:
To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to Standard of Care (SoC) in terms of Overall survival (OS) in patients with bTMB \geq 20mut/Mb.	OS

Secondary objectives:	Outcome measures:
To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of OS	OS in patients with <ul style="list-style-type: none"> • bTMB \geq 16mut/Mb NSCLC • bTMB \geq 12 mut/Mb NSCLC • PD-L1-negative NSCLC (Section 2.1.7) • All patients with NSCLC • bTMB < 20 mut/Mb NSCLC • bTMB non-evaluable population

Secondary objectives:	Outcome measures:
<p>To further assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of PFS, ORR, DoR, OS12, OS18, OS24, APF12, and PFS2*</p> <p>*See list of abbreviations for PFS, ORR, DoR, OS12, OS18, OS24, APF12, and PFS2; see Sections 3 and 4 for definition and analysis details</p>	<p>PFS, ORR, DoR and APF12 using Investigator assessments according to RECIST 1.1 in patients with</p> <ul style="list-style-type: none"> • bTMB \geq 20mut/Mb NSCLC, • bTMB \geq 16mut/Mb NSCLC • bTMB \geq 12mut/Mb NSCLC • PD-L1 negative NSCLC • All patients with NSCLC <p>PFS2 using local standard clinical practice^a, OS12, OS18 and OS24 in patients with</p> <ul style="list-style-type: none"> • bTMB \geq 20mut/Mb NSCLC • bTMB \geq 16mut/Mb NSCLC • bTMB \geq 12mut/Mb NSCLC • PD-L1 negative • All patients with NSCLC
<p>To assess efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of ORR, PFS and OS in further PD-L1 defined populations</p>	<p>OS in patients with PD-L1 TC \geq 25% and PD-L1 TC \geq 50%</p> <p>PFS and ORR using Investigator assessments according to RECIST 1.1 in patients with PD-L1 TC \geq 25% and PD-L1 TC \geq 50%.</p>
<p>To further assess efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of ORR, PFS and OS (including OS12, OS18 and OS24).</p>	<p>OS (including OS12, OS18 and OS24) and PFS/ORR using Investigator assessments according to RECIST 1.1 in patients in analysis sets defined by tTMB cutoffs [tTMB determined on CCI [REDACTED]]</p> <ul style="list-style-type: none"> • tTMB \geq 14mut/Mb NSCLC • tTMB \geq 12mut/Mb NSCLC • tTMB \geq 10mut/Mb NSCLC • tTMB \geq 8mut/Mb NSCLC
<p>To assess the pharmacokinetics (PK) of MEDI4736 + tremelimumab combination therapy</p>	<p>Concentration of MEDI4736 and tremelimumab in blood and non-compartmental PK parameters, such as peak concentration and trough (as data allow; sparse sampling)</p>
<p>To investigate the immunogenicity of MEDI4736 and tremelimumab</p>	<p>Presence of anti-drug antibodies (ADAs) for MEDI4736 and tremelimumab (confirmatory results: positive and negative; titers)</p>

^a PFS2 will be defined as the time from the date of randomization to the earliest of the progression event subsequent to that used for the PFS endpoint, or death

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

An objective to meet China Health Authority requirement is to evaluate consistency in efficacy and safety among Chinese patients for benefit-risk assessment of MEDI4736 + tremelimumab combination therapy compared to SoC. Further details regarding the Chinese specific analysis, please refer to the China specific SAP.

1.2 Study design

This is a randomized, open-label, multi-center, global Phase III study to determine the efficacy and safety of MEDI4736 + tremelimumab combination therapy versus platinum-based SoC chemotherapy in the first-line treatment of patients with EGFR and ALK wild-type advanced or metastatic NSCLC. Crossover from SoC to MEDI4736 + tremelimumab combination therapy will not be permitted. A schematic diagram of the overall study design is shown in [Figure 1](#), and a detailed study flow chart is shown in [Figure 2](#).

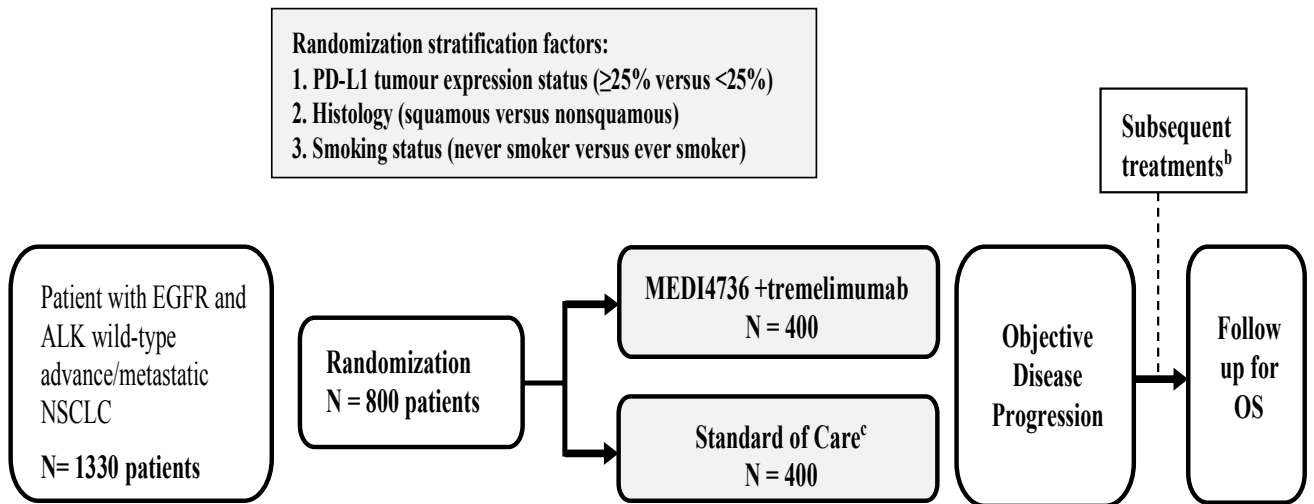
This study will enroll approximately 1330 patients at sites in North America, South America, Asia, Europe and Middle East/Africa countries to randomize approximately 800 patients (including approximately 336 patients with PD-L1-positive_{25%} NSCLC) to treatment. Global recruitment will be complete once approximately 800 patients have been randomized of which 30 patients are from mainland China. Once global enrolment is completed, the recruitment will continue in mainland China only. A total of approximately 160 patients from mainland China will be randomized.

Patients will provide a tumor tissue sample at enrolment to determine PD-L1 expression status (defined by the CCI). Mutation assessments for bTMB will be performed on the CCI.

- ≥ 20 mut/Mb tumor mutational burden in blood is considered as bTMB high
- < 20 mut/Mb tumor mutational burden in blood is considered as bTMB low
- $\geq 25\%$ PD-L1 and $\geq 1\%$ PD-L1 membrane expression in tumoral tissue are considered as relevant positive sub-groups
- $< 25\%$ PD-L1 is considered low/negative
- $< 1\%$ PD-L1 is considered negative

Patients will be randomized in a 1:1 ratio in a stratified manner according to PD-L1 tumor expression status ($\geq 25\%$ versus $< 25\%$), histology (squamous versus non-squamous), and smoking status (never smoker versus ever smoker) to receive treatment with MEDI4736 + tremelimumab combination therapy or SoC therapy.

Figure 1 Overall study design

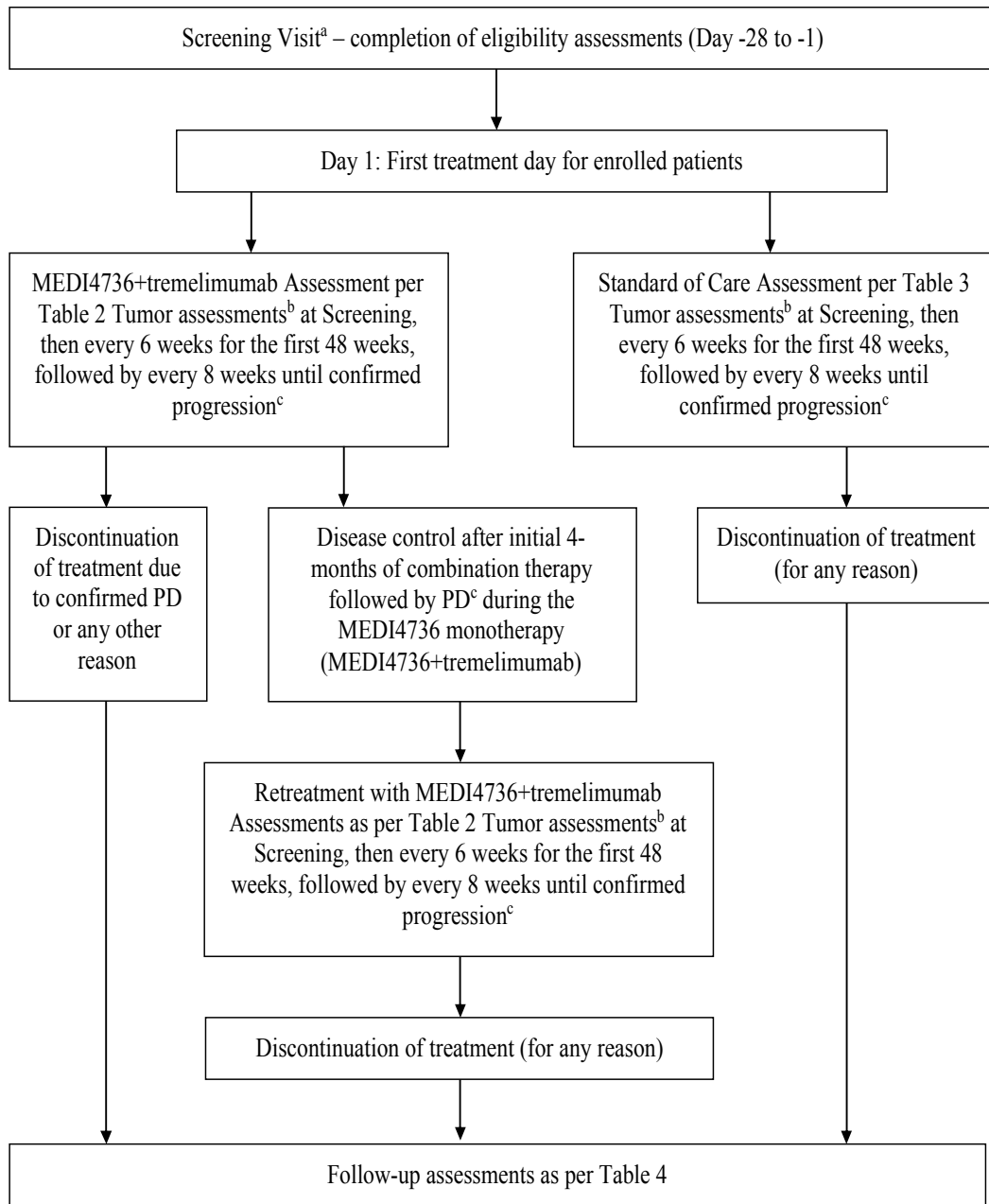


^a The same study design will be applied to the China cohort. The number of patients in Figure 1 reflects those for the global cohort. Enrollment in China will continue after the global cohort enrollment is completed

^b Offer of standard chemotherapy per Investigator's discretion

^c SoC is an Investigator choice from the following: paclitaxel + carboplatin, gemcitabine + cisplatin (or carboplatin) (squamous only), pemetrexed + cisplatin (or carboplatin) (non-squamous only), and for eligible patients, pemetrexed maintenance (non-squamous only following pemetrexed/platinum induction)

Figure 2 Study flow chart



^a Informed consent of study procedures and tumor biopsy sample may be obtained prior to the 28-day screening window, if necessary, in order to permit tumor biopsy sample acquisition and analysis prior to randomization
^b Tumor assessments were performed using RECIST 1.1
^c A confirmatory scan is always required following the initial demonstration of PD (See Section 5.1 of the CSP for more information)

Independent Data Monitoring Committee (IDMC)

An IDMC comprised of independent experts will be convened and will meet approximately 6 months after the study has started or after the first 30 patients have been randomized, whichever occurs first, to review safety assessments and make recommendations to continue, amend, or stop the study based on safety findings. The committee will meet at least every 6 months thereafter.

Details of the IDMC are provided in Section 5, and full details of the procedures and processes can be found in the IDMC Charter.

1.3 Number of patients

The study will plan to enroll approximately 1330 patients in order to randomize 800 eligible patients 1:1 to MEDI4736 + tremelimumab combination therapy or SoC globally. The 800 patients will comprise approximately 336 patients who have PD-L1-positive_{25%} tumors and approximately 520 patients who have PD-L1-positive_{1%} tumors. Global recruitment will be complete once approximately 800 patients have been randomized of which 30 patients are from mainland China. Once global enrollment is completed, the recruitment will continue in mainland China only. A total of approximately 160 patients from mainland China will be randomized.

The sample size calculation assumes a 3-month delay in separation of the OS curves between each group and with a 12-month recruitment period and a minimum follow up period of 43 months assumed, it is anticipated that this analysis will be performed 55 months after the first patient has been recruited. The assumed true OS average hazard ratios (HR) in bTMB_{≥20}, _{≥16}, _{≥12}, and PD-L1 negative population are based on D419AC00001 study, as well as on what would be considered a clinically meaningful improvement in each of the selected biomarker populations. The expected number of patients in each biomarker subgroup-in this study is based on observation from the D419AC00001 study.

The primary analysis will be performed when approximately 87% maturity in the population with bTMB \geq 20 mut/Mb is achieved.

MEDI4736 + tremelimumab versus SoC (OS in bTMB_{≥20} population)

Assuming the true OS average hazard ratio (HR) is 0.49 and the median OS in SoC chemotherapy alone arm is 10 months following an exponential distribution for both MEDI4736 + tremelimumab and SoC arm in the bTMB_{≥20} mut/Mb population, with approximately 140 patients, 122 OS events from the global cohort (approximately 87% maturity) will provide greater than 90% power to demonstrate statistical significance at the 2-sided alpha overall level of 5%. The smallest treatment difference that is statistically significant will be an HR of 0.70.

MEDI4736 + tremelimumab versus SoC (OS in bTMB_{≥16} mut/Mb population)

Assuming the true OS average HR is 0.62 and the median OS in SoC chemotherapy alone arm

is 10 months following an exponential distribution for both MEDI4736 + tremelimumab and SoC arm in the bTMB \geq 16 mut/Mb population, with approximately 212 patients, 193 OS events from the global cohort (approximately 91% maturity) will provide greater than 90% power to demonstrate statistical significance at the 2-sided alpha overall level of 5%. The smallest treatment difference that is statistically significant will be an HR of 0.75.

MEDI4736 + tremelimumab versus SoC (OS in bTMB \geq 12 population)

Assuming the true OS average HR is 0.65 and the median OS in SoC chemotherapy alone arm is 10 months following an exponential distribution for both MEDI4736 + tremelimumab and SoC arm in the bTMB \geq 12 mut/Mb population, with approximately 304 patients, 280 OS events from the global cohort (approximately 92% maturity) will provide greater than 90% power to demonstrate statistical significance at the 2-sided alpha overall level of 5%. The smallest treatment difference that is statistically significant will be an HR of 0.79.

MEDI4736 + tremelimumab versus SoC (OS in PD-L1 negative population)

Assuming the true OS average HR is 0.73 and the median OS in SoC chemotherapy alone arm is 10 months following an exponential distribution for both MEDI4736 + tremelimumab and SoC arm in the PD-L1 negative population, with approximately 194 patients, 182 OS events from the global cohort (approximately 94% maturity) will provide 57% power to demonstrate statistical significance at the 2-sided alpha overall level of 5%. The smallest treatment difference that is statistically significant will be an HR of 0.75.

Table 1 provides a summary of the statistical assumptions.

Table 1 Summary of statistical assumptions

Endpoint	Analysis population	Maturity	Alpha (%)	Power (%)	Average hazard ratio to detect
OS	bTMB \geq 20 population	87%	5	>90	0.49
OS	bTMB \geq 16 population	91%	5	>90	0.62
OS	bTMB \geq 12 population	92%	5	>90	0.65
OS	PD-L1 negative population	94%	5	57	0.73

OS: Overall survival

2. ANALYSIS SETS

2.1 Definition of analysis sets

2.1.1 Full analysis set (Intention to Treat (ITT))

The full analysis set (FAS) will include all randomized patients prior to the end of global recruitment. Any patients recruited in China, after global recruitment has ended, will not be included in the FAS (see Section 6 of the CSP). Efficacy analyses will be conducted in the FAS and in subsets of the FAS as described in this and following sections. Treatment groups will be compared on the basis of randomized IP, regardless of the treatment actually received. Patients who were randomized but did not subsequently go on to receive IP are included in the analysis in the treatment group to which they were randomized.

2.1.2 bTMB ≥ 20 mut/Mb analysis set

The bTMB ≥ 20 mut/Mb analysis set will include the subset of patients in the FAS whose blood TMB status is ≥ 20 mut/Mb at baseline as defined by the CCI [REDACTED] bTMB ≥ 16 mut/Mb analysis set

2.1.3 bTMB ≥ 16 mut/Mb analysis set

The bTMB ≥ 16 mut/Mb analysis set will include the subset of patients in the FAS whose blood TMB status is ≥ 16 mut/Mb at baseline as defined by CCI [REDACTED].

2.1.4 bTMB ≥ 12 mut/Mb analysis set

The bTMB ≥ 12 mut/Mb analysis set will include the subset of patients in the FAS whose blood TMB status is ≥ 12 mut/Mb at baseline as defined by CCI [REDACTED].

2.1.5 bTMB non-evaluable analysis set

The bTMB non-evaluable analysis set will include the subset of patients in the FAS

- (1) whose blood TMB status (TMB-high, TMB-low) at baseline cannot be determined by the CCI [REDACTED]. For these patients, assay results are either “failed” or “not done”.
- (2) whose sample for evaluation of bTMB is not available

2.1.6 bTMB evaluable analysis set

The bTMB-evaluable analysis set will consist of all subjects in the FAS who have a valid CCI [REDACTED] (TMB-high, TMBlow).

2.1.7 bTMB no-call analysis set

The bTMB no-call analysis set will consist of subjects in the FAS whose blood TMB status at baseline has been determined by the CCI [REDACTED] as a “no-call” for bTMB.

2.1.8 bTMB < 20 mut/Mb analysis set

The bTMB < 20 mut/Mb analysis set will include the subset of patients in the FAS whose blood TMB status is < 20 mut/Mb at baseline as defined by the CCI [REDACTED].

2.1.9 PD-L1-negative analysis set

The PD-L1-negative population analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1-negative as defined by the CCI [REDACTED] (i.e., <1% PD-L1-membrane expression in tumoral tissue).

2.1.10 PD-L1-based analysis sets

- Positive25%:- $\geq 25\%$ PD-L1 membrane-expression in tumoral tissue.
- Positive50%:- $\geq 50\%$ PD-L1 membrane-expression in tumoral tissue.

2.1.11 tTMB - based analysis sets

Tissue TMB-based analysis sets will be defined by the cutoff points 8, 10, 12 and 14. The analysis sets will be tTMB ≥ 8 mut/Mb, tTMB ≥ 10 mut/Mb, tTMB ≥ 12 mut/Mb, tTMB ≥ 14 mut/Mb. The tTMB evaluable analysis set, tTMB non-evaluable analysis set and tTMB no-call analysis set are defined in the same way as for bTMB, except applied to the results of the assay for tTMB. The analysis sets tTMB < 8 mut/Mb, tTMB < 10 mut/Mb, tTMB < 12 mut/Mb, tTMB < 14 mut/Mb will be used for subgroup analyses.

2.1.12 Safety analysis set

The safety analysis set (SAS) will consist of patients recruited prior to the end of global recruitment who received at least 1 dose of IP. Any patients recruited in China, after global recruitment has ended, will not be included in the safety analysis set (see Section 6 of the CSP). Safety data will not be formally analyzed but summarized using the SAS, according to the treatment received, that is, erroneously treated patients (e.g., those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received. If a patient received any dose of either MEDI4736 or tremelimumab, they will be reported under the MEDI4736 +tremelimumab combination arm.

Note that for analysis sets defined in 2.1.2 – 2.1.9, the corresponding safety analysis set includes all Safety analysis set subjects who are also in the respective analysis set. As examples, some specific safety analysis subset definitions have been specified in detail below.

2.1.13 bTMB ≥ 20 mut/Mb safety analysis set

The bTMB ≥ 20 mut/Mb safety analysis set will include the subset of patients in the SAS whose blood TMB status is ≥ 20 mut/Mb as defined by the CCI [REDACTED].

2.1.14 bTMB ≥ 16 mut/Mb safety analysis set

The bTMB ≥ 16 mut/Mb safety analysis set will include the subset of patients in the SAS whose blood TMB status is ≥ 16 mut/Mb as defined by the CCI [REDACTED].

2.1.15 bTMB \geq 12 mut/Mb safety analysis set

The bTMB \geq 12 mut/Mb safety analysis set will include the subset of patients in the SAS whose blood TMB status is \geq 12 mut/Mb as defined by the CCI [REDACTED].

2.1.16 PD-L1-negative safety analysis set

The PD-L1-negative safety analysis set will include the subset of patients in the SAS whose PD-L1 status is PD-L1-negative as defined by the CCI [REDACTED] (i.e., $<1\%$ PD-L1–membrane expression in tumoral tissue).

2.1.17 bTMB-evaluable safety analysis set

The bTMB-evaluable safety analysis set will consist of all subjects in the SAS who have a valid bTMB value.

2.1.18 Pharmacokinetic analysis set

All patients who received at least 1 dose of MEDI4736 or tremelimumab per the protocol for whom any post dose PK data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analysis will be included in the PK analysis set. The population will be defined by the Study Physician, Pharmacokineticist, and Statistician prior to any analyses being performed.

2.1.19 ADA evaluable set

Subjects in the safety analysis set with non-missing baseline ADA sample and at least 1 post-baseline ADA sample will be included in the ADA-evaluable set.

Definitions of the analysis sets for each outcome variable are provided in Table 2.

Table 2 Summary of outcome variables and analysis populations

Outcome variable	Population
Efficacy data	
OS	bTMB \geq 20 mut/Mb analysis set
OS	bTMB \geq 16 mut/Mb analysis set
OS	bTMB \geq 12 mut/Mb analysis set
OS	PD-L1-negative analysis set
OS	FAS (ITT population)
OS	bTMB $<$ 20 mut/Mb analysis set
OS	bTMB non-evaluable analysis set
OS	bTMB evaluable analysis set
OS	bTMB no-call analysis set

Outcome variable	Population
OS	PD-L1 \geq 25%
OS	PD-L1 \geq 50%
OS	tTMB \geq 8 mut/Mb analysis set
OS	tTMB \geq 10 mut/Mb analysis set
OS	tTMB \geq 12 mut/Mb analysis set
OS	tTMB \geq 14 mut/Mb analysis set
PFS, ORR	bTMB \geq 20 mut/Mb analysis set
PFS, ORR	bTMB \geq 16 mut/Mb analysis set
PFS, ORR	bTMB \geq 12 mut/Mb analysis set
PFS, ORR	PD-L1-negative analysis set
PFS, ORR	FAS (ITT population)
PFS, ORR	bTMB < 20 mut/Mb analysis set
PFS, ORR	bTMB-non-evaluable
PFS, ORR	PD-L1 \geq 25%
PFS, ORR	PD-L1 \geq 50%
PFS, ORR	tTMB \geq 8 mut/Mb analysis set
PFS, ORR	tTMB \geq 10 mut/Mb analysis set
PFS, ORR	tTMB \geq 12 mut/Mb analysis set
PFS, ORR	tTMB \geq 14 mut/Mb analysis set
APF12, OS12, OS18, OS24 and PFS2, DoR	bTMB \geq 20 mut/Mb analysis set
APF12, OS12, OS18, OS24 and PFS2, DoR	bTMB \geq 16 mut/Mb analysis sets
APF12, OS12, OS18, OS24 and PFS2, DoR	bTMB \geq 12 mut/Mb analysis set
APF12, OS12, OS18, OS24 and PFS2, DoR	PDL1-negative analysis set
APF12, OS12, OS18, OS24 and PFS2, DoR	FAS (ITT population)
DOR	bTMB < 20 mut/Mb analysis set
DOR	bTMB-non-evaluable
DOR	tTMB \geq 8 mut/Mb analysis set
DOR	tTMB \geq 10 mut/Mb analysis set
DOR	tTMB \geq 12 mut/Mb analysis set
DOR	tTMB \geq 14 mut/Mb analysis set

Outcome variable	Population
BOR	bTMB \geq 20 mut/Mb analysis set
BOR	bTMB \geq 16 mut/Mb analysis set
BOR	bTMB \geq 12 mut/Mb analysis set
BOR	PD-L1-negative analysis set
BOR	FAS (ITT population)
Demography	FAS (ITT population)
	bTMB \geq 20 mut/Mb analysis set
	bTMB \geq 16 mut/Mb analysis set
	bTMB \geq 12 mut/Mb analysis set
	PD-L1-negative analysis set
	bTMB no-call analysis set
	bTMB evaluable analysis set
	bTMB non-evaluable analysis set
	tTMB evaluable analysis set
	tTMB non-evaluable analysis set
PK data	PK analysis set
Immunogenicity data	ADA evaluable set
Safety Data	
Key Safety and Deaths	Safety Analysis Set
	bTMB \geq 20 mut/Mb safety analysis set
	bTMB \geq 16 mut/Mb safety analysis set
	bTMB \geq 12 mut/Mb safety analysis set
	PD-L1-negative safety analysis set
	bTMB < 20 mut/Mb safety analysis set
Exposure	Safety Analysis Set
	bTMB \geq 20 mut/Mb safety analysis set
AEs	Safety Analysis Set
	bTMB \geq 20 mut/Mb safety analysis set
Laboratory measurements	Safety Analysis Set
	bTMB \geq 20 mut/Mb safety analysis set
Vital signs	Safety Analysis Set
	bTMB \geq 20 mut/Mb safety analysis set
ECGs	Safety Analysis Set
	bTMB \geq 20 mut/Mb safety analysis set

FAS: Full analysis set

2.2 Violations and Deviations

The important protocol deviations will be listed and summarized by randomized 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 will be performed excluding patients with deviations that may affect the efficacy of the trial therapy if > 10% of patients within the analysis set:

- did not have the intended disease or indication or
- did not receive any randomized therapy.

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

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

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

- Deviation 1: Patients randomized but who did not receive study treatment.
- Deviation 2: Patients who deviate from key entry criteria as per the CSP. These are inclusion criteria 3, 4, 5 and exclusion criteria 3, 4, 8, 17.
- Deviation 3: Baseline RECIST scan > 42 days before date of randomisation.
- Deviation 4: No baseline RECIST 1.1 assessment on or before date of first dose.
- Deviation 5: Received prohibited concomitant medications (including other anti-cancer agents). Please refer to the Clinical Study Protocol (CSP) section 7.7 for those medications that are detailed as being ‘excluded’ from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to database lock to identify those likely to have an impact on efficacy.
- Deviation 6: Patients randomized who received treatment other than that to which they were randomized.

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

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

Misrandomisations in terms of errors in treatment dispensing, in addition to incorrect stratifications, will also be summarized and listed separately to the important protocol deviations. A misrandomisation is when a patient is not randomized or treated according to the randomisation schedule. It is envisaged that there will be 2 sub categories of this:

- Patients who receive no treatment whatsoever for a period of time due to errors in dispensing of medication. Note, this is not due to tolerability issues where patients may stop taking drug.
- The patient receives a treatment pack with a different code to their randomisation code. However, the actual treatment may still match the randomized treatment. For example, a patient is given randomisation code 0001, which according to the randomisation schedule is MEDI4736. However, at the randomisation visit they are given treatment pack 0003, but this still contains MEDI4736.

The summary will include all patients with a dispensing error but will also include information on how many of those patients received at least one dose of the wrong treatment at any time. Patients who receive the wrong treatment at any time will be included in the safety analysis set as described in Section 2.1. During the study, decisions on how to handle errors in treatment dispensing 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 details in Appendix D in the CSP) tumor response data will be used to determine each patient's visit response. It will also be used to determine if and when a patient has progressed and also their best objective response.

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

progression per RECIST 1.1 (irrespective of the reason for stopping treatment or subsequent therapy). If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits.

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

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

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

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

Following confirmed progression, patients should continue to be followed up for survival every 1 to 2 months as outlined in the study plan (Table 4 in the CSP). Exceptions are patients with confirmed PD who continue to receive IP at the discretion of the Investigator (after consultation with AstraZeneca) and who meet the criteria for treatment in the setting of PD; these patients will have scans for RECIST 1.1 assessments every 6 weeks (relative to the date of randomization per Table 2 and Table 3 in the CSP) for the first 48 weeks of treatment and then every 8 weeks until disease progression. Subsequent anticancer therapy information will be collected at the timepoints indicated in Table 4 in the CSP.

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

3.1.1 Investigator RECIST 1.1-based assessments: Target lesions

At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, or PD depending on the status of their disease compared with baseline and previous

assessments. Baseline will be assessed within the 28 days prior to randomization. If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of not evaluable (NE; unless there is evidence of progression in which case the response will be assigned as PD).

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (except lymph nodes which must have short axis ≥ 15 mm) with CT or MRI and which is suitable for accurate repeated measurements.

A patient can have a maximum of 5 measurable lesions recorded at baseline with a maximum of 2 lesions per organ (representative of all lesions involved suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). If more than one baseline scan is recorded then measurements from the one that is closest and prior to the date of randomisation will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

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

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

Table 3 TL visit responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all TLs. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10 mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive disease (PD)	A $\geq 20\%$ increase in the sum of diameters of TLs and an absolute increase of ≥ 5 mm, 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

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 TLs are recorded at baseline

Rounding of TL data

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

Missing TL data

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

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

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

Lymph nodes

For lymph nodes, if the size reduces to < 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 0mm 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. 0mm or < 10mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node LD increases by 20% but remains < 10mm.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response will be set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.
- Step 3: If not all lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD.
- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR.

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure a value of 5mm will be entered into the database and used in TL calculations, unless the radiologist has indicated and entered a smaller value that can be reliably measured. If a TL response of PD results then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

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

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

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

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and if $\leq 1/3$ of the TLs have missing measurements then scale up as described in the ‘Scaling’ section below. If the scaling results in a visit response of PD then the subject would be assigned a TL response of PD.
- Step 3: If after both steps PD has not been assigned, then, if appropriate, a scaled sum of diameters will be calculated (as long as $\leq 1/3$ of the TLs have missing measurements), treating the lesion with intervention as missing, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10 mm for lymph nodes) and the lesions that have been subject to intervention also have a value of 0 recorded. If scaling up is not appropriate due to too few non-missing measurements then the visit response will be set as NE.

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

Scaling (applicable only for irradiated lesions/lesion intervention)

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

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

Example of scaling

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

Lesion 5 is missing at the follow-up visit.

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

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

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

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

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of TLs

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

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

3.1.2 Investigator RECIST 1.1-based assessments: Non-target lesions and new lesions

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

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

Table 4 NTL Visit Responses

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

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

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

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

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

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

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

If the question ‘Any new lesions since baseline’ 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 tumor assessments where possible until objective disease progression is observed.

3.1.3 Investigator RECIST 1.1-based assessments: Overall visit response

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

Table 5 Overall visit responses

Target lesions	Non-target lesions	New lesions	Overall visit response
CR	CR or NA	No (or NE)	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	Non-PD or NE or NA	No (or NE)	PR
SD	Non-PD or NE or NA	No (or NE)	SD
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NE	Non-PD or NE or NA	No (or NE)	NE
NA	CR	No (or NE)	CR
NA	Non-CR/Non-PD	No (or NE)	SD
NA	NE	No (or NE)	NE

3.2 Outcome Variables

This study will analyze the primary endpoint of OS. In addition, the analyses of the secondary endpoints of PFS, ORR, BoR, DoR, and APF12 will be based on Investigator tumor assessments according to RECIST 1.1. PFS2 will be defined by local clinical practice, and survival rate (OS12, OS18 and OS24) will also be analyzed.

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

3.2.1 Primary endpoint - overall survival

OS is defined as the time from the date of randomization until death due to any cause (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 electronic case report form (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). Kaplan-Meier estimates of OS12, OS18 and OS24 (survival rate, that is, proportion of patients alive at 12 months, 18 months and 24 months) will also be derived as secondary endpoints.

3.2.2 Secondary endpoints

3.2.2.1 Progression-free survival

PFS (per RECIST 1.1 using Investigator assessments) will be defined as the time from the date of randomization until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomized therapy or receives another anticancer therapy prior to progression (i.e., date of PFS event or censoring – date of randomization + 1). Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more missed or non-evaluable visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment. If the patient has no evaluable visits or does not have baseline data, they will be censored at date of randomization (Day 1) unless they die within 2 visits (2x6 weeks for tumor assessments + 2x7 days for visit window) of baseline, in which case they will be deemed to have had an event, with date of death as the event date.

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

3.2.2.2 Objective response rate

ORR (per RECIST 1.1 using Investigator assessments) is defined as the number (%) of patients with at least 1 visit response of CR or PR prior to disease progression. If any patients do not have measurable disease at baseline then the analysis of ORR will exclude these patients, so that the denominator is a subset of the FAS population who have measurable disease at baseline. Data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who go off treatment without progression, receive a subsequent therapy, and then respond will not be included as responders in the ORR.

3.2.2.3 Duration of response

DoR (per RECIST 1.1 using Investigator assessments) will be defined as the time from the date of first documented response (CR or PR) until the first date of documented progression or death in the absence of disease progression (i.e. date of PFS event or censoring - date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the RECIST 1.1 PFS endpoint. The denominator for DoR will be defined as described for ORR (see Section 3.2.2.2).

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.

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

3.2.2.4 Proportion of patients alive and progression free at 12 months from randomization

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

3.2.2.5 Time from randomization to second progression

Time from randomisation to second progression (PFS2) is defined as the time from the date of randomisation to the earliest of the progression events (subsequent to that used for the primary variable PFS) 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 3.2.2.1 for details). The date of second progression will be recorded by the Investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: objective radiological imaging, symptomatic progression or death. RECIST assessments will not be collected for assessment of PFS2. The date of the PFS2 assessment and investigator opinion of progression status (progressed or non-progressed) at each assessment will be recorded in the eCRF. Patients alive and for whom a second disease progression has not been observed should be censored at the last time known to be alive and without a second disease progression, that is, censored at the latest of the PFS or PFS2 assessment date if the patient has not had a second progression or death.

3.2.2.6 Best objective response

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in Appendix D in the CSP. It is the best response a patient has had following randomization during their time in the study up until RECIST progression or the last evaluable assessment in the absence of RECIST progression.

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

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 up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorisation of BoR will be based on RECIST using the following response categories: CR, PR, SD, PD and NE.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 6 weeks minus 1 week, (i.e. at least 35 days (to allow for an early assessment within the assessment window), after randomisation (i.e. study day 36)). For CR/PR, the initial overall visit assessment which showed a response will use the latest of the dates contributing towards a particular overall visit assessment. The denominator will be consistent with that used in the ORR analysis.

For patients who die with no evaluable RECIST assessments, if the death occurs ≤ 90 days (ie, $2*(6 \text{ weeks} \pm 3 \text{ days})$) after randomization, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >90 days (i.e., $2*(6 \text{ weeks} \pm 3 \text{ days})$) after the date of randomization then BoR will be assigned to the NE category.

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

3.2.2.7 Change in tumor size

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

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

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

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

3.3 Safety

Safety and tolerability will be assessed in terms of adverse events (AEs) (including serious adverse events [SAEs]), deaths, laboratory data, vital signs, electrocardiograms (ECGs) and exposure. These will be collected for all patients. Safety data will be summarized from the treatment period for the immunotherapy agents alongside the SOC agents. Initial treatment and retreatment will generally be combined into one treatment period. Safety data from the retreatment period for the MEDI4736 + tremelimumab group may also be summarized via a small set of headline summaries should there be sufficient number of patients retreated to warrant this. Any safety summaries representing the retreatment 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 retreatment period.

Data from the treatment period on the immunotherapy agents (MEDI4736 + tremelimumab) will be compared against SoC in the main presentations of safety data and safety data from the retreatment period may also be summarized 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 (i.e., the last dose of MEDI4736 +/- tremelimumab) overall and between date of start dose and 30 days following last dose of the SoC agents. Note that for one version of the safety outputs the period of time after the administration of subsequent therapy will not be considered ‘on treatment’ (see Section 4.2.12 for further information).

All AEs will be listed however only TEAEs will be summarized

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

3.3.1 Adverse events (AEs)

AEs and SAEs will be collected from the time of the patient signing the informed consent form until the follow-up period is completed (90 days after the last dose of MEDI4736 +/- tremelimumab and 30 days after last dose of SoC). AEs and SAEs collected prior to randomization will be reported as pre-randomization AEs and SAEs. If an event that starts post the defined safety follow up period noted above is considered to be due to a late onset toxicity to study drug then it should be reported as an AE or SAE as applicable.

The Medical Dictionary for Regulatory Activities (MedDRA) (MedDRA version 21.1) will be used to code the AEs. A decision may be made to use a previous MedDRA version for consistency across similar studies. 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 (i.e. the last dose of MEDI4736 +/- tremelimumab) or 30 days after the last dose of the SoC agents. For the MEDI4736 + tremelimumab arm, in the 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 each treatment group a causality of related or missing for either component (MEDI4736 or tremelimumab or SoC) will be taken as related to study drug.

AEs of special interest

An adverse event of special interest (AESI) is one of scientific and medical interest specific to understanding of the Investigational Product and may require close monitoring and rapid communication by the investigator to the sponsor. An AESI may be serious or non-serious. The rapid reporting of AESIs allows ongoing surveillance of these events in order to characterize and understand them in association with the use of this investigational product.

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

Some clinical concepts (including some selected individual preferred terms and higher level terms) have been considered “AEs of special interest” (AESI) to the MEDI4736 program. These AESIs have been identified as Pneumonitis, Hepatic events, Diarrhea/Colitis, intestinal perforations, adrenal insufficiency, Type 1 diabetes mellitus, hyperthyroid events, hypophysitis, hypothyroid events, thyroiditis, renal events, dermatitis/rash, pancreatic events, myocarditis, myasthenia gravis, Guillain-Barre syndrome, myositis, infusion/hypersensitivity reaction, other rare/miscellaneous events. 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.

3.3.2 Treatment exposure

Exposure will be defined with the initial treatment period and the retreatment period for the MEDI4736 + tremelimumab group combined as described below.

- Total (or intended) exposure is defined as:

The total treatment period from first dose date of study drug to the earliest of “last dose date of study drug + 27 days” or death date or DCO.

Actual exposure of 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 summarized by combining the SoC treatments together. Actual exposure will be calculated for SoC also.

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

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

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

Total (or intended) exposure of Gemcitabine

Total (or intended) exposure is defined as the total treatment period from first dose date of study drug (Day 1) to the earliest of “last dose date of study drug” + 20 days or death date or DCO.

Dose reductions are not permitted per Section 6.8 of the CSP for the immunotherapy agents (MEDI4736 + tremelimumab). The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Exposure will also be measured by the number of cycles received. For all five choices of SoC regimen, a cycle corresponds to a period of 21 days, but for each immunotherapy agents a cycle corresponds to one dose of treatment. If a cycle is prolonged due to toxicity, this should still be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

Calculation of duration of dose delays (for actual exposure), MEDI4736 and tremelimumab:

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

Patients who permanently discontinue during a dose delay

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

3.3.3 Dose intensity

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

Relative dose intensity (RDI) will be defined as follows for MEDI4736 and tremelimumab:

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

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

3.3.4 Laboratory data

Laboratory data will be collected throughout the study, from screening to the follow-up visits as described in the CSP. Blood and urine samples for determination of haematology, clinical chemistry, and urinalysis will be collected as described in Section 5.2.1 of the CSP. For the definition of baseline and the derivation of post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 3.3.8 below will be used.

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

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

Corrected calcium (mmol/L) = Total calcium (mmol/L) + $([40 - \text{albumin (G/L)}] \times 0.02)$

Males: $(140 - \text{age in years at baseline}) * \text{weight in kg at visit} / (72 * \text{Creatinine in mg/dL at visit})$

Females: $0.85 * (140 - \text{age in years at baseline}) * \text{weight in kg at visit} / (72 * \text{Creatinine in mg/dL at visit})$

Use rules as in Section 3.3.7 for selection if multiple values available. If creatinine values are available in units different from mg/dL then convert to mg/dL (e.g. if in $\mu\text{mol/L}$ then divide by 88.4). If weight is missing at visit then last available weight will be used.

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

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

Local reference ranges will be used for the primary interpretation of laboratory data at the local laboratory. Project reference ranges will be used throughout for reporting purposes. The denominator used in laboratory summaries of CTC grades will only include evaluable patients, in other words those who had sufficient data to have the possibility of an abnormality.

For example:

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

3.3.5 ECGs

ECG data, available at screening and obtained as clinically indicated thereafter 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 3.3.8 below will be used. Categorical summaries of change from baseline in overall ECG assessments (recorded as “abnormal” and “normal”) will be created if sufficient number of ECG assessments are recorded.

At each time point the Investigator’s assessment of the ECG will be collected locally. Heart rate, duration of QRS complex, RR, PR and QT intervals will be collected centrally via a digital read. This digital copy of all ECGs will be held centrally by a central ECG provider, and the data from this review will be stored for analysis if necessary at the end of the study. If it is necessary to analyze this data then QTcF (Fridericia) will be calculated programmatically using the reported ECG values (RR and QT).

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

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

3.3.6 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 3.3.8 below will be used.

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

3.3.7 General considerations for safety assessments

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

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarized. Inclusion within the time window should be based on the actual date 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 + tremelimumab combination therapy (with 4 weeks between scheduled assessments) are:

Day 29, visit window 2 – 42

Day 57, visit window 43 – 70

Day 85, visit window 71 – 98

Day 113, visit window 99 – 126

Day 141, visit window 127 – 154

Day 169, visit window 155 – 182

Day 197, visit window 183 – 210

Day 225, visit window 211 – 238

Day 253, visit window 239 – 266

Day 281, visit window 267 – 294

Day 309, visit window 295 – 322

Day 337, visit window 323 – 350

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

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).
- Listings should display all values contributing to a time point for a patient.
- For visit based summaries:

- If there is more than one value per patient within a time window then the closest value to the scheduled visit date should be summarized, or the earlier in the event the values are equidistant from the nominal visit date. If there are two values recorded on the same day and the parameter is CTCAE gradable then the record with the highest toxicity grade should be used. Alternatively, if there are two records recorded on the same day and the toxicity grade is the same (or is not calculated for the parameter) then the average of the two records should be used. The listings should highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all 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 summarized 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 retreatment period for the MEDI4736 + tremelimumab group, then baseline is similarly defined as the last non-missing measurement prior to the first dose on the retreatment 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 (i.e. some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline as this is the most conservative. In the scenario where there are two assessments on day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline. Where safety data are summarized over time, study day will be calculated in relation to date of first treatment.

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.

3.4 Other Biomarker Variables

If applicable, the relationship of exploratory biomarkers to primary and secondary endpoints such as OS, PFS, ORR and DoR will be explored and presented for a subset of patients in the FAS population who are evaluable for each biomarker. This will be assessed using similar summary and graphical representations to those that are outlined for the efficacy outputs.

3.5 Pharmacokinetic and Immunogenicity variables

Analyses to evaluate the pharmacokinetics and immunogenicity of MEDI4736 and tremelimumab will be performed in collaboration between the reporting CRO (IQVIA) and AstraZeneca/MedImmune Clinical Pharmacology group or designee.

3.5.1 Population pharmacokinetics and exposure-response/safety analysis

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

3.5.2 Pharmacokinetic non-compartmental analysis

The PK analyses will be performed at AstraZeneca. The actual sampling times will be used in the PK calculations. PK concentration data and summary statistics will be tabulated. Individual and mean blood concentration-time profiles will be generated. PK parameters will be determined using standard non-compartmental methods. 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.

3.5.3 Immunogenicity analysis

Serum samples for ADA assessments will be conducted utilizing a tiered approach (screen, confirm, titer), and ADA data will be collected at scheduled visits shown in the CSP. ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titer will be reported as well. In addition, the presence of neutralizing antibody (nAb) will be tested for all ADA-positive samples using a ligand-binding assay. The nAb results will be reported as positive or negative. A patient is defined as being ADA-positive if a positive ADA result is available at any time, including baseline and all post-baseline measurements; otherwise ADA negative.

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

- ADA positive at any visit (at baseline and/or post-baseline). The percentage of these subjects in a population is known as ADA prevalence.

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

3.6 Health Resource Use

A payer analyses plan (PAP) will contain a description of analyses supplementary to this SAP, conducted in order to support payer reimbursement submissions.

4. ANALYSIS METHODS

4.1 General principles

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

- H0: No difference between MEDI4736 + tremelimumab combination therapy and SoC
- H1: Difference between MEDI4736 + tremelimumab combination therapy and SoC

The primary endpoint is OS in patients with bTMB ≥ 20 mut/Mb. The analysis will be performed when there is approximately 87% maturity in the bTMB ≥ 20 mut/Mb analysis set.

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

- Descriptive statistics will be used for all variables, as appropriate, and will be presented by treatment group. Continuous variables will be summarized by the number of observations (n), mean, standard deviation, median, minimum, and maximum. For log transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables will be summarized by frequency counts and percentages for each category.
- Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment arm.

- For continuous data the mean, geometric mean, CV, geometric CV 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, maximum and n will be displayed with the same accuracy as the original data.
- For categorical data, percentages will be rounded to 1 decimal place.
- All summaries will be presented in tabular format by treatment group, unless otherwise specified. Data will be presented in data listings by treatment group and patient number.
- A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days.
- P-values will be rounded to 4 decimal places. P-values less than 0.00005 (e.g. 0.00002) will not be rounded to 4 decimal places (e.g. 0.0000) but instead be displayed as <0.0001. P-values output as <0.0001 by statistical software will not be rounded and will be displayed in the same way ('<0.0001').
- SAS® version 9.4 will be used for all analyses.

Baseline will be the last assessment of the variable under consideration prior to the intake of the first dose of IP, except for efficacy variables. For efficacy variables, see definition of efficacy endpoints (Sections 4.2.3 – 4.2.9). All data collected will be listed. Disposition, baseline characteristics and IVRS stratification factors data will be summarised based on the bTMB ≥ 20 mut/Mb and FAS analysis sets. Demography and efficacy data will be summarized and analyzed based on the bTMB ≥ 20 mut/Mb, bTMB ≥ 16 mut/Mb, bTMB ≥ 12 mut/Mb, FAS, the PD-L1-negative analysis sets, and bTMB-evaluable and bTMB non-evaluable analysis sets. In addition selected efficacy data will also be summarized based on the PDL1-positive_{25%}, the PDL1-positive_{50%}, bTMB no-call, bTMB <20 mut/Mb, bTMB non-evaluable, tTMB ≥ 16 mut/Mb, tTMB ≥ 12 mut/Mb, tTMB ≥ 10 mut/Mb and tTMB ≥ 8 mut/Mb analysis sets. PK data will be summarized and analyzed based on the PK analysis set. Safety and treatment exposure data will be summarized based on the Safety analysis set and where appropriate, for the subsets of the Safety analysis set for patients within the bTMB ≥ 20 mut/Mb, bTMB ≥ 16 mut/Mb, bTMB ≥ 12 mut/Mb, FAS, the PD-L1-evaluable and the PD-L1-negative analysis sets. ADA data will be summarised based on the bTMB ≥ 20 mut/Mb and FAS analysis sets.

As felt essential to understanding of efficacy and safety, outputs may also be summarized by treatment group for bTMB ≥ 20 mut/Mb, bTMB ≥ 16 mut/Mb, bTMB ≥ 12 mut/Mb, FAS and the PD-L1-negative analysis sets, and also within the PDL1-positive_{25%} and the PDL1-positive_{50%}, bTMB <20 mut/Mb, non-evaluable bTMB population and tTMB analysis sets. Safety data will be summarized from the treatment period for the immunotherapy agents alongside the SoC agents. Initial treatment and re-treatment will generally be combined into one treatment period.

4.2 Analysis methods

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

The following table (Table 6) details which endpoints are to be subjected to formal statistical analysis, together with pre-planned sensitivity analyses, making it clear which analysis is regarded as sensitivity analysis for that endpoint. Note, all endpoints compare MEDI4736 + tremelimumab combination therapy versus SoC.

Table 6 Overview of Main Pre-planned statistical and sensitivity analysis to be conducted

Endpoints analyzed	Notes
OS	Unstratified Log-rank test for Primary analysis in: <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set Secondary analysis in: <ul style="list-style-type: none"> • bTMB \geq 16 mut/Mb analysis set • bTMB \geq 12 mut/Mb analysis set • bTMB < 20 mut/Mb analysis set • bTMB non-evaluable analysis set • bTMB evaluable analysis set • bTMB no-call analysis set • tTMB \geq 14 mut/Mb analysis set • tTMB \geq 12 mut/Mb analysis set • tTMB \geq 10 mut/Mb analysis set • tTMB \geq 8 mut/Mb analysis set Stratified Log-rank test for: <ul style="list-style-type: none"> • FAS (ITT population) • PD-L1-negative analysis set • PD-L1 \geq 25% analysis set • PD-L1 \geq 50% analysis set Sensitivity analysis: max-combo test for: <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set

Endpoints analyzed	Notes
OS12, OS18 and OS24	<p>Kaplan-Meier estimates and 95% confidence intervals of survival rate at 12 months, 18 months and 24 months (following method described by Glimm et al 2010)</p> <p>Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med 2010;29:219-28.</p> <p>Klein et al 2007) in:</p> <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set • bTMB \geq 16 mut/Mb analysis set • bTMB \geq 12 mut/Mb analysis set • bTMB $<$ 20 mut/Mb analysis set • bTMB non-evaluable analysis set • bTMB evaluable analysis set • bTMB no-call analysis set • tTMB \geq 14 mut/Mb analysis set • tTMB \geq 12 mut/Mb analysis set • tTMB \geq 10 mut/Mb analysis set • tTMB \geq 8 mut/Mb analysis set • FAS (ITT population) • PD-L1-negative analysis set • PD-L1 \geq 25% analysis set • PD-L1 \geq 50% analysis set

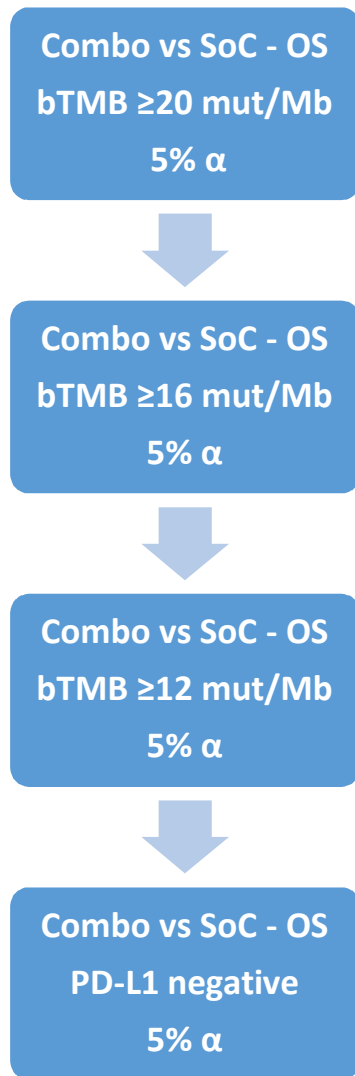
Endpoints analyzed	Notes
PFS	<p><u>Log-rank tests for:</u></p> <p>Main analyses: Investigator RECIST 1.1 assessments in:</p> <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set • bTMB \geq 16 mut/Mb analysis set • bTMB \geq 12 mut/Mb analysis set • PDL1-negative analysis set • bTMB $<$ 20 mut/Mb analysis set • bTMB-non-evaluable analysis set • FAS (ITT population) • PD-L1 \geq 25% • PD-L1 \geq 50% • tTMB \geq 14 mut/Mb analysis set • tTMB \geq 12 mut/Mb analysis set • tTMB \geq 10 mut/Mb analysis set • tTMB \geq 8 mut/Mb analysis sets <p>Sensitivity analysis: Investigator RECIST 1.1 modified for confirmation of progression in bTMB \geq 20 mut/Mb analysis set</p>
ORR	<p><u>Logistic regression</u> using Investigator RECIST 1.1 assessments adjusting for same factors as the primary endpoint in:</p> <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set • bTMB \geq 16 mut/Mb analysis set • bTMB \geq 12 mut/Mb analysis set • PDL1-negative analysis set • bTMB $<$ 20 mut/Mb analysis set • bTMB-non-evaluable analysis set • FAS (ITT population) • PD-L1 \geq 25% • PD-L1 \geq 50% • tTMB \geq 14 mut/Mb analysis set • tTMB \geq 12 mut/Mb analysis set • tTMB \geq 10 mut/Mb analysis set • tTMB \geq 8 mut/Mb analysis sets

Endpoints analyzed	Notes
DoR	<p>Analysis methods as described by Ellis et al 2008 using Investigator RECIST 1.1 assessments in:</p> <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set • bTMB \geq 16 mut/Mb analysis set • bTMB \geq 12 mut/Mb analysis set • PDL1-negative analysis set • FAS (ITT population) • PDL1-negative analysis set • bTMB < 20 mut/Mb analysis set • bTMB-non-evaluable analysis set • PD-L1 \geq 25% • PD-L1 \geq 50% • tTMB \geq 14 mut/Mb analysis set • tTMB \geq 12 mut/Mb analysis set • tTMB \geq 10 mut/Mb analysis set • tTMB \geq 8 mut/Mb analysis sets
APF12	<p>Use the Kaplan-Meier estimates of PFS at 12 months (following method described by Glimm et al 2010 Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med 2010;29:219-28. Klein et al 2007) in:</p> <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set • bTMB \geq 16 mut/Mb analysis set • bTMB \geq 12 mut/Mb analysis set • PDL1-negative analysis set • FAS (ITT population)
PFS2	<p><u>Log-rank test</u> in:</p> <ul style="list-style-type: none"> • bTMB \geq 20 mut/Mb analysis set • bTMB \geq 16 mut/Mb analysis set • bTMB \geq 12 mut/Mb analysis set • PDL1-negative analysis set • FAS (ITT population)

4.2.1 Multiple testing strategy

In order to strongly control the type I error at two-sided 5%, a multiple testing procedure (MTP) with gatekeeping strategy will be used across the primary endpoint (OS) in the bTMB ≥ 20 mut/Mb populations and secondary analyses of OS in bTMB ≥ 16 mut/Mb, bTMB ≥ 12 mut/Mb, OS in PD-L1-negative analysis sets. If the higher level hypothesis in the MTP is rejected for superiority, the following hypothesis will then be tested in the MTP as shown in Figure 3.

Figure 3 Multiple testing procedure for controlling the Type I error rate



If the analysis indicates superiority in OS for the bTMB ≥ 20 mut/Mb analysis set then subsequent analysis of secondary OS endpoints will be performed in accordance with the hierarchical testing strategy in the analysis sets bTMB ≥ 16 mut/Mb, bTMB ≥ 12 mut/Mb,

PD-L1-negative. If the null hypothesis is not rejected, then results will be presented as secondary analyses with nominal p-values. The above testing procedure will ensure strong control of the family-wise error rate (Glimm et al 2010).

4.2.2 Analysis of primary endpoint - overall survival

The primary OS analysis will be performed in the bTMB ≥ 20 mut/Mb analysis set. An unstratified log-rank test will be performed to study the effect of MEDI4736 + tremelimumab combination therapy versus SoC.

The unstratified secondary OS analysis will also be performed in the bTMB ≥ 16 mut/Mb analysis set, bTMB ≥ 12 mut/Mb analysis set, bTMB < 20 mut/Mb analysis set, bTMB non-evaluable analysis set, bTMB evaluable analysis set, bTMB no-call analysis set, tTMB ≥ 8 mut/Mb analysis set, tTMB ≥ 10 analysis set, tTMB ≥ 12 analysis set and tTMB ≥ 14 .

The stratified secondary OS analysis will also be performed in the PDL1-negative analysis set, FAS (ITT population),

Analyses in the FAS (ITT population) will be performed using a stratified log-rank test adjusting for PD-L1 tumor expression ($\geq 25\%$ versus $< 25\%$), histology (squamous versus non-squamous), and smoking status (never smoker versus ever smoker). The effect of MEDI4736 + tremelimumab combination therapy versus SoC treatment will be estimated by the HR together with its appropriately sized CI and p-value. Analyses in the PDL1-negative analysis set, and PD-L1-positive_{25%}, PD-L1-positive_{50%} analysis sets will be performed using a stratified log-rank test adjusting for histology (squamous versus non-squamous), and smoking status (never smoker versus ever smoker).

The HR and its CI will be estimated from the Cox proportional hazards model (Cox 1972) (with Efron method to handle ties and the stratification variables, if included in the strata statement, to be applied as specified in the previous paragraph) and the CI calculated using a profile likelihood approach. The covariates in the statistical modelling will be based on the values entered into IVRS at randomization, even if it is subsequently discovered that these values were incorrect.

Kaplan-Meier plots of OS will be presented by treatment arm, and by treatment arm and bTMB subgroup, PD-L1 tumor status subgroup, or tTMB subgroup as appropriate. Summaries of the number and percentage of patients who have died, who are still in survival follow-up, who are lost to follow-up, and who have withdrawn consent will be provided along with the median OS, and 95% confidence intervals for the median OS, for each treatment. The effect of treatment will be estimated by the HR together with its appropriately sized CI and p-value.

Kaplan Meier plots of OS using reverse censoring indicator, by treatment arm based on the bTMB ≥ 20 mut/Mb analysis set will also be presented.

Proportional hazards will be tested by examining plots of complementary log-log (event

times) versus log (time) based on the bTMB \geq 20 mut/Mb analysis set.

The max-combo test will be conducted as a sensitivity analysis on the OS data in the primary analysis set bTMB \geq 20 mut/Mb, to test for treatment differences in the case of nonproportional hazards. The analysis will be based on adaptive procedure involving selection of best test statistics with log-rank (G0,0) and the Fleming-Harrington (FH) test (G0,1, G1,0, and G1,1) with alpha correction (Duke-Margolis, 2018) [Duke-Margolis 2018](#)

Multiple Imputation method in handling missing data

To further assess impact on of missing bTMB data and to evaluate analysis robustness in primary analysis (bTMB \geq 20 population), multiple imputation methods may be performed to estimate the expected treatment effect in the bTMB high population identified. The purpose of this analysis is to perform the supportive sensitivity analysis given various assumptions on the missing data (MAR (missing at random) and MNAR (missing not at random)).

The missing bTMB values will be imputed using predictive models containing all potential prognostic baseline covariates and relevant clinical outcome and the OS analysis will be performed based on primary analysis method to each imputed datasets and the results will be combined, a OS HR as well as confidence interval will be presented as supportive sensitivity analysis.

Additional sensitivity analyses may be conducted on the primary endpoint by conducting analyses on subsets of tTMB analysis sets defined in Section 2.1.11, where subjects with low tissue volume in the baseline sample tested for tTMB are excluded from the analysis set.

Subgroup analyses

Subgroup analysis will be conducted comparing OS between MEDI4736 + tremelimumab combination therapy versus SoC in the following subgroups of primary population (TMB \geq 20 mut/Mb) and FAS analysis sets as deemed appropriate (but not limited to):

- Sex (male versus female)
- Age at randomization (<65 versus \geq 65 years of age)
- Histology (squamous versus non-squamous)
- Smoking status (never smoker versus ever smoker)
- Race (Asian versus non-Asian)
- Subsets of FAS only: bTMB (\geq 12 versus < 12 mut/Mb, \geq 16 versus < 16 mut/Mb, \geq 20 versus < 20 mut/Mb), bTMB evaluable vs non-evaluable analysis sets vs bTMB no-call analysis set.

- Subsets of FAS only: tTMB (≥ 8 versus < 8 mut/Mb, ≥ 10 versus < 10 mut/Mb, ≥ 12 versus < 12 mut/Mb, ≥ 14 versus < 14 mut/Mb), tTMB evaluable vs non-evaluable analysis sets vs tTMB no-call analysis set.
- Subsets of FAS only: PD-L1 using cutpoints of 1%, 10%, 25% and 50% tumor expression ($\geq 1\%$ versus $< 1\%$, $< 10\%$ versus $\geq 10\%$, $\geq 25\%$ versus $< 25\%$ and $< 50\%$ versus $\geq 50\%$) and immune cell expression (IC $\geq 25\%$ versus $< 25\%$)
- Subsets of FAS only: PD-L1 and bTMB subgroups combined: bTMB ≥ 20 and PD-L1 $\geq 1\%$, bTMB ≥ 20 and PD-L1 $< 1\%$, bTMB < 20 and PD-L1 $\geq 1\%$, bTMB < 20 and PD-L1 $< 1\%$.
- Subsets of FAS only: PD-L1 and bTMB subgroups combined: bTMB ≥ 20 and PD-L1 $\geq 25\%$, bTMB ≥ 20 and PD-L1 $< 25\%$, bTMB < 20 and PD-L1 $\geq 25\%$, bTMB < 20 and PD-L1 $< 25\%$.
- Subsets of FAS only: PD-L1 and bTMB subgroups combined: bTMB ≥ 16 and PD-L1 $\geq 1\%$, bTMB ≥ 16 and PD-L1 $< 1\%$, bTMB < 16 and PD-L1 $\geq 1\%$, bTMB < 16 and PD-L1 $< 1\%$.
- Subsets of FAS only: PD-L1 and bTMB subgroups combined: bTMB ≥ 16 and PD-L1 $\geq 25\%$, bTMB ≥ 16 and PD-L1 $< 25\%$, bTMB < 16 and PD-L1 $\geq 25\%$, bTMB < 16 and PD-L1 $< 25\%$.
- ECOG (0 versus ≥ 1)
- Baseline Liver Metastases (Yes versus No)

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

No adjustment to the significance level for testing of the subgroup and sensitivity analysis will be made since all these analyses will be considered supportive of the analysis of OS; all p-values obtained from subgroup analyses will be considered nominal

Cox proportional hazards modelling will be employed in the bTMB ≥ 20 mut/Mb to assess the effect of subgroup factors on the HR estimate; a model will be constructed, containing treatment alone, to ensure that any output from the Cox modelling is likely to be consistent with the results of the log-rank test. Additionally, for each subgroup, the HR (MEDI4736 + tremelimumab combination therapy: SoC) and 95% CI will be calculated from a single model that contains treatment and subgroup factor. These will be presented on a forest plot. The subgroup analyses for the stratification factors will be based on the values entered into the interactive voice response system (IVRS), even if it is subsequently discovered that these values were incorrect, all other factors will be based on values recorded on the eCRF as indicated above.

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 analyzed. In this case, only descriptive summaries will be provided.

Exploratory: Effect of covariates on the HR estimate

Cox proportional hazards modeling will be employed to assess the effect of covariates on the HR estimate. The result from the initial model and the model containing additional covariates will be presented.

Additional covariates for this model will include: age at randomization, histology, smoking status, race, and number of target lesions at baseline (1,2,>=3). If feasible, all covariates will be entered into the model, but if not, a most important subset of variables will be determined by a standard variable-selection process.

This multivariate Cox model will be performed in bTMB \geq 20 mut/Mb population. The stratification factor PD-L1 status (<25% versus \geq 25%) will not be included in this analysis. This analysis evaluates the treatment effect adjusting for any potential imbalances in baseline prognostic factors.

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

Any quantitative interactions identified using this procedure will then be tested to rule out any qualitative interaction using the approach of Gail and Simon 1985 (Gail and Simon 1985).

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

Exploratory analyses of OS adjusting for the impact of subsequent immunotherapy or other investigational treatment may be performed, if a sufficient proportion of patients switch. Methods such as Rank Preserving Structural Failure Time (Robins and Tsiatis 1991), Inverse Probability of Censoring Weighting (Robins 1993), the two-stage model, and other methods in development will be explored. The decision to adjust and the final choice of methods will be based on a blinded review of the data and the plausibility of the underlying assumptions. Baseline and time-dependent characteristics will be explored, and summaries of baseline characteristics will be summarized by treatment arm, splitting between those that have and have not switched at the time of the analysis. Further detail will be provided in the Payer Analysis Plan. These analyses are intended to support reimbursement appraisals.

4.2.3 Progression-free survival

The main PFS analysis will be based on the programmatically derived RECIST 1.1 using the Investigator tumor assessments in the bTMB \geq 20, bTMB \geq 16, bTMB \geq 12, and PDL1-

negative analysis sets, and analysis sets mentioned in Table 6 for the PFS endpoint. PFS will be analyzed using a log-rank test, using the same methodology described for the OS endpoint within the respective analysis sets. The effect of MEDI4736 + tremelimumab combination therapy versus SoC will be estimated by the HR together with its corresponding 95% CI and p-value.

Kaplan-Meier plots will be presented by treatment arm. Summaries of the number and percentage of patients experiencing a PFS event and the type of event (RECIST 1.1 or death) and censoring will be provided along with median PFS, and 95% confidence intervals for median PFS, for each treatment.

An exploratory analysis of PFS using Investigator assessment based on RECIST 1.1 modified for confirmation of progression will be performed in the bTMB₂₀ ≥ 20 mutations/Mb analysis set. The log-rank test used for the analysis of PFS will be repeated for the exploratory analysis.

Sensitivity analysis for the bTMB ≥ 20 mut/Mb analysis set will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled timepoints. The midpoint between the time of progression and the previous evaluable RECIST assessment will be analyzed using a log-rank test. For patients whose death was treated as PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust even in highly asymmetric assessment schedules ([Sun and Chen 2010](#)).

Attrition bias for the bTMB ≥ 20 mut/Mb analysis set will be assessed by repeating the PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following 2 or more non-evaluable tumor assessments will be included. In addition, patients who take subsequent therapy prior to progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier plot of the time to censoring where the censoring indicator of the PFS analysis is reversed in the bTMB ≥ 20 mut/Mb analysis set.

Subgroup analysis will be conducted comparing PFS (per RECIST 1.1 using Investigator assessments) between MEDI4736 + tremelimumab combination therapy versus SoC using the same methodology (but as deemed unnecessary, may not include all subgroups) as described for the primary endpoint of OS.

Exploratory analyses

PFS based on RECIST 1.1 modified for confirmation of progression will be performed for exploratory purposes using the algorithm described above for the RECIST 1.1 Investigator assessments, but following a modification whereby any objective disease progression must be confirmed by the next scheduled scan. The confirmatory scan must be no sooner than 4 weeks after the initial suspected progression. If disease progression is confirmed (or disease progression occurs and no further scans are recorded) then the date of progression will be

when it was originally observed. Patients with a single disease progression and no further tumor assessment scans will be treated as having PD in the analysis. In the absence of significant clinical deterioration, the investigational site is advised to continue the patient on their randomized MEDI4736 + tremelimumab combination therapy until progression has been confirmed. If progression is not confirmed, the patient should continue their randomized MEDI4736 + tremelimumab combination therapy and on-treatment assessments. Treatment through PD in the SoC group is at the Investigator's discretion; however, a confirmatory scan is required for all patients in the SoC group, even if a subsequent treatment is started.

Additional supportive summaries/analysis

In addition, for OS, the number of patients prematurely censored will be summarized by treatment arm together, based on the bTMB ≥ 20 mut/Mb analysis set. A patient would be defined as prematurely censored if their survival status was not defined at the DCO. 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 summarized using median time from randomisation to date of censoring (date last known to be alive) in censored patients only, presented by treatment group.

For PFS, days between RECIST assessments will be tabulated by treatment group, based on the bTMB ≥ 20 mut/Mb analysis set.

4.2.4 Objective response rate

The ORR will be based on the programmatically derived RECIST using the Investigator tumor data. The ORR will be compared between MEDI4736 + tremelimumab versus SoC using logistic regression models adjusting for the same factors as the primary endpoint (PD-L1 tumor expression, histology and smoking status), as appropriate. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favor MEDI4736+tremelimumab) together with its associated profile likelihood CI (e.g. using the option 'LRCI' in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). This analysis will be performed in the the bTMB ≥ 20 , bTMB ≥ 16 , bTMB ≥ 12 , and PDL1-negative analysis sets, and analysis sets mentioned in Table 6 for the ORR endpoint. The analysis of the patients in the PD-L1- based analysis sets will be performed using a logistic regression model adjusting for only histology and smoking status.

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.

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

Summaries will be produced that present the number and percentage of patients with a tumor response (CR/PR). Overall visit response data will be listed for all patients (i.e., the FAS).

For each treatment arm, best objective response (BoR) will be summarized by n (%) for each category (CR, PR, SD, PD, and NE). No formal statistical analyses are planned for BoR.

4.2.5 Duration of response

In order to analyze the DoR between MEDI4736 + tremelimumab and SoC, descriptive data will be provided for the DoR in responding patients (i.e. 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).

4.2.6 Survival rate at 12 months, 18 months, and 24 months

OS12, OS18 and OS24 and 95% confidence interval will be presented by treatment arm.

The proportion of patients alive at 12 months (OS12) will be defined as the Kaplan-Meier estimate of OS at 12 months, and similarly for OS18 and OS24.

4.2.7 Proportion of patients alive and progression free at 12 months

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

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

This analysis will be performed in the bTMB ≥ 20 , bTMB ≥ 16 , bTMB ≥ 12 , PDL1-negative analysis sets and the FAS and in the bTMB < 20 mut/Mb, bTMB non-evaluable analysis sets

4.2.8 Time from randomization to second progression

PFS2 will be analyzed using an unstratified log-rank tests for bTMB and tTMB based analysis sets, and using stratified log-rank tests in the FAS and PDL1-negative, using the same methodology as described for the OS endpoint in the respective analysis sets. The effect of MEDI4736 + tremelimumab combination therapy versus SoC will be estimated by the HR together with its corresponding CI and p-value. Kaplan-Meier plots will be presented by treatment arm. Summaries of the number and percentage of patients who have an event as well as who were censored will be provided along with the medians for each treatment.

4.2.9 Subsequent Therapy

A summary table of first subsequent therapies by treatment group will be provided, as well as response to first subsequent therapy by treatment group. For supportive purposes, the time to the start of subsequent therapy will be analyzed using the same methodology and model as for PFS. The HR for the treatment effect together with its 95% CI will be presented. In addition, a Kaplan-Meier plot of the time to the start of subsequent therapy will be presented by treatment

group and the time between progression and starting subsequent therapy will be assessed. No multiplicity adjustment will be applied as these are viewed as supportive endpoints. Subsequent therapies received after discontinuation of treatment will be summarized and listed by treatment group. Patients who subsequently received an immunotherapy agent or entered an immunotherapy trial will be summarized and listed by treatment arm..

Time from treatment discontinuation to start of subsequent therapy will be summarized. A summary table of subsequent therapies by treatment arm will be provided. This analysis will be performed in the bTMB ≥ 20 mut/Mb and FAS analysis sets; also in other analysis sets as deemed necessary.

Subsequent disease-related anti-cancer therapies received after discontinuation of study treatment will have summaries produced by treatment group.

4.2.10 Change in tumor size in the bTMB ≥ 20 mut/Mb analysis set

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

Tumor size will also be presented graphically using waterfall plots for each treatment arm, to present each subject's best percentage change in tumor size as a separate bar, with the bars ordered from the largest increase to the largest decrease. A reference line at the +20% and -30% change in TL tumor size level will be added to the plots, which corresponds with the definition of progression and 'partial' response, respectively. 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 Non-squamous) of each patient will be indicated. Additional waterfall plots showing percentage change in tumor size at specific timepoints may be produced if it is felt that these are warranted to provide greater clarity.

The above outputs will be programmed for the investigator RECIST assessments.

4.2.11 Healthcare resource use

An exploratory health economic analysis may be undertaken to examine the impact of disease and treatment on resource use to primarily support the economic evaluation of MEDI4736+tremelimumab in comparison to SoC, and will be outlined in the Payer Analysis Plan (PAP). This would include providing descriptive statistics as appropriate, including means, medians, and ranges.

4.2.12 Safety data

Safety and tolerability data will be presented by treatment group using the full safety

population and the subset of patients in the bTMB ≥ 20 mut/Mb safety analysis set. Key safety tables will also be presented for bTMB ≥ 16 mut/Mb safety analysis set, bTMB ≥ 12 mut/Mb safety analysis set, PDL1-negative safety analysis set.

Safety data will be summarized only. No formal statistical analyses will be performed on the safety data. Data from all cycles of treatment will be combined in the presentation of safety data. Initial treatment and retreatment will generally be combined into one treatment period. AEs (both in terms of MedDRA preferred terms and CTCAE grade) will be listed individually by patient. The number of patients experiencing each AE will be summarized by treatment arm and CTCAE grade. Additionally, data presentations of the rate of AEs per person-years at risk may be produced. Any safety summaries examining specifically the impact of retreatment with MEDI4736 + tremelimumab will be produced separately, if required.

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

The following sections describe the planned safety summaries. However, additional safety tables may be required to aid interpretation of the safety data. For example, if an imbalance is seen in AEs or laboratory abnormalities that could be due to the differential follow-up periods (showing up more of the background/disease related AEs/abnormalities), additional summaries may be produced using a 30 day follow up period for both treatment arms to further explore / explain.

Adverse Events

All AEs, both in terms of MedDRA v21.1 preferred term and CTCAE grade, will be summarized descriptively by count (n) and percentage (%) for each treatment group. The MedDRA v21.1 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 'pretreatment'. However, any AE occurring before the administration of the first dose on Study Day 1 that increases in severity after the first dose will be regarded as treatment emergent and thus will be included in the majority of summary tables.

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

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

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

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved) and investigator's assessment of severity and relationship to study drug. Frequencies and percentages of patients reporting each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from any on episode level summaries which may be produced).

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

- All AEs
- All AEs possibly related to study medication (as determined by the reporting investigator)
- AEs with CTCAE grade 3 or 4
- Adverse events of CTCAE maximum grade
- AEs with CTCAE grade 3 or 4, possibly related to study medication (as determined by the reporting investigator)
- AEs with outcome of death
- AEs with outcome of death possibly related to study medication (as determined by the reporting investigator)
- All SAEs

- All SAEs possibly related to study medication (as determined by the reporting investigator)
- SAEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication, possibly related to study medication (as determined by the reporting investigator)
- AEs leading to hospitalization
- AEs leading to dose delay of study medication
- Immune mediated AEs (as determined by the reporting investigator)
- Infections are included in AE/SAEs of this list as well as Infection pooled term analysis below, so deleted it here. However, AESI should be included. This list contains parameters that are not usually included in CSR, eg SAEs leading to d/c, AE leading to hospitalization. What is the rationale for inclusion here?
- Infusion reaction AEs (as determined by the reporting investigator)

An overall summary of the number and percentage of patients in each category will be presented. In addition, a truncated AE table of most common AEs and another table showing the most common AEs with CTCAE grade 3 or 4 – at least 1% in D+T or SoC arm, showing all events that occur in at least 5% of patients in D+T or SoC arm will be summarized 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 summarized by preferred term within each system organ class for the output summarising all AEs. For each preferred term, the event rate is defined as the number of patients with that AE divided by the total treatment duration (days) summed over patients and then multiplied by 365.25 x 100 to present in terms of per 100 patient years.

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

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

In addition, all AEs will be listed.

Infection Adverse events

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

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

Overall Infection AE summaries will be presented, including the number and percentage of patients in each of these categories.

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 before starting study treatment
 - Related to disease under investigation only
 - AE with outcome of death only
 - Unknown reason for death
- Death after starting study treatment
 - Related to disease under investigation only
 - Related to disease under investigation [a] and a TEAE [b] with outcome of death
 - TEAE onset on or prior to subsequent therapy or no subsequent therapy
 - TEAE onset after start of subsequent therapy
 - TEAE [b] with outcome of death only
 - TEAE onset on or prior to subsequent therapy or no subsequent therapy
 - TEAE onset after start of subsequent therapy

Death after end of safety follow up period

Related to disease under investigation and an AE with outcome of death
AE with outcome of death only
Not related to disease under investigation
Unknown relationship to disease under investigation
Unknown reason for death
Other deaths (patients who died and are not captured in earlier categories)

This output may be updated to classify deaths in one of the categories.

Separate summaries of deaths on-treatment or during safety follow-up period will be produced for the safety and bTMB20 safety analysis sets

All deaths will be listed for the safety and bTMB20 safety analysis sets, as well as deaths due to disease under investigation

Adverse events of special interest

Preferred terms used to identify adverse events of special interest (as defined in section 3.3.1) will be listed before DBL and documented in the Study Master File. Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI 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:

- All AESIs by PT
All AESI by category, sub-category and PT
At least adverse event of special interest presented by outcome
- AESI by maximum reported CTCAE grade by category, sub-category and PT
At least one adverse event of special interest possibly related to study medication by maximum reported CTCAE grade
- At least one adverse event of special interest leading to discontinuation of study medication
Time to onset of first AESI
AESI by category, sub-category and PT and onset

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

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

Summary of long term tolerability

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

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

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

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

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

Laboratory assessments

Data obtained up until the 90 days following discontinuation of immunotherapy agents (ie, the last dose of MEDI4736/tremelimumab) or 30 days following discontinuation of the Standard of Care agent or until the initiation of the first subsequent therapy following discontinuation of treatment (whichever occurs first) will be used for reporting based on the safety and bTMB \geq 20mut/Mb analysis sets. This will more accurately depict laboratory toxicities attributable to study treatment only as a number of toxicities up to 90 days following discontinuation of immunotherapy agents or 30 days following discontinuation of the Standard of Care agent possibly attributable to subsequent therapy.

However, to assess the longer term toxicity profile, all summaries of laboratory data may also be produced containing data collected up until 90 days following discontinuation of the immunotherapy agents or up until 30 days following discontinuation of the Standard of Care agent (i.e., without taking subsequent therapy into account).

A small selection of summaries of laboratory data may also be produced containing data from initiation of the first subsequent therapy following discontinuation of study treatment until 90 days following discontinuation of immunotherapy agents or until 30 days following discontinuation of the Standard of Care agent (ie summarising the laboratory data collected on patients taking subsequent therapy during the safety collection follow-up window post discontinuation of study treatment). These outputs will only be produced if the number of laboratory toxicities observed warrant the inclusion of such outputs for interpretational purposes. Any data post 90 days last dose for immunotherapy agents or post 30 days last dose for Standard of Care agents will not be summarized.

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

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

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

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

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

- Haematology: Haemoglobin (Low), Leukocytes (Low), Lymphocytes (Low), Absolute Neutrophils Count (Low), Absolute Platelets Count (Low)

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

For the parameters with no CTCAE grading that are listed in the CSP, shift tables from baseline to worst value on-treatment will be provided. Additional summaries will include a shift table for urinalysis (Color, direct Bilirubin, Blood, Glucose, Ketones, Protein) comparing baseline value to maximum on-treatment value, provided there is sufficient data such that warrants a summary.

Lab parameters, as published in the literature, that are associated (prognostic or predictive) with clinical outcomes including but not limited to neutrophil-lymphocyte-ratio or platelet-lymphocyte-ratio etc. may be analysed in an exploratory manner in regard to OS, PFS, ORR, DoR comparing MEDI4736 + tremelimumab with SoC in various defined populations and subgroups.

Reversibility of creatinine clearance (CrCl) will be tabulated.

Liver Enzyme Elevations and Hy's law

Based on the safety and bTMB ≥ 20 mut/Mb analysis sets, the following summaries will include the number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study
 - ALT $\geq 3x$ – $\leq 5x$, $> 5x$ – $\leq 8x$, $> 8x$ – $\leq 10x$, $>10x$ – $\leq 20x$, and $>20x$ Upper Limit of Normal (ULN) during the study
 - AST $\geq 3x$ – $\leq 5x$, $> 5x$ – $\leq 8x$, $> 8x$ – $\leq 10x$, $>10x$ – $\leq 20x$, and $>20x$ ULN during the study
 - Total bilirubin $\geq 2x$ – $\leq 3x$, $>3x$ – $\leq 5x$, $>5x$ ULN during the study
 - ALT or AST $\geq 3x$ – $\leq 5x$, $>5x$ – $\leq 8x$, $>8x$ – $\leq 10x$, $>10x$ – $\leq 20x$, $>20x$ ULN during the study
 - ALT or AST $\geq 3x$ ULN and Total bilirubin $\geq 2x$ 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 $\geq 3x$ ULN plus Total bilirubin $\geq 2x$ ULN or AST $\geq 3x$ ULN plus Total bilirubin $\geq 2x$ ULN at any visit.

Liver biochemistry test results over time for patients with elevated ALT or AST (ie $\geq 3x$ ULN), and elevated Total bilirubin (ie $\geq 2x$ ULN) (at any time) will be plotted. Individual

patient data where ALT or AST (ie $\geq 3x$ ULN) plus Total bilirubin (ie $\geq 2x$ 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 $3 \times \text{ULN}$ for ALT, AST, and $2 \times \text{ULN}$ for Total bilirubin. In each plot, Total bilirubin will be in the vertical axis.

Thyroid function tests

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).

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

Number (percentage) of subjects experiencing events outside normal limit ranges will be tabulated, with the denominator for percentages equal to the number of patients with at least one post baseline TSH test result.

ECGs

ECG data obtained up until the safety follow-up will be included in the summary tables based on the safety and $\text{bTMB} \geq 20 \text{mut/Mb}$ analysis sets. Overall evaluation of ECG is collected at each visit in terms of normal or abnormal, and the relevance of the abnormality is termed as “clinically significant” or “not clinically significant”. A shift table of baseline evaluation to worst evaluation will be produced if a sufficient number of ECG assessments are recorded.

Vital signs

Vital signs data obtained up until the 30 day safety follow-up visit will be included in the summary tables based on the safety and $\text{bTMB} \geq 20 \text{mut/Mb}$ analysis sets.

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 summarized over time in terms of absolute values and changes from baseline at each scheduled measurement by actual treatment group.

Vital signs will be listed by visit and timepoint

Time to Subsequent Therapy from discontinuation of study treatment

A descriptive summary will be produced for time to subsequent therapy from discontinuation of study treatment. This summary is supportive of the Adverse Event and Laboratory data outputs.

Other Safety Data

Data from positive pregnancy tests will be listed only.

Overdose listing will also be prepared.

4.2.13 ECOG performance status

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

4.2.14 PK data (MEDI4736+tremelimumab arm only)

Pharmacokinetic concentration data will be listed for each patient and each dosing day, and a summary provided for all evaluable patients in the PK analysis population.

Immunogenicity analysis

A summary of the number and percentage of patients who develop detectable ADA to MEDI4736 or tremelimumab by ADA categories (Section 3.5.3) in different treatment arms will be presented based on the ADA evaluable set. Immunogenicity results will be listed for all patients in the safety analysis set regardless of ADA-evaluable status. ADA titer and nAb data will be presented for samples confirmed positive for the presence of ADA to durvalumab and/or tremelimumab. AEs in ADA positive patients by ADA positive category will be listed. The effect of ADA on PK, safety and efficacy will be examined by descriptive summaries if data allow.

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

4.2.15 PK/PDx relationships (MEDI4736+tremelimumab)

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.

4.2.16 Biomarker data

The relationship of PD-L1 expression and bTMB/tTMB, and if appropriate, other exploratory biomarkers (e.g., IFN γ) with clinical outcomes (including but not restricted to) OS, PFS and ORR may be presented.

CCI

Summaries and analysis for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the CSR in a separate report.

4.2.17 Demographic and baseline characteristics data

- Demographic and Disease Characteristics at baseline will be summarized for all patients in the bTMB ≥ 20 mut/Mb, bTMB ≥ 16 mut/Mb, bTMB ≥ 12 mut/Mb, PD-L1-negative and FAS analysis sets and also bTMB no-call analysis set, bTMB-evaluable and bTMB-non-evaluable, tTMB-evaluable and tTMB-non-evaluable analysis sets by treatment group.

The following will be summarized for patients in the FAS and in the bTMB ≥ 20 mut/Mb analysis set, and also in additional analysis sets described in Section 2 whenever deemed necessary:

- Patient disposition (including screening failures and reason for screening failure, components of durvalumab and tremelimumab combination therapy, and for retreatment patients)
- Important protocol deviations
- Inclusion in analysis populations, and reasons for exclusion
- 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, body mass index (BMI) and body mass index group)
- Patient recruitment by region, country and center
- Previous disease-related treatment modalities
- Number of regimens of previous chemotherapy at baseline
- Previous anti-cancer therapy
- Disease characteristics at baseline (ECOG performance status, mutation detected, liver metastases, CNS metastases, PD-L1 subgroups, best response to previous therapy, overall disease classification)
- Disease characteristics at diagnosis (Primary tumour location, histology type, AJCC [American Joint Committee on Cancer] staging)
- Extent of disease at baseline
- TNM classification at baseline
- Medical history (past and current)
- Surgical history

- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer therapy
- Nicotine use, categorised (never, current, former)
- Stratification factors as per IVRS (PD-L1 and smoking status) and eCRF (Histology type) data

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

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

4.2.18 Treatment exposure

The following summaries related to study treatment will be produced for safety and bTMB ≥ 20 mut/Mb safety analysis sets, by randomized treatment group:

- Total exposure of each treatment group.
- Actual exposure of each treatment group.
- Total number of cycles received.
- Number of dose delays and reductions and reasons for dose delays/reductions, for MEDI4736 and tremelimumab, and SoC agents. Dose interruptions will be based on investigator initiated dosing decisions.
- Number of dose delays and duration of delays of MEDI4736 or tremelimumab or SoC. In addition, delays due to AEs and due to reasons other than AEs will be summarized separately.
- Number of infusions received.
- RDI (relative dose intensity) of MEDI4736 and tremelimumab.
- Exposure over time will be summarized and plotted.

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

Dose calculation of tremelimumab

Patient weight at baseline should be used for dosing calculations unless there is a $\geq 10\%$ change in weight. Dosing day weight can be used for dosing calculations instead of

baseline weight per institutional standard. The volume of tremelimumab (in mL) to add to the IV bag is calculated as follows: in combination with MEDI4736: $1 \text{ mg/kg} \times \text{patient weight (kg)} \div \text{tremelimumab concentration (nominal: 20 mg/mL)}$.

4.2.19 Subsequent Therapy

Subsequent disease-related anti-cancer therapies received after discontinuation of study treatment will have summaries produced by treatment group.

5. INTERIM ANALYSES

No interim analyses are planned for this study, although safety monitoring will be conducted by the IDMC. Details of the plan and communication process are provided in the IDMC Charter.

5.1 Independent Data Monitoring Committee

This study will use an external IDMC to assess ongoing safety analyses. The committee will meet approximately 6 months after the study has started or after the first 30 patients have been randomized, whichever occurs first, to review the safety data from the study. The IDMC will meet at least every 6 months thereafter. Following each meeting, the IDMC will report to the sponsor and may recommend changes in the conduct of the study.

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

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

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

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

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

6. CHINA COHORT

The global cohort has approximately 1330 patients enrolled from sites in North America, South America, Asian, Europe and Middle East/Africa- countries to randomize approximately 800 patients. The China cohort consists of all Chinese patients from sites accredited by CFDA and enrolled prior to the last patient last visit (LPLV) of the global cohort. The China cohort has approximately 160 randomized patients. The global cohort consists of all patients enrolled by the documented date of last patient first visit (LPFV) of the global cohort. Global recruitment will be complete once approximately 800 patients have been randomized of which 30 patients are from mainland China. Once global enrolment is completed, recruitment across all sites except for those in mainland China will be closed, and the recruitment of China patients will continue until the total number of randomized China patients is reached. Hence, a patient randomized in the China cohort prior to the LPFV of the global cohort enrolment will be included in both the (globally recruited) FAS and the China FAS. A patient randomized in the China cohort after the LPFV of the global cohort enrolment will be included only in the China FAS.

Details of the analysis, including the vendor to perform the analysis, will be specified in the China supplementary SAP, which is to be finalized before the global cohort data locks for analysis.

7. CHANGES OF ANALYSIS FROM PROTOCOL

Not applicable.

8. REFERENCES

Berry et al 1991

Berry G, Kitchin RM, Mock PA. A comparison of 2 simple hazard ratio estimators based on the logrank test. *Stat Med* 1991;10(5):749-55.

Collett 2003

Collett D. *Modelling survival data in medical research*: 2nd ed. Chapman and Hall/CRC; 2003.

Cox 1972

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

Duke-Margolis 2018

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

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.

Gail and Simon 1985

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

Glimm et al 2010

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

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.

Pintilie 2006.

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

Lan and DeMets 1983

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

Robins 1993

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

Robins and Tsiatis 1991

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

Selke and Siegmund 1983

Selke 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. *Stat Med* 1991;10(11):1665-77.

9. APPENDIX

None.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d419ac00003-sap-ed-3		
Document Title:	Statistical Analysis Plan Edition 3	
Document ID:	CCI [REDACTED]	
Version Label:	3.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	Signed by	Meaning of Signature
06-Jun-2019 13:14 UTC	PPD [REDACTED]	Content Approval
06-Jun-2019 19:40 UTC	PPD [REDACTED]	Author Approval

CCI [REDACTED]