AstraZeneca

Statistical Analysis Plan

Drug Substance	MEDI4736 and tremelimumab
Study Code	D4193C00002
Edition Number	04
Date	18 January 2018

A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 Monotherapy and MEDI4736 in Combination with Tremelimumab Versus Standard of Care Therapy in Patients with Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck (SCCHN)

A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 Monotherapy and MEDI4736 in Combination with Tremelimumab Versus Standard of Care Therapy in Patients with Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck (SCCHN)

Study Statistician [redacted]

Date

A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 Monotherapy and MEDI4736 in Combination with Tremelimumab Versus Standard of Care Therapy in Patients with Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck (SCCHN)

Global Product Statistician [redacted]

[redacted] Date

TABLE OF CONTENTS PAGE

TITLE PAGE 1
SIGNATURE OF STUDY STATISTICIAN 2
SIGNATURE OF GLOBAL PRODUCT STATISTICIAN 3
TABLE OF CONTENTS 4
LIST OF ABBREVIATIONS 8
AMENDMENT HISTORY 11
1. STUDY DETAILS 14
2. STUDY OBJECTIVES 15
2.1 Co-Primary Objectives 15
2.1.1 Secondary Objectives 15
2.1.2 Safety Objectives 16
[redacted]
2.2 Study Design 17
2.3 Number of Patients 23
3. ANALYSIS SETS 25
3.1 Definition of Analysis Sets 25
3.1.1 Full Analysis Set (FAS) 25
3.1.2 PD-L1-Negative Analysis Set 25
3.1.3 PD-L1-Positive Analysis Set 25
3.1.4 Safety Analysis Set 25
[redacted]
3.2 Violations and Deviations 26
4. PRIMARY AND SECONDARY VARIABLES 28
4.1 Derivation of RECIST Visit Responses 28
4.1.1 Site Investigator Assessment Using RECIST 1.1 30
4.1.1.1 Site Investigator Assessment Using RECIST 1.1: Target Lesions (TLs) 30
4.1.1.2 Non-Target Lesions (NTLs) and New lesions 34
4.1.1.3 Site Investigator Assessment Using RECIST 1.1: Overall Visit Response 36
4.2 Outcome Variables 36
4.2.1 Primary Endpoint: Overall Survival (OS) 37
4.2.2 Secondary Endpoints 37
4.2.2.1 Progression-Free Survival (PFS) 37
4.2.2.2 Objective Response Rate (ORR) 38

4.2.2.3 Duration of Response (DoR) 38
4.2.2.4 Disease Control Rate (DCR) 39
4.2.2.5 Proportion of Patients Alive and Progression Free at 6 Months (APF6) 39
4.2.2.6 Proportion of Patients Alive and Progression Free at 12 Months (APF12) 39
4.2.2.7 Proportion of Patients Alive at 12 Months (OS12) 39
4.2.2.8 Proportion of Patients Alive at 18 Months (OS18) 39
4.2.2.9 Proportion of Patients Alive at 24 Months (OS24) 39
4.2.2.10 Best Objective Response (BoR) 40
4.2.2.11 Time to Response 40
4.2.2.12 Time from Randomization to First Subsequent Therapy or Death (TFST) 40
4.2.2.13 Time from Randomization to Second Subsequent Therapy or Death (TSST) 41
4.2.2.14 Time from Randomization to Second Progression-Free Survival (PFS2) 41
4.2.2.15 Change in Tumor Size 41

4.3 Safety 42
4.3.1 Adverse Events (AEs) 42
4.3.2 Treatment Exposure 43
4.3.3 Dose Intensity 45
4.3.4 Laboratory Data 46
4.3.5 ECGs 46
4.3.6 Vital Signs 47
4.3.7 General Considerations for Safety Assessments 47

4.4 Patient reported outcome 50
4.4.1 EORTC QLQ-C30 50
4.4.2 EORTC QLQ-H&N35 53
[redacted]
[redacted]
4.6 Biomarker Variables 57
[redacted]
5. ANALYSIS METHODS 57

5.1 General Principles 58
5.1.1 Multiple Testing Strategy 60

5.2 Analysis Methods 61
5.2.1 Overall Survival (OS) 61
5.2.2 Progression Free Survival (PFS) 66
5.2.3 Objective Response Rate (ORR) 66

5.2.4 Best Objective Response (BoR) 66
5.2.5 Disease Control Rate (DCR) 66
5.2.6 Duration of Response (DoR) 67
5.2.7 Proportion of Patients Alive and Progression Free at 6 and 12 Months 67
5.2.8 Time to Response (TTR) 67
5.2.9 Time from Randomization to Second Progression (PFS2) 67
5.2.10 Time from Randomization to First and Second Subsequent Therapy or Death 67
5.2.11 Change in Tumor Size 67
5.2.12 Patient Reported Outcome (PRO) 68
5.2.12.1 EORTC QLQ-C30 68
5.2.12.2 EORTC QLQ-H&N35 69
[redacted]
5.2.12.7 PRO Compliance 70
5.2.13 Safety 70
5.2.13.1 Adverse Events 70
5.2.13.2 Laboratory Assessments 75
5.2.13.3 ECGs 77
5.2.13.4 Vital Signs 77
5.2.14 WHO/ECOG Performance Status 77
[redacted]
5.2.17 Demographic, Initial Diagnostics and Baseline Characteristics Data 78
5.2.18 Treatment Exposure 79
5.2.19 Subsequent Therapy 80
6. INTERIM ANALYSIS 80
7. CHANGES OF ANALYSIS FROM PROTOCOL 81
8. REFERENCES 83
9. APPENDIX (NOT APPLICABLE) 84

LIST OF TABLES

Table 1 Co-Primary Objectives 15
Table 2 Secondary Objectives 15
Table 3 Safety Objectives 16
[redacted]
Table 5 Summary of Statistical Assumptions 24
Table 6 Summary of Outcome Variables and Analysis Populations 25
Table 7 TL Visit Responses 31
Table 8 NTL visit responses 34
Table 9 Overall Visit Responses 36
Table 10 Change from BL and Visit Response for EORTC QLQ C30 51
Table 11 Change from BL and Visit response for EORTC QLQ H&N35 54
Table 12 Pre-planned statistical and sensitivity analyses to be conducted 59
Table 13 Changes of Analysis from Protocol 81

LIST OF FIGURES

Figure 1 Overall Study Design 19
Figure 2 Study Flow Chart for MEDI4736 + Tremelimumab Combination Therapy and MEDI4736 Monotherapy Groups 20
Figure 3 Study Flow Chart for Standard of Care Group 22
Figure 4 Levels of the Multiple Testing Strategy 61

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
[redacted]	[redacted]
AE	Adverse Event
AESI	Adverse Events of Special Interest
ALP	Alkaline phosphatase
ALT	Alanine Aminotransferase
APF6	Proportion of Patients Alive and Progression Free at 6 Months from Randomization
APF12	Proportion of Patients Alive and Progression Free at 12 Months from Randomization
AST	Aspartate Aminotransferase
Baseline	Refers to the last assessment prior to intake of the first dose of IP, except for Efficacy where baseline refers to the last visit prior to enrolment.
BoR	Best Objective Response
CI	Confidence Interval
CR	Complete Response
CRA	Clinical Research Associate
CRF / eCRF	Case Report Form (electronic)
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Event
CTLA-4	Cytotoxic T–Lymphocyte-Associated Antigen 4
CTM	Clinical Team Manager
DBP	Diastolic Blood Pressure
DCO	Data Cut-Off
DCR	Disease Control Rate
DoR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EDoR	Expected Duration of Response
EORTC	European Organisation for Research and Treatment of Cancer
[redacted]	[redacted]

Abbreviation or special term	Explanation
FAS	Full Analysis Set
[redacted]	[redacted]
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
IDMC	Independent Data Monitoring Committee
IMT	Immunomodulatory Therapy
IP	Investigational Product
irRC	Immune-Related Response Criteria
ITT	Intent-to-Treat
IV	Intravenous
KM	Kaplan-Meier
MD	Medical Doctor
MedDRA	Medical Dictionary for Regulatory Activities
MEDI4736	Immune-Mediated Therapy
mg	Milli-gram
MMA	Medical Monitoring Associate
MRI	Magnetic Resonance Imaging
NA	Not Applicable
nAb	Neutralizing antibody
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NTL	Non-target lesions
ORR	Objective response rate
OS	Overall survival
OS12	Proportion of patients alive at 12 months from randomization
OS18	Proportion of patients alive at 18 months from randomization
OS24	Proportion of patients alive at 24 months from randomization
PD	Progressive disease
PD-1	Programmed cell death 1
PD-L1	Programmed cell death ligand 1
PD-L1 –ve	Patients with PD-L1–negative tumor expression status
PD-L1 +ve	Patients with PD-L1–positive tumor expression status
PDx	Pharmacodynamic(s)

Abbreviation or special term	Explanation
PFS	Progression free survival
[redacted]	[redacted]
[redacted]	[redacted]
PR	Partial response
PRO	Patient reported outcome
q12W	Every 12 weeks
q2W	Every 2 weeks
q4W	Every 4 weeks
QLQ-C30 v3	30-item core quality of life questionnaire, version 3
QLQ-H&N35	35-item head and neck quality of life questionnaire
QoL	Quality of life
QTc	QT interval corrected for heart rate
QTcF	QT interval corrected for heart rate using Fridericia's formula
RECIST 1.1	Response Evaluation Criteria In Solid Tumors version 1.1
RR	Response rate
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SCCHN	Squamous cell carcinoma of the head and neck
SD	Stable disease
SDV	Source data verify
SBP	Systolic blood pressure
TL	Target lesions
TSST	Time to second subsequent therapy
TTR	Time to treatment response
ULN	Upper limit of normal
WHO	World Health Organization

AMENDMENT HISTORY

Date	Brief description of change
24 May 2016	In line with the Clinical Study Protocol (CSP) Amendments • Changes in primary endpoints (OS instead of both OS and PFS) • Removal of blinded independent central review • Changes in multiple testing procedures • Addition of an interim analysis • Simplification of PRO endpoints for secondary analysis • Text in sample size section updated In line with project wide developments • Removal of non-critical sensitivity analyses • Removal of non-critical secondary analyses • Reduction of the lists of deviations regarded as important protocol deviations • Change in definition for Total and Actual exposure. • Removal of PID • Addition of summaries of AESIs requiring concomitant medication and particularly the relationship of AESIs to the use of immunosuppressive agents • Clarification of statistical method used to compute point estimate and confidence intervals for primary analysis • Inclusion of additional prognostic factors, and changes in statistical methods for subgroup analyses • Removal of formal analysis of Duration of Response endpoint Other changes • Addition of Time to response endpoint • Additional minor clarifications.

Date	Brief description of change
21 February 2017	In line with the Clinical Study Protocol (CSP) Amendment Version 5 • Changes from a single primary objective of MEDI4736 +tremelimumab versus SoC in terms of OS to co-primary objectives of MEDI4736 +tremelimumab versus SoC in terms of OS and MEDI4736 monotherapy versus SoC in terms of OS. • Changes in multiple testing procedures. • Text in sample size section updated. In line with project wide developments • Hazard ratio and confidence interval derived at landmarks as per Klein's method were removed. • Description of haemorrhages adverse events added. Other changes • Addition of time to first and second subsequent therapy. • Harmonization of PRO's statistical analysis wording. • Additional minor clarifications.

Date	Brief description of change
5 January 2018	• The cut point for determining the PD-L1 positive and PD-L1 negative groups was confirmed to be 25%. • [redacted] • Text defining what constitutes a suspected Hy's law case has been clarified to show that ALT/AST elevations can be at any visit prior to, or at the same time as the bilirubin elevation (i.e. do not have to be concurrent). • The method to control for tied observations in the statistical analyses has been specified along with the method fitting stratification variables (Efron and strata statements) to avoid ambiguity. • Wording stating confidence intervals would be (1-α)% simplified to 95%. • Patients with high baseline PRO scores are excluded from analyses of time to deterioration. The threshold was changed from ≤95 to ≤90. • The definition of PFS2 has been updated to clarify that the second progression is the first progression following a first subsequent therapy. • The lab assessments to be included in the TFLs has been updated. • The conversion from leukocyte percentages to counts has been added. • ECG text clarified to reflect that they're collected ad hoc. • "Subject too heavily affected by symptoms of disease under investigation" added as a deterioration event for HRQoL. • Text was added to section 4.3.7 to clarify the imputation rules for cases where there may be a missing AE start date. • AESI list updated in section 4.3.1 to align with CSP amendment. • Table 6 Summary of Outcome Variables and Analysis Populations incorrectly implied DCR and DoR would be analysed in each of the subset of patients with by PD-L1 positive and negative. Thus these were removed.

1. STUDY DETAILS

The target population for this study is patients with recurrent or metastatic SCCHN who have progressed during or after only one palliative systemic treatment regimen for recurrent or metastatic disease that must have contained a platinum agent or who have progressed within 6 months of the last dose of platinum given as part of multimodality therapy of curative intent. These patients have poor prognosis with limited Standard of Care (SoC) therapeutic options, which only have transient and limited benefit (ORR: 4% to 13%, progression-free survival [PFS]: 2.5 months, and OS: 5.5 months) (Shin and Khuri 2013, Vermorken et al 2008). Thus, there is a significant unmet medical need for additional treatment options for use in this patient population.

Multiple lines of evidence suggest that SCCHN tumors, like many other malignancies, create a highly immunosuppressive environment and may be amenable to therapeutic intervention with immune-modulating agents. Two specific pathways, the programmed cell death ligand 1 (PD-1–PD-L1) axis and the Cytotoxic T–lymphocyte-associated antigen 4 (CTLA-4) pathway have been successfully targeted by immunomodulating therapies (IMT) to obtain tumor reduction.

MEDI4736, an antibody that blocks the interaction between PD-L1 and its receptors, may relieve PD-L1-dependent immunosuppressive effects and, therefore, enhance the cytotoxic activity of anti-tumor T-cells. This hypothesis is supported by emerging clinical data from Study 1108 which indicates encouraging response rates (RRs), duration of response (DoR), and radiographic tumor shrinkage in 12 of 46 evaluable patients with a manageable safety profile which justifies this confirmatory study.

Combining immunotherapy agents has been shown to result in improved RR relative to those for monotherapy. The rationale for combining MEDI4736 and tremelimumab is that the mechanisms of CTLA-4 and PD-1 are non-redundant, suggesting that targeting both pathways may have additive or synergistic activity (Pardoll 2012), resulting in higher RRs in SCCHN. Evidence of clinical activity and manageable safety profile has been observed in Study D4190C00006.

The co-primary objectives of the study are:

- To assess the efficacy of **MEDI4736 + tremelimumab combination therapy versus SoC** in patients with squamous cell carcinoma of the head and neck (SCCHN), regardless of programmed cell death ligand 1 (PD-L1) status, in terms of overall survival (OS)
- To assess the efficacy of **MEDI4736 monotherapy versus SoC** in patients with SCCHN, regardless of PD-L1 status, in terms of overall survival (OS)

2. STUDY OBJECTIVES

2.1 Co-Primary Objectives

Table 1 Co-Primary Objectives

Co-primary Objectives:	Outcome Measures:
To assess the efficacy of MEDI4736 + tremelimumab combination therapy versus SoC in terms of OS	OS in all patients, regardless of PD-L1 status
To assess the efficacy of MEDI4736 monotherapy versus SoC in terms of OS	OS in all patients, regardless of PD-L1 status

OS Overall survival; PD-L1 programmed cell death ligand 1; SoC Standard of Care.

2.1.1 Secondary Objectives

Table 2 Secondary Objectives

Key Secondary Objectives:	Outcome Measures:
To further assess the efficacy of MEDI4736 + tremelimumab combination therapy versus SoC in terms of OS	OS in PD-L1-negative patients
To assess the efficacy of MEDI4736 monotherapy versus SoC in terms of OS	OS in PD-L1-positive patients
To further assess the efficacy of MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy versus SoC in terms of PFS, PFS2, ORR, TTR, TFST, TSST, DoR, DCR, APF6, APF12, OS12, OS18, and OS24	PFS, PFS2, ORR, TTR, DoR, DCR, APF6, and APF12 using the site Investigator's assessments **according to RECIST 1.1** OS12, OS18, OS24, TFST and TSST
To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to MEDI4736 monotherapy in terms of PFS, ORR, and OS	PFS and ORR in PD-L1–negative patients using the site Investigator's assessments according to RECIST 1.1 OS in PD-L1–negative patients
To explore symptoms and HRQoL in patients treated with MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy versus SoC using the EORTC QLQ-C30 v3 and the H&N35 module	EORTC QLQ-C30: global health QoL, functioning (physical), and symptoms (fatigue) EORTC QLQ-H&N35: symptoms (pain, swallowing) Changes in World Health Organization/Eastern Cooperative Oncology Group performance status

APF6 Proportion of patients alive and progression free at 6 months from randomization; APF12 Proportion of patients alive and progression free at 12 months from randomization; DCR Disease control rate; DoR Duration of response; EORTC European Organisation for Research and Treatment of Cancer; QLQ-C30 v3 30-item core quality of life questionnaire, version 3; H&N35 35-item head and neck quality of life questionnaire; HRQoL Health-related quality of life; OS12 Overall survival at 12 months; OS18 Overall survival at 18 months; OS24 Overall survival at 24 months; PD-L1 programmed cell death ligand 1; PFS Progression-free survival; PFS2 Time from randomization to second PFS; QoL Quality of life; RECIST 1.1 Response Evaluation Criteria in Solid Tumors version; SoC Standard of Care; TTR Time to treatment response; TFST Time to first subsequent therapy; TSST Time to second subsequent therapy

2.1.2 Safety Objectives

Table 3 Safety Objectives

Safety Objective:	Outcome Measures:
To assess the safety and tolerability profile of MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy compared to SoC	AEs, physical examinations, laboratory findings, and vital signs

AE adverse event; SoC Standard of Care

2.2 Study Design

This is a randomized, open-label, multi-center, global Phase III study to determine the efficacy and safety of MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy versus SoC therapy in the treatment of patients with recurrent or metastatic SCCHN who have progressed during or after only one palliative systemic treatment regimen for recurrent or metastatic disease that must have contained a platinum agent or who have progressed within 6 months of the last dose of platinum given as part of multimodality therapy of curative intent (i.e., those who are refractory to prior platinum therapy).

Patients will undergo a screening assessment on their tumor tissue sample to determine HPV status (for patients with oropharyngeal cancer only).

Inclusion criteria include a valid PD-L1 result. PD-L1 expression levels are assessed by IHC with the VENTANA PD-L1 (SP263) Assay. PD-L1 positive is defined as ≥25% of tumor cells with membrane staining for PD-L1 at any intensity. PD-L1 low/negative is defined as <25% of tumor cells with membrane staining for PD-L1 at any intensity.

If the patient's PD-L1 status has already been assessed using the VENTANA PD-L1 SP263 IHC assay as a part of the screening process for another AstraZeneca/ MedImmune study, this test result can be used for the determination of eligibility. The specified expression cut-off level will be used for the purpose of stratification and therefore included in the stratified log rank tests for OS. However the actual cut-off level for the subgroup analyses of OS by PD-L1 status and also as the basis for determining the PD-L1-negative subgroup in the MTP might be done using different cutoff (see sections 3.1.2 and 3.1.3).

Patients will be randomized in a stratified manner according to prognostic factors: PD-L1 status, tumor location/HPV status, and smoking status. HPV status and smoking status are known prognostic markers for patients with head and neck cancer. Patients will be randomized in a 1:1:1 ratio to receive treatment with MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy, or SoC therapy.

Patients in the MEDI4736 monotherapy treatment group will receive 10 mg/kg MEDI4736 via intravenous (IV) infusion q2w until PD. Patients in the MEDI4736 + tremelimumab treatment group will receive 20 mg/kg MEDI4736 via IV infusion q4w for up to 4 months and tremelimumab 1 mg/kg via IV infusion q4w for 4 months. Upon completion of 4 months of combination therapy, patients in the MEDI4736 + tremelimumab arm will then continue dosing with MEDI4736 monotherapy at 10 mg/kg q2w beginning 4 weeks after the last dose of combination therapy is administered until PD. Patients in the SoC treatment group will receive monotherapy with 1 of the following therapies at the Investigator's discretion until PD: cetuximab, a taxane (i.e., docetaxel or paclitaxel), methotrexate, or a fluoropyrimidine (i.e., 5-FU, TS-1, or capecitabine).

Patients randomized to 1 of the IMT treatment groups will receive treatment with either MEDI4736 or MEDI4736 + tremelimumab beginning on Day 0 until objective disease progression according to RECIST 1.1. (unless, in the Investigator's opinion, the patient continues to receive benefit from the treatment), initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or another discontinuation criterion is met. Patients with PD who, in the Investigator's opinion, continue to receive benefit from their assigned IMT treatment and who meet the criteria for treatment in the setting of PD may continue to receive their assigned IMT treatment after consultation with the Sponsor and at the Investigator's discretion. All patients randomized to 1 of the IMT treatment groups require documentation of objective disease progression according to RECIST 1.1. A second scan obtained at a minimum of 4 weeks later to confirm progression is required for treatment management decisions only and only where it is clinically feasible. Disease response assessment should be solely based on RECIST 1.1 with response of PD entered for the first scan that meets progression criteria as outlined by RECIST 1.1. Patients with confirmed progression in the monotherapy arm or in the combination portion of therapy in the MEDI4736 + tremelimumab arm cannot continue therapy if the progression occurred during dosing after objective response in the target lesions (i.e., the response and progression events both occurred while receiving active IP during the same treatment period in the target lesions).

Patients who the Investigator determines may not continue IMT treatment after PD will enter follow-up and will be followed up until death. Patients who have discontinued IMT treatment due to toxicity or symptomatic deterioration, those who have no objective disease progression according to RECIST 1.1, or who have commenced subsequent anticancer therapy, will be followed up until death.

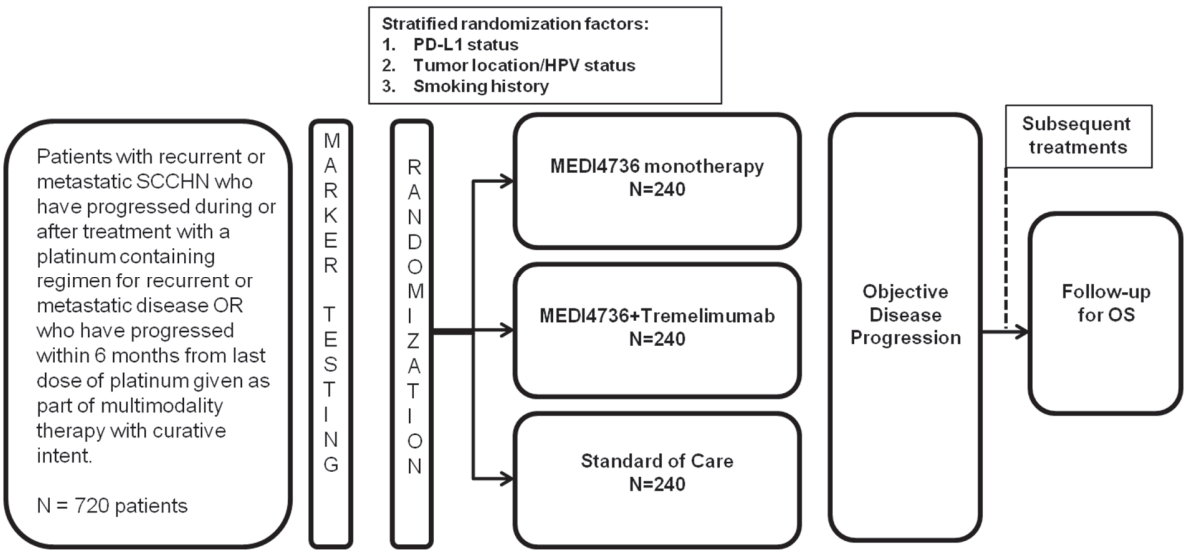
Patients randomized to the SoC arm will receive the Investigator-chosen SoC treatment beginning on Day 0 until PD, initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or another discontinuation criterion is met. For equivocal findings of progression (e.g., very small and uncertain new lesions; cystic changes or necrosis in existing lesions), treatment may continue until the next scheduled assessment. All patients randomized to the SoC treatment group require documentation of objective disease progression according to RECIST 1.1. A second scan obtained at a minimum of 4 weeks later to confirm progression is required in patients where it is clinically feasible and only for treatment management

decisions. Disease response assessment should be solely based on RECIST 1.1 with response of PD entered for the first scan that meets progression criteria as outlined by RECIST 1.1. Patients who the Investigator determine may not continue SoC treatment will enter follow-up. Patients who have discontinued SoC treatment due to PD, toxicity or symptomatic deterioration, who have no objective disease progression or who have commenced subsequent anticancer therapy, will be followed up for study endpoints (objective disease progression according to RECIST 1.1 or death).

Tumor assessments will be performed using computed tomography (CT) or magnetic resonance imaging (MRI) at the times specified in the protocol. (tables 2 and 3). RECIST 1.1 measurements as given by the site Investigator's assessments will be used to derive the secondary variables of PFS, ORR, DoR, DCR, APF6, and APF12.

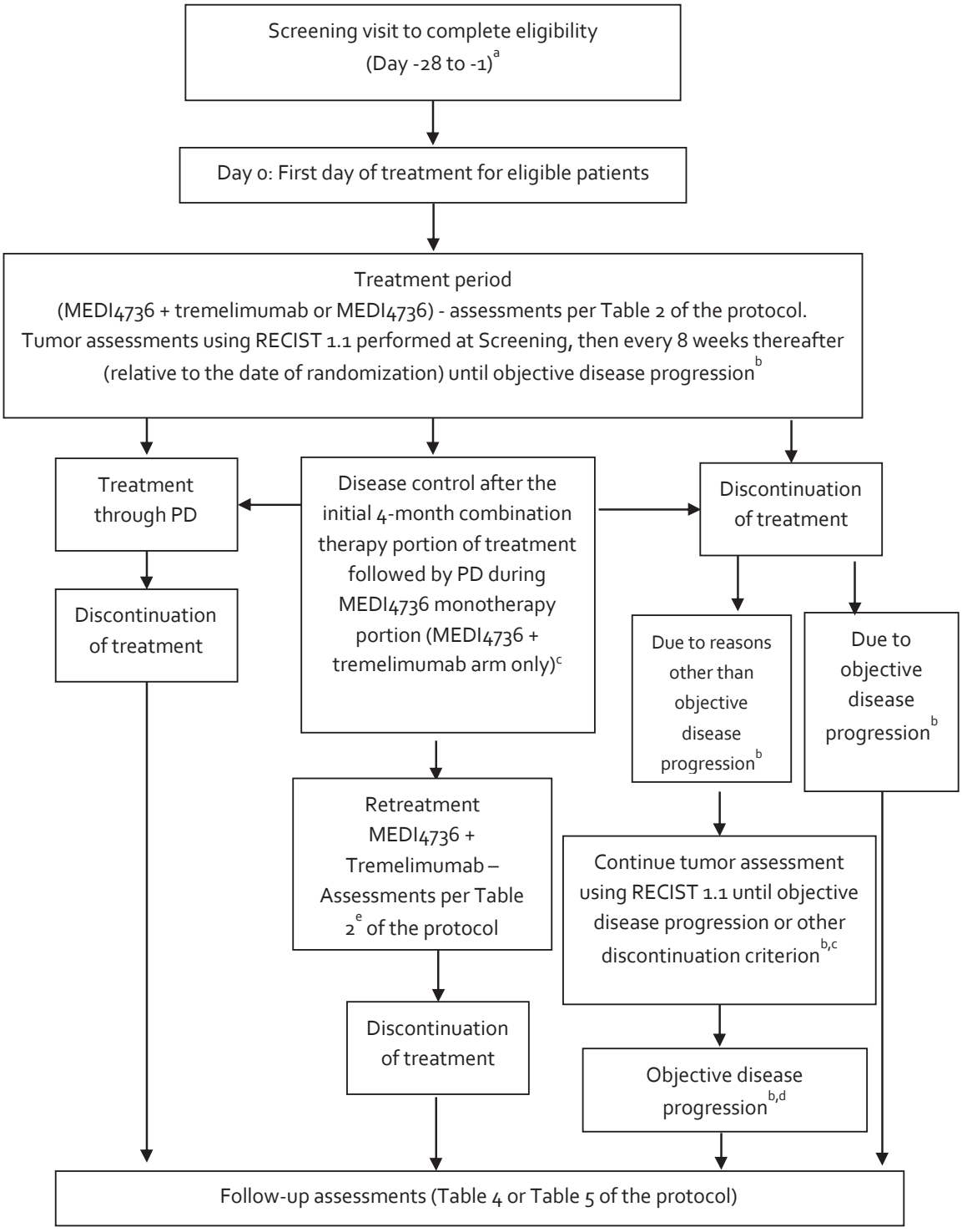
Below is a schematic diagram of the overall study design (Figure 1). A flow chart for the IMT treatment groups (MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy) is presented in Figure 2, followed by a flow chart for the SoC group which is presented in Figure 3.

Figure 1 Overall Study Design



Note: For all Standard of Care therapies, a particular treatment (cetuximab, taxane, methotrexate, or fluoropyrimidine-based regimen) will not be used in patients who have previously received that treatment for recurrent/metastatic disease or who have experienced recurrence or progression of disease within 6 months of prior multimodal therapy using that particular treatment.
*: For HPV status, only patients with oropharyngeal cancer will be tested.
HPV human papillomavirus; PD-L1 programmed cell death ligand 1; OS overall survival; SCCHN Squamous Cell Carcinoma of the Head and Neck.

Figure 2 **Study Flow Chart for MEDI4736 + Tremelimumab Combination Therapy and MEDI4736 Monotherapy Groups**



[a] Screening assessments may be performed over multiple visits. Imaging and procedures performed before signing the ICF may be used for screening purposes if the patient consents to their use. However, randomization must occur within 28 days of all procedures (with the exception of tumor biopsy [if required] and PD-L1 testing) used for screening purposes. If the patient's PD-L1 status has already been assessed using the VENTANAPD-L1 (SP263) IHC assay as a part of the screening process for another AstraZeneca/MedImmune study, this test result can be used for the determination of eligibility.

[b] A confirmatory scan is required following the initial demonstration of PD, if clinically feasible. A second scan obtained at a minimum of 4 weeks later to confirm progression is required for treatment management decisions only and only where it is clinically feasible. Disease response assessment should be solely based on RECIST 1.1 with response of PD entered for the first scan that meets progression criteria as outlined by RECIST 1.1. Administration of IMT treatment will continue between the initial assessment of progression and confirmation for progression. Patients in the IMT treatment groups with confirmed PD who continue to receive study treatment at the discretion of the Investigator (following consultation with the Sponsor) will continue to follow the assessments in Table 2 of the protocol until treatment is discontinued. This will not be considered retreatment but will be considered a part of the initial therapy. For all patients who are treated through progression, the Investigator should ensure that patients do not have any significant, unacceptable, or irreversible toxicities that indicate that continuing treatment will not further benefit the patient, and that the patient still meets all of the inclusion criteria and none of the exclusion criteria for this study including re-consenting to continue treatment. Patients with rapid tumor progression or with symptomatic progression that requires urgent medical intervention (e.g., central nervous system metastasis, respiratory failure due to tumor compression, or spinal cord compression) will not be eligible to continue to receive IP.

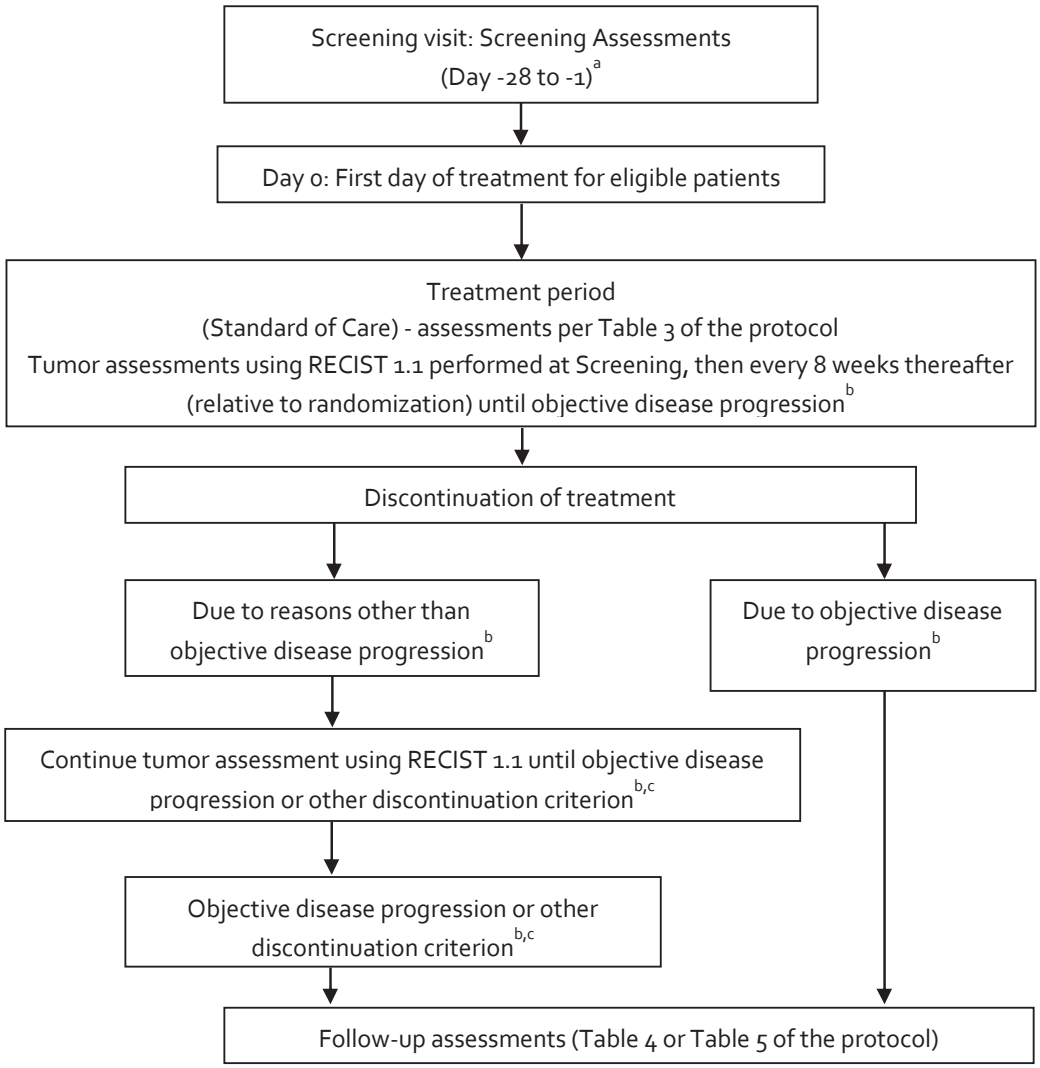
[c] Patients assigned to the MEDI4736 + tremelimumab combination therapy arm who complete the 4 cycles of the combination therapy portion of the treatment regimen that achieve and maintain disease control (i.e., CR, PR, or SD) and subsequently progress on monotherapy MEDI4736 may restart their assigned treatment once Before restarting MEDI4736 + tremelimumab combination treatment, the Investigator should ensure that patients do not have any significant, unacceptable, or irreversible toxicities that indicate that continuing treatment will not further benefit the patient, and that the patient still meets all of the inclusion criteria and none of the exclusion criteria for this study including re-consenting to restart treatment. To restart study treatment, the patient must not have received an intervening cancer therapy post study treatment discontinuation. Patients should have a baseline tumor assessment within 28 days of restarting study treatment; all further scans should occur q8w (relative to the date of restarting study treatment). Patients cannot obtain retreatment if the progression occurred during dosing with the combination portion of therapy after objective response in the target lesions (i.e., the response and progression events both occurred while receiving active IP during the same treatment period in the target lesions). Retreatment in the combination arm can only occur if PD occurs during the monotherapy portion of dosing.

[d] Patients with objective disease progression according to RECIST 1.1 who discontinue IP should have scans conducted according to local practice until the patient commences a new treatment (these scans are optional).

[e] Patients who progress on the MEDI4736 + tremelimumab arm may be eligible for retreatment if they progress during the monotherapy portion of dosing.

CR Complete response; IP Investigational product; PD Progressive disease; PD-L1 Programmed cell death ligand-1; PR Partial response; q8w Every 8 weeks; RECIST 1.1 Response Evaluation Criteria in Solid Tumors version 1.1; SD Stable disease.

Figure 3 Study Flow Chart for Standard of Care Group



[a] Screening assessments may be performed over multiple visits. Imaging and procedures performed before signing the ICF may be used for screening purposes if the patient consents to their use. However, randomization must occur within 28 days of all procedures (with the exception of tumor biopsy [if required] and PD-L1 testing) used for screening purposes. If the patient's PD-L1 status has already been assessed using the VENTANAPD-L1 (SP263) IHC assay as a part of the screening process for another AstraZeneca/MedImmune study, this test result can be used for the determination of eligibility.

[b] A confirmatory scan is required if clinically feasible following the initial demonstration of PD. A second scan obtained at a minimum of 4 weeks later to confirm progression is required for treatment management decisions only and only where it is clinically feasible. This scan should occur preferably at the next scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinical deterioration if clinically feasible. For equivocal findings of progression (e.g., very small and uncertain new lesions; cystic changes or necrosis in existing lesions), treatment may continue until the next scheduled assessment. Disease response assessment should be solely based on RECIST 1.1 with response of PD entered for the first scan that meets progression criteria as outlined by RECIST 1.1. In the SoC group, it is at the Investigator's discretion whether or not a patient with uncertain PD continues treatment until PD is confirmed; however, a scan with objective disease progression according to RECIST 1.1 is required for all patients in the SoC group, even if a subsequent treatment is started.

[c] Patients with objective disease progression according to RECIST 1.1 who discontinue SoC treatment should have scans conducted according to local practice until the patient commences a new treatment (these scans are optional).

2.3 Number of Patients

The sample size for this study was selected to be consistent with the research hypothesis as described in Section 8.5 of the Clinical Study Protocol (CSP).

The study will enroll approximately 1200 patients in order to identify 720 eligible patients with recurrent or metastatic SCCHN who have progressed during or after only one palliative systemic treatment regimen for recurrent or metastatic disease that must have contained a platinum agent or who have progressed within 6 months of the last dose of platinum given as part of multimodality therapy with curative intent. Patients will be randomized in a 1:1:1 fashion to MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy, or SoC (240 patients in each treatment group). An approximate 40% attrition rate due to screen failures and a <2% attrition rate due to withdrawals during treatment are expected. The study is sized to characterize the OS benefit of MEDI4736 + tremelimumab combination therapy versus SoC in all patients, regardless of PD-L1 status and to characterize the OS benefit of MEDI4736 monotherapy versus SoC in all patients, regardless of PD-L1 status. The sizing assumes a 3-month delay in separation of the survival curves between each arm, hence the use of average hazard ratios (HRs).

The analysis of OS is expected to be performed after approximately 11 months of follow-up:

- When 375 death events have occurred in 480 patients (78% maturity) across the MEDI4736 + tremelimumab combination therapy and SoC arms, regardless of PD-L1 status AND

- When 375 death events have occurred in 480 patients (78% maturity) across the MEDI4736 monotherapy and SoC arms, regardless of PD-L1 status.

Interim analysis for OS will be performed when a total of 300 events (80% of required events) have been accumulated across the MEDI4736 + tremelimumab combination therapy and SoC arms, regardless of PD-L1 status and a total of 300 events (80% of required events) have been accumulated across the MEDI4736 monotherapy and SoC arms, regardless of PD-L1 status.

MEDI4736 + tremelimumab combination therapy <u>or</u> MEDI4736 monotherapy, in all patients, regardless of PD-L1 status (co-primary objectives)

If OS at 18 months was 25% with either the MEDI4736 + tremelimumab combination therapy **<u>or</u>** MEDI4736 monotherapy and 10% with SoC (with 5.5 month median OS), and assuming the true average OS HR is 0.69, the study will have 90% power to demonstrate statistical significance at the 2.2% level (using a 2-sided test) for the comparison of either MEDI4736 + tremelimumab combination therapy **<u>or</u>** MEDI4736 monotherapy versus SoC, allowing for 1 interim analysis conducted at approximately 80% of the target events with the smallest treatment difference that could be statistically significant being an average HR of 0.79. With an assumed 15-month recruitment period and a minimum follow-up period of 11 months from "last patient in", it is anticipated that the final analysis will be performed 26 months after the first patient has been recruited.

MEDI4736 + tremelimumab combination therapy in PD-L1-negative patients (key secondary objective)

Assuming 70% of randomized patients are PD-L1-negative patients (i.e., 504 PD-L1-negative patients with 168 in each treatment arm), and if OS at 18 months was 28% with the MEDI4736 + tremelimumab combination therapy and 10% with SoC (with 5.5 month median OS), and assuming the true average OS HR is 0.64, the study will have 90% power to demonstrate statistical significance at the 2.2% level (using a 2-sided test) for the comparison of MEDI4736 + tremelimumab combination therapy versus SoC in the PD-L1-negative patients, with the smallest treatment difference that could be statistically significant being an average HR of 0.75. With a 15-month recruitment period and a minimum follow-up period of 11 months assumed for PD-L1-negative patients, it is anticipated that this analysis will be performed 26 months after the first patient has been recruited.

MEDI4736 monotherapy in PD-L1-positive patients (other secondary objective)

Assuming 30% of randomized patients are PD-L1-positive patients (i.e., 216 PD-L1-positive patients with 72 in each treatment arm), and if OS at 18 months was 41% with either MEDI4736 monotherapy and 10% with SoC (with 5.5 month median OS), and assuming the true average OS HR is 0.49, the trial will have 90% power to demonstrate statistical significance at the 2.5% level (using a 2-sided test) for the comparison of either MEDI4736 monotherapy versus SoC in PD-L1 positive patients, with the smallest treatment difference that could be statistically significant being an average HR of 0.63. With a 15-month recruitment period and a minimum follow-up period of 11 months assumed, it is anticipated that this analysis will be performed 26 months after the first patient has been recruited. A summary of the statistical assumptions is provided in Table 5.

Table 5 Summary of Statistical Assumptions

	N	Overall HR	Landmarks[a]	Events (maturity)	Power	Critical values HR (Landmarks)
MEDI4736 + tremelimumab combination therapy versus SoC						
All patients, regardless of PD-L1 status	480	0.69	10 vs. 25%	375 (78%)	90%	0.79 (10 vs. 19%)
PD-L1–negative	336	0.64	10 vs. 28%	258 (77%)	90%	0.76 (10 vs. 21%)
MEDI4736 monotherapy versus SoC						
All patients, regardless of PD-L1 status	480	0.69	10 vs. 25%	375 (78%)	90%	0.79 (10 vs. 19%)
PD-L1–positive	144	0.49	10 vs. 41%	102 (71%)	90%	0.63 (10 vs. 28%)

[a] OS landmark is 18 months; the assumed recruitment and minimum follow-up periods for all patients, regardless of PD-L1 status, are 15 and 11 months, respectively (i.e., 26-month study duration); Sample size estimate in the PD-L1 negative comparison assume that 70% of the patients enrolled have PD-L1 negative disease. Sample size estimate in the PD-L1 positive comparison assume that 30% of the patients enrolled have PD-L1 positive disease.

HR hazard rate; PD-L1 programmed cell death ligand 1; SoC Standard of Care

3. ANALYSIS SETS

3.1 Definition of Analysis Sets

Four analysis sets are defined for this study. Table 6 gives a summary of outcome variables and analysis populations.

3.1.1 Full Analysis Set (FAS)

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

3.1.2 PD-L1-Negative Analysis Set

The PD-L1-negative analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1-negative as defined by the VENTANA PD-L1 (SP263) IHC assay. The cut-off level to determine the PD-L1-negative analysis set will be the same as that for stratification purposes (<25% PD‑L1–membrane expression in tumor tissue).

3.1.3 PD-L1-Positive Analysis Set

The PD-L1-positive analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1-positive as defined by the VENTANA PD-L1 (SP263) IHC assay. The cut-off level to determine the PD-L1-positive analysis set will be the same as that for stratification purposes (≥25% PD‑L1–membrane expression in tumor tissue).

3.1.4 Safety Analysis Set

The Safety Analysis Set (SAS) will consist of all patients who received at least 1 dose of study treatment. Safety data will not be formally analyzed but summarized using the safety analysis set according to the treatment received, that is, erroneously treated patients (e.g., those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received.

[redacted]

Table 6 Summary of Outcome Variables and Analysis Populations

Outcome variable	Population
Efficacy data	

Table 6 Summary of Outcome Variables and Analysis Populations

Outcome variable	Population
OS	Full analysis set (ITT population)
PFS, PFS2, ORR, TTR, TFST, TSST, DoR, DCR, AFP6, APF12, OS12, OS18, OS24, PROs, and symptom endpoints	Full analysis set (ITT population)
OS, PFS, ORR, TTR,	PD-L1-negative analysis set
OS, PFS, ORR, TTR,	PD-L1-positive analysis set
Demography	Full analysis set (ITT population)
EORTC QLQ-C30 and H&N35	Full analysis set
[redacted]	[redacted]
Safety Data	
Exposure	Safety analysis set
AEs	Safety analysis set
Laboratory measurements	Safety analysis set
WHO performance status	Safety analysis set
Vital signs	Safety analysis set

AE Adverse event; APF6(12) Proportion of patients alive and progression free at 6 (12) months; DoR Duration of response; ITT Intent-to-Treat; ORR Objective response rate; OS Overall survival; OS12(18, 24) Proportion of patients alive at 12 (18, 24) months after randomization; PD-L1 Programmed cell death ligand 1; PFS Progression-free survival; PFS2 Second progression; [redacted] PRO Patient-reported outcome; TTR Time to response; TFST Time to first subsequent therapy; TSST Time to second subsequent therapy

3.2 Violations and Deviations

The important deviations specified below will be listed and summarized. None of the deviations will lead to patients being excluded from the analysis sets described in Section 2.1. If the deviations are serious enough to have the potential to impact the primary analysis, sensitivity analyses may be performed. 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 randomized to the study.

The deviations below will be programmatically derived.

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.

- Patients who deviate from key entry criteria (Deviation 1). These are inclusion criteria 3, 4, 6, 8 and exclusion criteria 1, 7, 9 as per clinical study protocol D4193C00002, fifth edition, dated 07SEP2016.

  - Inclusion criteria:

3) Histologically or cytologically confirmed recurrent or metastatic SCCHN (oral cavity, oropharynx, hypopharynx, or larynx) not amenable to therapy with curative intent (surgery or radiation therapy with or without chemotherapy). Patients who refuse radical resection are eligible.

4) Tumor progression or recurrence during or after only one palliative systemic treatment regimen for recurrent or metastatic disease that must have contained a platinum agent OR progression within 6 months of the last dose completion of platinum given as part of containing multimodality therapy with curative intent.

6) Confirmed PD-L1-positive or -negative SCCHN by the VENTANA PD-L1 (SP263) IHC assay on newly acquired tumor tissue (preferred) or archival tissue (≤3 years old).

- If the patient's PD-L1 status has already been assessed using the VENTANA PD-L1 (SP263) IHC assay as a part of the screening process for another AstraZeneca/MedImmune study, this test result can be used for the determination of eligibility, provided the PD-L1 status was obtained on tissue within the last 3 years.

- Note: A positive PD-L1 sample is measured using a defined cut-off based on ≥25% of tumor cells with membrane staining of any intensity for PD-L1. A negative PD-L1 sample is defined as <25% of tumor cells with membrane staining for PD-L1.

8) At least 1 lesion, not previously irradiated, that can be accurately measured at baseline as ≥10 mm in the longest diameter (except lymph nodes which must have a short axis ≥15 mm) with CT or MRI and that is suitable for accurate repeated measurements as per RECIST 1.1 guidelines. Lesions in a previously irradiated field can be used as measurable disease provided that there has been demonstrated progression in the lesion.

- Exclusion criteria:

1) Histologically or cytologically confirmed squamous cell carcinoma of any other primary anatomic location in the head and neck not specified in the inclusion criteria, patients with SCCHN of unknown primary, and non-squamous histologies (e.g., nasopharynx or salivary gland).

7) Current or prior use of immunosuppressive medication within 14 days before the first dose of their assigned IP. The following are exceptions to this criterion:

-Intranasal, inhaled, topical steroids, or local steroid injections (e.g., intra-articular injection)

-Systemic corticosteroids at physiologic doses not to exceed 10 mg/day of prednisone or its equivalent

-Steroids as pre-medication for hypersensitivity reactions (e.g., CT scan pre-medication)

9) Active or prior documented autoimmune or inflammatory disorders (including inflammatory bowel disease, diverticulitis with the exception of a prior episode that has resolved or diverticulosis, celiac disease, or other serious GI chronic conditions associated with diarrhea; systemic lupus erythematosus; Wegener syndrome [granulomatosis with polyangiitis]; myasthenia gravis; Graves' disease; rheumatoid arthritis; hypophysitis; uveitis, etc.) within the past 3 years prior to the start of treatment. The following are exceptions to this criterion:

- Patients with vitiligo or alopecia

- Patients with hypothyroidism (e.g., following Hashimoto syndrome) stable on hormone replacement or any skin condition not requiring systemic treatment

- Baseline RECIST scan > 40 days before randomization (Deviation 2). Note that although the screening period for baseline RECIST assessment was 28 days, an additional 12 day window should be applied thus only baseline RECIST assessments of greater than 40 days will be deemed as constituting an important deviation.

- No baseline RECIST 1.1 assessment on or before date of randomization (Deviation 3).

- Received prohibited systemic anti-cancer agents (Deviation 4). Please refer to the CSP section 7.7 for the systemic anti-cancer agents that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review prior to database lock.

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 will be made prior to database lock. For example, details of disallowed concomitant medication use will be reviewed by a physician and may be determined as key.

4. PRIMARY AND SECONDARY VARIABLES

4.1 Derivation of RECIST Visit Responses

For all patients, the RECIST version 1.1 (see further Appendix F of 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. RECIST 1.1 assessments will be performed using CT/MRI assessments of the neck (from base of skull) through the chest and abdomen (including liver). Additional anatomy should be imaged based on signs and symptoms of individual patients at baseline and follow-up.

The baseline assessments should be performed no more than 28 days before start of study treatment and ideally should be performed as close as possible to the start of study treatment (Tables 2 and 3 of CSP). Follow-up assessments will be performed every 8 weeks for the first 48 weeks (relative to the date of randomization) and then every 12 weeks as indicated in the schedule of procedures presented in the protocol (Tables 2 to 5 of CSP). Follow-up assessments will be performed every 8 weeks for the first 48 weeks (relative to the date of randomization) and then every 12 weeks as indicated in the schedule of procedures presented in the protocol (Tables 2 to 5 of CSP) until confirmed objective disease progression per RECIST 1.1. The confirmatory scans should preferably be performed at the next scheduled visit (relative to the date of randomization) and no less than 4 weeks after the initial assessment of PD (in the absence of clinically significant deterioration) and prior to initiation of alternate therapy. Confirmation of progression in patients in the SOC arm is preferable and should be done if clinically feasible. If an unscheduled assessment was performed and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits (relative to the date of randomization). All confirmatory scans should be recorded on the database.

Progression would be considered confirmed if the following criteria are met:

- ≥20% increase in the sum of diameters of target lesions (TL) compared with the nadir at 2 consecutive visits with an absolute increase of 5 mm.

- The assessment of progression of ≥20% increase in the sum diameters of target lesions compared with the nadir is at the first progression time point relative to the nadir (the smallest sum of diameters and this may be at baseline or subsequent follow-up visit). The confirmed scan confirms the persistence of the ≥20% increase relative to the nadir.

- And/or significant progression (worsening) of non-target lesions (NTL) or new lesions at the confirmatory PD time point compared with the first time point where progression of non-target lesions or new lesions identified.

- And/or additional new unequivocal lesions at the confirmatory PD time point compared with the first time point new lesions were identified.

In the absence of significant clinical deterioration, the investigator should continue IMT treatment until progression is confirmed. If progression is not confirmed, then the patient should continue on IMT treatment and on treatment assessments. Treatment through PD in the Standard of Care group is at the Investigator's discretion; however, a confirmatory scan is required for all patients in the Standard of Care group even if a subsequent treatment is started.

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

<u>IMT Treatment groups</u>

Patients will have scans every 8 weeks while on treatment (relative to the date of randomization) until study treatment is stopped. Patients with rapid tumor progression or with symptomatic progression that requires urgent medical intervention (e.g., central nervous system metastasis, respiratory failure due to tumor compression, or spinal cord compression) will not be eligible to continue to receive study treatment. IP should be discontinued if there is confirmed PD while receiving IMT following a previous response (CR or PR) to IP in the target lesions (i.e., the response and progression events both occurred while receiving active IP during the same treatment period in the target lesions).

RECIST outcomes will be calculated using a computer program for site investigator data.

4.1.1 Site Investigator Assessment Using RECIST 1.1

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

At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, or PD depending on the status of their disease compared with baseline and previous assessments. Baseline will be assessed within the 28 days prior to randomization. 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 objective disease progression according to RECIST 1.1 in which case the response will be assigned as PD). Imaging and procedures performed before signing the ICF may be used for screening purposes if the patient consents.

4.1.1.1 Site Investigator Assessment Using RECIST 1.1: Target Lesions (TLs)

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is $\geq$ 10 mm in the longest diameter (except lymph nodes which must have short axis $\geq$ 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 has the latest date prior to and including randomization will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

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

Measurable disease (i.e. at least one TL) is one of the entry criteria for the study. However, if a patient with non-measurable disease is enrolled in the study, the evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new

lesions (see Section 4.1.1.2 for further details). If a patient does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

Table 7 TL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all target lesions since baseline. Any pathological lymph nodes selected as target lesions must have a reduction in short axis to <10 mm.
Partial Response (PR)	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum of diameters.
Progressive Disease (PD)	At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm.
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD
Not Evaluable (NE)	Only relevant in certain situations (i.e. if any of the target lesions were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a target lesion response
Not Applicable (NA)	No target lesions are recorded at baseline

Rounding of TL data

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

Missing TL data

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

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

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

Lymph nodes

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

TL visit responses subsequent to CR

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

- Step 1: If all lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node LD increases by 20% but remains < 10 mm.

- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to NE irrespective of whether when referencing the sum of TL diameters, the criteria for PD is also met.

- Step 3: If not all lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD

- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR

TL too large to measure

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

TL too small to measure

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

Lesion intervention

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

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

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and scale up as described below, as long as there remain ≤ 1/3 of the TLs with missing measurements. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then if appropriate, a scaled sum of diameters will be calculated (as long as ≤ 1/3 of the TLs with interventions), and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention also has a value of 0 recorded.

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 lesion intervention)

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

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

Example of scaling

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

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

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

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

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

Lesions that split in two or more parts

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

Change in method of assessment of TLs

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

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

4.1.1.2 Non-Target Lesions (NTLs) and New lesions.

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

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

Table 8 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.

Table 8 NTL visit responses

Visit Responses	Description
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 discontinuation of therapy. 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.

If a new lesion is equivocal, for example because of its small size, the treatment and tumor assessments should be continued until the new lesion has been confirmed. If repeat scans confirm there is a new lesion, then the progression date should be declared using the date of the initial scan.

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 and should be treated as NE in the derivation of overall visit response.

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.

4.1.1.3 Site Investigator Assessment Using RECIST 1.1: Overall Visit Response

Table 9 Overall Visit Responses how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 9 Overall Visit Responses

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

CR Complete response, PR Partial response, SD Stable disease, PD Progression of disease, NE Not evaluable, NA Not applicable (only relevant if there were no TL/NTL at baseline).

4.2 Outcome Variables

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

RECIST outcomes will be derived using the programmatically derived overall visit response from investigator RECIST assessments.

4.2.1 Primary Endpoint: Overall Survival (OS)

Overall Survival (OS) is defined as the time from the date of randomization until death due to any cause. 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.

Note: Survival calls will be made in the week following the date of data cut-off for the analysis, and if patients are confirmed to be alive or if the death date is post the data cut-off date, these patients will be censored at the date of data cut-off. Death dates may be found by checking publicly available death registries.

If a patient is known to have died but only a partial date is available then the date will be imputed and used for derivations as follows:

- Missing day - Impute the 1st of the month. unless month is same as month of first dose of study drug then impute first dose date
- Missing day and month – impute 1st January.

If the patient is known to be alive after an imputed death date. The date of death will be imputed as last date known to be alive + 1.

4.2.2 Secondary Endpoints

4.2.2.1 Progression-Free Survival (PFS)

PFS (per RECIST 1.1 as assessed by the site investigator) 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 therapy or receives another anticancer therapy prior to progression (i.e., date of event or censoring - date of randomization +1). Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the 2 missed visits. Given the scheduled visit assessment scheme (i.e. assessments are scheduled to occur every 8 weeks), two missing visits will equate to 18 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks). If the patient has no evaluable visits or does not have baseline data, they will be censored at Day 1 unless they die within 2 visits of randomization, then they will be treated as an event with date of death as the event date.

Given the scheduled visit assessment scheme (i.e. eight-weekly), two missing visits will equate to 18 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks).

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 investigator assessments, the date of progression will be determined based on the earliest of the RECIST 1.1 assessment/scan dates of the component that indicates objective disease progression according to RECIST 1.1.

- 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.

In the absence of significant clinical deterioration, the investigational site is advised to continue the patient on their randomized MEDI4736 + tremelimumab combination therapy, MEDI4736 monotherapy until objective disease progression according to RECIST 1.1. If progression is not confirmed, the patient should continue their randomized MEDI4736 + tremelimumab combination therapy or MEDI4736 monotherapy treatment and on-treatment assessments. Treatment through PD in the SoC group is at the Investigator's discretion; however, a second scan to confirm PD is requested for all patients in the SoC group, if clinically feasible.

4.2.2.2 Objective Response Rate (ORR)

ORR (per RECIST 1.1 as assessed by the site investigator) is defined as the number (%) of patients with at least 1 visit response of CR or PR and will be based on all randomized patients. Therefore, data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of ORR.

Patients with stable disease, who go off treatment to receive a subsequent therapy and then respond, will not be included as responders in the ORR.

Patients who continue IMT through progression and respond during the IMT re-challenge period would not be included as responders in ORR assessment.

ORR will be obtained using the algorithm described above for the RECIST site Investigator tumor data. The denominator for ORR will be all randomized patients (i.e., the FAS population).

4.2.2.3 Duration of Response (DoR)

DoR (per RECIST 1.1 as assessed by the site investigator) will be defined as the time from the date of first documented response until the first date of documented progression or death in the absence of disease progression. 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 above).

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.

DoR will be obtained using the algorithm described above for the RECIST site Investigator tumor data.

4.2.2.4 Disease Control Rate (DCR)

DCR at 6 months is defined as the Percentage of patients who have a best objective response (BoR) of CR or PR in the first 6 months or who have demonstrated SD for a minimum interval of 24 weeks (-7 days, i.e., 161 days) following the start of treatment with IP. As additional analysis, DCR at 6 months will also be assessed as percentage of patients who have a best objective response (BoR) of CR or PR in the first 6 months or who have demonstrated SD for a minimum interval of 16 weeks (-7 days, i.e., 105 days) following the start of treatment with IP.

DCR at 12 months is defined as the percentage of patients who have a BoR of CR or PR in the first 12 months or who have demonstrated SD for a minimum interval of 48 weeks (-7 days, i.e., 329 days) following the start of treatment with IP.

DCR will be determined programmatically based on RECIST 1.1 site investigator's tumor data using all data up until the first progression event.

4.2.2.5 Proportion of Patients Alive and Progression Free at 6 Months (APF6)

The APF6 will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed by the site investigator) at 6 months.

4.2.2.6 Proportion of Patients Alive and Progression Free at 12 Months (APF12)

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

4.2.2.7 Proportion of Patients Alive at 12 Months (OS12)

The OS12 will be defined as the Kaplan-Meier estimate of OS at 12 months.

4.2.2.8 Proportion of Patients Alive at 18 Months (OS18)

The OS18 will be defined as the Kaplan-Meier estimate of OS at 18 months.

4.2.2.9 Proportion of Patients Alive at 24 Months (OS24)

The OS24will be defined as the Kaplan-Meier estimate of OS at 24 months.

4.2.2.10 Best Objective Response (BoR)

BoR is calculated based on the overall visit responses from each RECIST 1.1 assessment. It is the best response a patient has had during their time in the study up until objective disease progression according to RECIST 1.1.

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

BoR will be determined programmatically based on RECIST 1.1 using site Investigator data using all data up until objective disease progression according to RECIST 1.1.

For patients, whose progression event is death, BoR will be calculated based upon all evaluable RECIST 1.1 assessments prior to death.

For patients who die with no evaluable RECIST 1.1 assessments, if the death occurs ≤17 weeks (i.e., 16 Weeks ±7 days) after enrollment, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >17 weeks (i.e., 16 weeks ±7 days) after the date of enrollment then BoR will be assigned to the NE category.

Progression events that have been censored due to them being >17 weeks after the last evaluable assessment will not contribute to the BoR derivation.

4.2.2.11 Time to Response

Time to response (per RECIST 1.1 as assessed by the site investigator) is defined as the time from the date of first dose until the date of first documented response. The date of first documented response should coincide with that used for the RECIST 1.1 DoR endpoint.

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

4.2.2.12 Time from Randomization to First Subsequent Therapy or Death (TFST)

Time to the first subsequent therapy or death will be defined as the time from the date of randomization to the earlier of either the start date of the first subsequent anticancer therapy after discontinuation of randomized treatment or the date of death (i.e., the date of first subsequent cancer therapy, death, or censoring defined as the date of randomization + 1 day).

Any patient not known to have received a first subsequent anticancer therapy will be censored at the last date that the patient was known not to have received a first subsequent anticancer therapy. If a patient terminated the study before the first subsequent therapy for a reason other than death, the patient will be censored at the earliest of either the patient's last known date to be alive or the study termination date.

Patients not receiving randomized treatment would have TFST calculated in the same way, i.e. time from date of randomization to the subsequent anticancer therapy or death.

4.2.2.13 Time from Randomization to Second Subsequent Therapy or Death (TSST)

As a supportive summary to PFS2, time to second subsequent therapy (TSST) or death will be defined as the time from the date of randomization to the earlier of either the start date of the second subsequent anticancer therapy after discontinuation of first subsequent therapy, or the date of death. Any patient not known to have had a second subsequent anticancer therapy will be censored at the last date when the patient was known not to have received a second subsequent anticancer therapy. If a patient terminated the study for reason other than death before second subsequent anticancer therapy, the patient will be censored at the earliest of the last known to be alive and termination dates.

4.2.2.14 Time from Randomization to Second Progression-Free Survival (PFS2)

PFS2 will be defined as the time from the date of randomization to the earliest of the progression event subsequent to first subsequent therapy or death. 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. Second progression status will be reviewed (every 6 weeks for the first 24 weeks relative to the date of randomization and then every 8 weeks thereafter) following the progression event used for the primary variable PFS (the first progression) and status recorded. Patients alive and for whom a second disease progression has not been observed should be censored at the earliest of; date of withdrawal, date last known alive, DCO or, if a patient has not had a first subsequent therapy; the date last known not to have received a first subsequent therapy (TFST censoring date).

4.2.2.15 Change in Tumor Size

For supportive purposes percentage change from baseline in tumor size will be derived at each scheduled tumor assessment visit (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, if no decrease, as the smallest increase in tumor size from baseline.

This is based on RECIST1.1 target lesion measurements taken at baseline and at the timepoint of interest. Tumor size is defined as the sum of the longest diameters of the target lesions for the BICR data based upon RECIST assessments. Target lesions are measurable tumor lesions. Baseline for RECIST is defined to be the last 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. (week X - baseline) / baseline * 100). More details on target lesions and measurements can be found in Section 4.1.

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 ± 1 week of the protocol scheduled visit will be used for that visit.

4.3 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. Data from all cycles of treatment will be combined in the presentation of safety data. 'On treatment' will be defined as assessments between date of start dose and 90 days following discontinuation of IP (i.e., the last dose of MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy, or SoC). For AEs, on treatment (or treatment-emergent AEs) will be defined as any AEs that started after dosing or prior to dosing and which worsen following exposure to the treatment (see section 5.2.13.1 for a more detailed definition).

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

4.3.1 Adverse Events (AEs)

AEs and SAEs will be collected from the time of signature of informed consent throughout the treatment period and up to 90 days after the last dose of IP (MEDI4736 + tremelimumab, MEDI4736, or SOC) or until initiation of another therapy (excluding palliative radiation as anti-cancer therapy). Any AE occurring before treatment with IP will be included in the data listings but will not be included in the summary tables of AEs.

A separate data listing of AEs occurring more than 90 days after discontinuation of MEDI4736 in combination with tremelimumab or MEDI4736 monotherapy will be produced. These events will not be included in AE summaries.

The Medical Dictionary for Regulatory Activities (MedDRA) dictionary (using the latest or current MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 4.03 or higher).

AEs of Special Interest (AESIs)

Adverse events of special interest (AESIs) are events of scientific and medical interest specific to the further understanding of the MEDI4736 and tremelimumab safety profile and require close monitoring and rapid communication by the Investigator to the Sponsor. MEDI4736 and tremelimumab AESIs may be serious or non-serious. The rapid reporting of these AESIs allows ongoing analysis of these events in order to characterize and understand them in association with the use of these IPs.

The AESI's have been identified in the CSP (Version 6) as:

- Diarrhea/Colitis and intestinal perforation
- Pneumonitis/ILD
- Hepatitis/transaminase increases
- Endocrinopathies (i.e., events of hypophysitis/hypopituitarism, adrenal insufficiency, and hyperthyroidism and hypothyroidism, and type I diabetes mellitus)
- Rash/Dermatitis
- Nephritis/blood creatinine increases
- Pancreatitis/serum lipase and amylase increases
- Myocarditis
- Myositis/polymyositis
- Neuropathy/neuromuscular toxicity (eg, Guillain-Barré, and myasthenia gravis)
- Other inflammatory responses that are rare/less frequent with a potential immune-mediated etiology include, but are not limited to, pericarditis, sarcoidosis, uveitis and other events involving the eye, skin, hematological and rheumatological events.

An AstraZeneca medically qualified expert, after consultation with the Global Patient Safety Physician, has reviewed the AEs of special interest and identified which preferred terms contribute to each AESI. A further review will take place prior to database lock to ensure new terms not already included in the older MedDRA version are captured within the categories for the new higher MedDRA version. The list will be provided by AZ prior to database lock.

4.3.2 Treatment Exposure

Exposure will be defined separately for MEDI4736 monotherapy, MEDI4736 on the MEDI4736+ tremelimumab combination arm, tremelimumab on the MEDI4736+tremelimumab combination arm including retreatment, and standard of care as follows:

Total (or intended) exposure of MEDI4736 (monotherapy)

- Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of “last dose date of study drug + 13 days” or death date or DCO.

Total (or intended) exposure of MEDI4736 (combination)

- Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of “last dose date of study drug + (13 days or 27 days)” or death

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

Total (or intended) exposure of tremelimumab (combination)

- Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of "last dose date of study drug + 27 days" or death date or DCO. For patients who are retreated; the time from last dose prior to retreatment until first dose of retreatment will be excluded from the total duration.

Actual exposure of MEDI4736/tremelimumab
- Actual exposure is defined as above, but excluding total duration of dose delays.

Total (or intended) exposure for the SOC treatments
- The total (or intended) exposure for the SOC treatments 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 not be calculated for SOC.

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

Exposure will also be measured by the number of cycles received. For the immunotherapy and SOC arms, a cycle corresponds to a 4 week 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. Each immunotherapy agent will be measured in terms of number of doses given.

Patients who permanently discontinue during a dose interruption

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

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

MEDI4736 (monotherapy):

- Since Patients in the MEDI4736 monotherapy treatment group will receive 10 mg/kg MEDI4736 via IV infusion q2w until disease progression, the duration of dose delays will be calculated as follow:

  For all dosing dates:
  Duration of delay = Sum of (Date of the dose - Date of previous dose – 14 days)

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

MEDI4736 (given in combination):

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

  For Cycle 2 to Cycle 5 (for Week 4 to Week 16) doses:

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

  For Cycle 5 and beyond (for Week 18 and beyond) doses:

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

  Duration of delay = Duration1 + Duration2

Tremelimumab (given in combination):

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

  For Cycle 2 to Cycle 4 (for Week 4 to Week 12) doses:

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

4.3.3 Dose Intensity

Dose intensity will be derived for the initial treatment period and the re-treatment period separately for the immunotherapy agents. It will also be derived for the SOC agents. Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation.

Relative dose intensity (RDI) will be defined as follows for MEDI4736, tremelimumab and all Standard of Care therapy:

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

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

4.3.4 Laboratory Data

Laboratory data will be collected throughout the study, from screening to the follow-up visits as described in Tables 2, 3, 4 and 5 of 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 derivation of post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 4.3.7 below will be used.

For the change from baseline summaries for laboratory data, the baseline value will be the latest result obtained prior to the start of study treatment. This can be on the same day of first dose where time is prior to first dose.

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 SI units. The following parameters have CTC grades defined for both high and low values: Potassium, Sodium, Magnesium, Glucose and Corrected calcium so high and low CTC grades will be calculated.

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

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

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

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

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 needs only to have 1 post dose-value recorded.

4.3.5 ECGs

ECG data obtained up until the 30 days from date of last dose of study treatment will be used for reporting. The Investigator's assessment of the ECG will be collected locally.

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

4.3.6 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 4.3.7 below will be used.

For the change from baseline summaries for vital signs, the baseline value will be the latest result obtained prior to the start of study treatment. This can be on the same day of first dose where time is prior to first dose.

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

4.3.7 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 1). 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. Note that in protocol Table 2 and Table 3, Day 0 is mentioned as the first day after screening (Day -28 to -1). Day 0 should be considered the same as reporting Day 1. The Statistical windows only referred to reporting Day 1 to avoid confusion.

For example, the visit windows for vital signs data in the MEDI4736 monotherapy are shown in the example below.

Visit	Week	Day	Statistical Window
Screening	NA	-28 to -1	D-28 – D-1
Baseline	NA	D0[1]	Low – D1[1]
V0	W0	D0	D1[2] – D7
V1	W2	D14	D8 – D21
V2	W4	D28	D22 – D35
V3	W6	D42	D36 – D49
V4	W8	D56	D50 – D63
V4a	W10	D 70	D64 – D77
V5	W12	D84	D78 – D91
V5a	W14	D 98	D92 – D105
V6	W16	D112	D106 – D119
V6a	W18	D 126	D120 – D133
V7	W20	D140	D134 – D147
V7a	W22	D 154	D148 – D161
V8	W24	D168	D162 – D175
V8a	W26	D182	D176 – D189
V9	W28	D196	D190 – D203
V9a	W30	D210	D204 – D217
V10	W32	D224	D218 – D231
V10a	W34	D238	D232 – D245
V11	W36	D252	D246 – D259
V11a	W38	D266	D260 – D273
V12	W40	D280	D274 – D287
V12a	W42	D294	D288 – D301
V13	W44	D308	D2302 – D315
V13a	W46	D322	D316 – D329
V14	W48	D336	D330 – D343
V15	W50	D350	D344 – High[3]

[1]Latest pre-dose value available
[2]post-dose values only
[3] Note that visits up to 13 days after the last dosing date will be considered as being on treatment for the purposes of visit windowing and may be assigned to an on-treatment visit. Visits after this will be considered as follow-up and may be assigned accordingly

- 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. 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 re-treatment period then baseline is similarly defined as the last non-missing measurement prior to the first dose on the re-treatment period. For laboratory data, any assessments made on day 1 will be considered pre-dose. Alternatively, if two visits are equally eligible to assess patient status at baseline (e.g., screening and baseline assessments both on the same date prior to first dose with no washout or other intervention in the screening period), the average can be taken as a baseline value. For non-numeric laboratory tests (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 study 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.

- For missing diagnostic dates, if day and/or month are missing use 01 and/or Jan. If year is missing, put the complete date to missing.

- For missing AE start date the following will be applied:

  - Missing day - Impute the first day of the month unless month is the same as month of the first dose of study drug then impute first dose date.

  - Missing day and month – impute 1st January unless year is the same as first dose date then impute first dose date.

  - Completely missing date – impute the first dose date.

- For missing end AE dates, the following will be applied:

- Missing day - Impute the last day of the month unless month is the same as month of the first dose of study drug then impute last dose date.

- Missing day and month – impute 31st December unless year is the same as first dose date then impute last dose date.

- Completely missing date – need to look at whether the AE/medication is still ongoing before imputing a date and also when it started in relation to study drug. If the ongoing flag is missing, then assume that AE is still present / medication is still being taken (i.e. do not impute a date). If the AE/medication has stopped and start date is prior to first dose date then impute the 1st dose date, if it started on or after first dose date then impute a date that is after the last dose date.

4.4 Patient reported outcome

PROs will be assessed using the EORTC QLQ-C30, EORTC QLQ-H&N35 [redacted]. All items/questionnaires will be scored according to published scoring guidelines. All PRO analyses will be based on the Full Analysis Set (FAS).

4.4.1 EORTC QLQ-C30

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

The change from baseline in HRQoL will be assessed using the EORTC QLQ-C30 global health QoL scale, which includes 2 items from the EORTC QLQ-C30: “How would you rate your overall health during the past week?” (Item 29), and “How would you rate your overall QoL during the past week?” (Item 30).

Definition of clinically meaningful changes

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

For each subscale, if <50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers et al 2001). If at least 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing.

The reason for any missing questionnaire will be identified and recorded. If there is evidence that the missing data are systematic, missing values will be handled to ensure that any possible bias is minimized.

Table 10 Change from BL and Visit Response for EORTC QLQ C30

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

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

Time to symptom deterioration

For each of the symptoms scales in the EORTC QLQ-C30, time to symptom deterioration will be defined as the time from randomization until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration. Death will be included as an event only if the death occurs within two visits of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by EORTC QLQ-C30) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if symptoms deteriorate after two or more missed PRO assessment visits or the patient dies after two or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated (prior to the two missed assessment visits). If a patient has no evaluable visits or does not have baseline data they will be censored at 0 days. The population for the analysis of time to symptom deterioration will include a subset of the FAS who have baseline scores of ≤90.

Time to HRQoL/Function deterioration

For HRQoL, time to deterioration will be defined as the time from the date of randomization until the date of the first clinically meaningful deterioration (a decrease in the function scales or the global health status/QoL from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to HRQoL/function deterioration. Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the HRQoL/function change could be evaluated.

Patients whose HRQoL (as measured by EORTC QLQ-C30) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the HRQoL/function could be evaluated. Also, if HRQoL deteriorates after two or more missed PRO assessment visits or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where HRQoL/function could be evaluated. If a patient has no evaluable visits or does not have baseline data they will be censored at 0 days. The population for the analysis of time to HRQoL/function deterioration will include a subset of the FAS who have baseline scores of ≥10.

If a patient did not completed the questionnaire with reason being; 'Subject too heavily affected by symptoms of disease under investigation' then this will be counted as a deterioration event.

Symptom improvement rate

The symptom improvement rate will be defined as the number (%) of patients with 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (a decrease from baseline score ≥10 for EORTC QLQ-C30 symptom scales) in that symptom from baseline. The denominator will consist of a subset of the FAS who have a baseline symptom score ≥10.

HRQoL/function improvement rate

The HRQoL/function improvement rate will be defined as the number (%) of patients with 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (an increase from baseline score ≥10 for EORTC QLQ-C30 functional scales and global health status/QoL) in that scale from baseline. The denominator will consist of a subset of the FAS who have a baseline HRQoL/function score ≥10.

PRO Compliance

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

- Received forms=number of EORTC QLQ-C30 forms plus the number not completed where the reason was 'Subject too heavily affected by symptoms of disease under investigation'

- Expected forms= number of patients still under HRQL follow-up at the specified assessment time excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under HRQL follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.

- Evaluable forms = subset of the expected EORTC QLQ-C30 forms with at least one subscale that can be determined; or where reason for non-completion was documented as 'Subject too affected by symptoms of disease under investigation'.

Thus, the overall compliance rate is defined as number of patients with an evaluable baseline and at least one evaluable follow-up form (as defined above), divided by the number of patients expected to have completed at least a baseline EORTC QLQ-C30 form.

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

4.4.2 EORTC QLQ-H&N35

The H&N35 is a head and neck cancer-specific module from the EORTC for head and neck cancer comprising 35 questions to assess head and neck cancer symptoms. The head and neck cancer module includes 11 single items and 7 multi-item scales that assess pain, swallowing, senses (taste and smell), speech, social eating, social contact, and sexuality. For all items and scales, high scores indicate increased symptomatology/more problems.

The scoring approach for the H&N35 is identical in principle to that for the symptom scales/single items of the EORTC QLQ-C30. As the wording is reversed on the H&N35, higher scores represent greater symptom severity.

Definition of clinically meaningful changes

Changes in score compared with baseline will be evaluated. The developers of the H&N35 have suggested that a minimum clinically meaningful change is a change in the score from baseline of >10 for scales/items from the H&N35 module (Bjordal et al 2000). For example, a clinically meaningful deterioration or worsening in dry mouth (as assessed by H&N35) is defined as an increase in the score from baseline of ≥10. At each post-baseline assessment, the change in symptoms/functioning from baseline will be categorized as improved, no change, or deterioration, as shown in Table 11. Since there is no well-established minimal clinically important difference for the H&N35 module, an exploratory analysis will be conducted to determine the most appropriate threshold in this patient population.

Table 11 Change from BL and Visit response for EORTC QLQ H&N35

Score	Change from baseline	Visit response
H&N35 symptom scales and items	≥+10	Deterioration
	≤-10	Improvement
	Otherwise	No change

HRQoL Health-related quality of life; H&N35 35-item head and neck quality of life questionnaire.

Time to symptom deterioration

For each of the symptom scales/items in the H&N35, time to symptom deterioration will be defined as the time from the date of the first dose until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration. Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by the H&N35) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if symptoms progress after 2 or more missed PRO assessment visits or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated. If a patient has no evaluable visits or does not have baseline data, they will be censored at 0 days. The population for analysis of time to symptom deterioration will include a subset of the FAS population who have baseline scores ≤90.

Symptom improvement rate

The symptom improvement rate will be defined as the number (%) of patients with 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (a decrease from baseline score >10 for H&N35 scales/items) in that symptom from baseline.

PRO Compliance

Summary measures of overall compliance and compliance over time will be derived for the EORTC QLQ-H&N35. These will be based upon the compliance derivation described for EORTC QLQ-C30.

[redacted]

[redacted]

[redacted]

4.6 Biomarker Variables

Biomarker status will be assessed for evaluable patients according to following criteria:

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

Additionally, a classification using a 1% cut point will be explored with the following criteria:

- Positive: ≥ 1% tumor cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control.
- Negative: < 1% tumor cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control.

[redacted]

5. ANALYSIS METHODS

Formal statistical analysis will be performed to test the main hypotheses:

- H_0: No difference in OS between either MEDI4736 + tremelimumab combination therapy **or** MEDI4736 monotherapy and SoC in all patients, regardless of PD-L1 status
- H_1: Difference in OS between either MEDI4736 + tremelimumab combination therapy **or** MEDI4736 monotherapy and SoC in all patients, regardless of PD-L1 status

The study will be considered positive if OS for either co-primary objective (MEDI4736 + tremelimumab versus SoC or MEDI4736 monotherapy versus SoC) is statistically significant. The analysis of OS will be performed when:

- Approximately 375 death events have occurred in 480 patients (78% maturity) across the MEDI4736 + tremelimumab combination therapy and SoC arms regardless of PD-L1 status AND when approximately 375 death events have occurred in 480 patients (78% maturity) across the MEDI4736 monotherapy and SoC arms, regardless of PD-L1 status.

Interim analysis for OS will be performed when 300 death events (80% of required events) have been accumulated across the MEDI4736 + tremelimumab combination therapy and SoC arms. It is expected that approximately 300 death events would have accumulated across the MEDI4736 monotherapy and SoC arms at this time.

5.1 General Principles

IVRS based stratification factors will be used for all analysis. If there is >10% discordance in stratification factors as recorded in IVRS versus the Case Report Form (CRF), then a sensitivity analysis of the primary endpoint OS will be performed using CRF based stratification factors.

Other general principles that will be followed throughout the study include the following:

- Descriptive statistics will be used for all variables, as appropriate, and will be presented by treatment group. Continuous variables will be summarized by the number of observations, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized by frequency counts and percentages for each category.

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

- For continuous data, the mean and median will be rounded to 1 additional decimal place compared to the original data. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.

- For categorical data, percentages will be rounded to 1 decimal place.

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

- SAS® version 9.1.3 or higher will be used for all analyses.

- Unless otherwise stated all PD-L1 analyses will use the 25% cut point as defined in Section 4.6.

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

Efficacy and PRO data will be summarized and analyzed based on the FAS. [redacted] Safety data will be summarized on the Safety Analysis Set.

Table 12 provides an overview of all pre-planned statistical and sensitivity analyses.

Table 12 Pre-planned statistical and sensitivity analyses to be conducted

Endpoints analyzed	Notes
Overall survival	Stratified log-rank test and Cox proportional model for: Co-primary objectives – MEDI4736 + tremelimumab combination therapy versus SoC for all patients, regardless of PD-L1 status (stratified for PD-L1 status, tumor location/HPV status, and smoking status) – MEDI4736 monotherapy versus SoC for all patients, regardless of PD-L1 status (stratified for PD-L1 status, tumor location/HPV status, and smoking status) Secondary objectives: – MEDI4736 + tremelimumab combination therapy versus SoC in PD-L1 negative patients (stratified for tumor location/HPV status, and smoking status) – MEDI4736 monotherapy versus SoC in PD-L1 positive patients (stratified for tumor location/HPV status, and smoking status) – MEDI4736 + tremelimumab combination therapy compared to MEDI4736 monotherapy in PD-L1 negative patients (stratified for tumor location/HPV status, and smoking status) Sensitivity analysis: Kaplan Meier plot of time to censoring where the censoring indicator of the primary analysis is reversed – attrition bias Subgroup analyses as specified below
Progression free survival	A similar analysis will be conducted as described above for overall survival. No subgroup analyses needed
Proportion of patients alive at 12 months, 18 months, and 24 months	Kaplan-Meier estimates of survival at 12 months, 18 months, and 24 months
Objective response rate	Logistic regression using site Investigator assessment (RECIST 1.1)
Duration of response	Descriptive statistical and Kaplan Meier plots
Proportion of patients alive and progression free at 6 and 12 months	Kaplan Meier estimates of progression free survival at 6 and 12 months
Time to treatment Response	Kaplan-Meier plots using site Investigator data (RECIST 1.1)
Time to first and second subsequent therapy (TFST, TSST)	Stratified log-rank test
Time from randomization to second progression (PFS2)	Stratified log-rank test
Disease control rate	Logistic regression using site Investigator data (RECIST 1.1)
Best objective response	N (%) using site Investigator data (RECIST 1.1)
Symptom improvement rate (EORTC QLQ-C30 and EORTC QLQ-H&N35 endpoints)	Logistic regression

Table 12 Pre-planned statistical and sensitivity analyses to be conducted

Endpoints analyzed	Notes
QoL/Function improvement rate (EORTC QLQ-C30 endpoints)	Logistic regression
Time to QoL/Function deterioration (EORTC QLQ-C30 endpoints)	Stratified log-rank test
Time to symptom deterioration (EORTC QLQ-C30 and EORTC QLQ-H&N35 endpoints)	Stratified log-rank test
[redacted]	[redacted]

CI Confidence interval; irRC Immune-related response criteria; RECIST 1.1 Response Evaluation Criteria in Solid Tumors version 1.1; OS Overall survival.

5.1.1 Multiple Testing Strategy

In order to strongly control the type I error at 5% (2-sided), a Multiple Testing Procedure (MTP) will be used across the co-primary objectives (OS in MEDI4736 + tremelimumab combination therapy versus SoC and OS in MEDI4736 monotherapy versus SoC), and across the analysis populations (all patients regardless of PD-L1 status and PD L1-negative population). If the highest level hypothesis in the MTP is rejected for superiority, the remaining hypothesis will then be tested as shown in Figure 4 below.

Hypotheses will be tested using an MTP with an alpha recycling strategy (Burman et al 2009). The levels of the MTP are shown in Figure 4. Of note, the comparisons of MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy will not be included in the MTP, and therefore will not be conducted under strict alpha control.

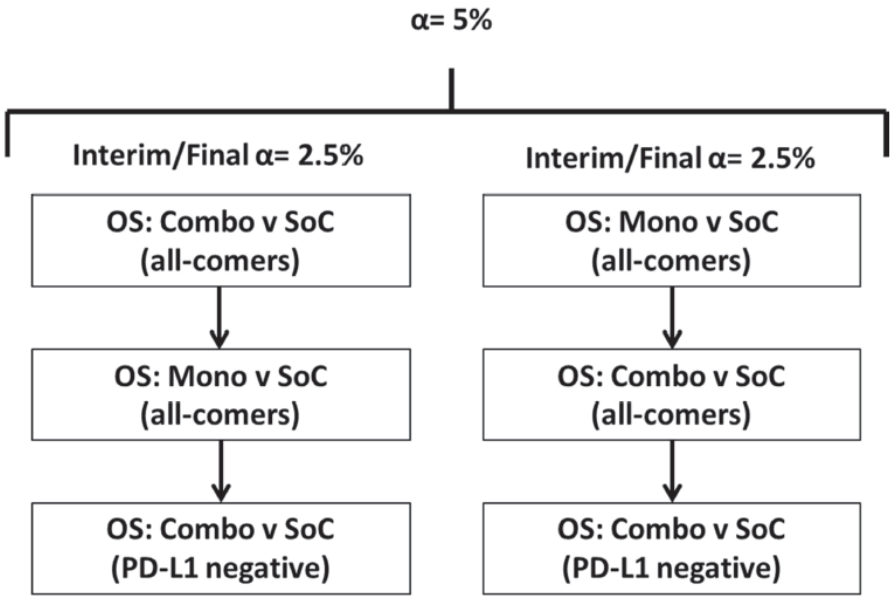
According to alpha (test mass) splitting and alpha recycling, the test mass that becomes available after each rejected hypothesis is recycled to the next hypotheses not yet rejected. Since OS is tested at multiple timepoints (i.e., 1 interim analysis and final analysis) the OS tests for the same comparison/population (i.e., shown in box 2 in the MTP) will be considered as 1 test family. As long as 1 test in the family can be rejected, the family is rejected, thus, the assigned total alpha to the family can be recycled to next MTP level.

The testing procedure stops when the entire test mass is allocated to a non-rejected hypothesis in the multiple testing procedure. Implementation of this pre-defined ordered testing procedure, including recycling, will strongly control type I error at 5% (2-sided), among all key hypotheses.

There will be one interim analysis of OS, conducted after approximately 80% of the target death events have been observed across MEDI4736 + tremelimumab combination therapy and SoC arms, and across MEDI4736 monotherapy and SoC arms, regardless of PD-L1 status. The alpha level allocated to OS for MEDI4736 + tremelimumab combination therapy versus

SoC and MEDI4736 monotherapy versus SoC (all patients, regardless of PD-L1 status), will be controlled at the interim and final analysis time points by using the Lan DeMets spending function (Lan and DeMets 1983) that approximates an O'Brien Fleming approach, where the alpha level applied at the interim depends upon the proportion of information available. If statistically significant, then testing for OS will continue in the PD-L1-negative population. A separate spending function (based on observed information fraction in this subgroup) will be used to adjust the alpha levels at the interim and final analyses in the PD-L1-negative population.

Figure 4 Levels of the Multiple Testing Strategy



Combo: MEDI4736 + tremelimumab.
Mono: MEDI4736 monotherapy.
OS Overall survival; PD-L1 negative Patients with PD-L1–negative tumors; SoC Standard of Care.
No other study objectives will be considered in the formal multiple testing procedure.

5.2 Analysis Methods

5.2.1 Overall Survival (OS)

Analysis of the Co-Primary Variables

The primary analysis of OS in MEDI4736 + tremelimumab and MEDI4736 monotherapy, in all patients, regardless of PD-L1 status will be done using a stratified log-rank test stratified by PD-L1 status (positive and negative), tumor location/HPV status (oropharyngeal cancer with HPV positive status, oropharyngeal cancer with HPV negative status, and non-oropharyngeal cancer regardless of HPV status), and smoking status (>10 and ≤10 packyears). The effect of treatment will be estimated by the HR together with its corresponding 95 % CI and p-value, using the Breslow approach for handling ties (Breslow 1974), as described in the multiple strategy (Section 5.1).

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

The HR and its CI will be estimated from a stratified Cox proportional hazards model (Cox 1972) with treatment as the only factor and the same strata information as above (with ties = Efron and any stratification variables included in the strata statement).

Secondary Variables

OS will be analyzed in PD-L1-negative patients using a stratified log-rank test, adjusting for tumor location/HPV status, and smoking status.

For the comparisons of MEDI4736 + tremelimumab combination therapy versus SoC in PDL1-negative patients, treatment effects will be estimated by the HR together with their corresponding 95% CIs and p-values using the appropriate alpha, as described in the multiple strategy above (Section 5.1). The HR and its CI will be estimated from a stratified Cox proportional hazards model (Cox 1972) with treatment as the only factor and the same strata information as above.

Kaplan-Meier plots of OS will be presented by treatment arm. Summaries of the number and percentage of patients who have died, those still in survival follow-up, those lost to follow-up, and those who have withdrawn consent will be provided along with the median OS for each treatment.

A similar analysis will be conducted to compare MEDI4736 monotherapy versus SoC in the PD-L1 positive patients as described above.

In addition, a secondary analysis of OS will be performed to compare MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy in the PD-L1-negative population as described above.

Analysis of OS in Subgroup Populations

For each one of the following subpopulations:

- PD-L1 status (positive, negative) from eCRF using a 25% cut point
- PD-L1 status (positive, negative) using a 1% cut point
- Tumor location/HPV Status (oropharyngeal cancer with HPV positive status, oropharyngeal cancer with HPV negative status, and non-oropharyngeal cancer regardless of HPV status) from eCRF
- Primary tumor site (oral cavity, oropharynx, hypopharynx and larynx)
- Smoking Status (>10 and ≤10 pack-years) from eCRF

- Smoking history (current, former, never)
- Use of chewing tobacco, oral snuff, and sublingual nicotine (Yes vs. No)
- Sex (male and female)
- Age at randomization (<65, ≥65 – 75, and ≥75 years of age)
- Race (Asian, non-Asian)
- ECOG Performance status (0 and ≥1)
- Prior radiation therapy (Yes vs. No)
- Time to recurrence from platinum-containing multimodality therapy (<6 months, ≥6 months or presentation with metastatic disease)
- Prior lines of systemic therapy (1, 2, and ≥3)
- Metastatic disease at baseline (stage IVc and other stages)
- Extent of Disease (recurrent and metastatic)
- Standard of Care (cetuximab, taxane, methotrexate, fluoropyrimidine-based regimen) if there is sufficient data for the analysis.

The comparisons will include MEDI4736 + tremelimumab combination therapy versus SoC and MEDI4736 monotherapy versus SoC for all patients regardless of PD-L1 status. For PD-L1 negative patients and PD-L1 positive patients, the comparisons will include MEDI4736 + tremelimumab combination therapy versus SoC MEDI4736 monotherapy versus SoC, and MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy. Treatment effect will be estimated by the HR together with its corresponding 95% CI using an unstratified Cox model with treatment as the only covariate. For standard of care subgroup analysis, only the SoC of interest will be included as a treatment factor (e.g. for comparison of MEDI4736 monotherapy versus cetuximab, the treatment factors will be MEDI4736 monotherapy and cetuximab).

Other baseline variables may also be assessed if there is clinical justification or if an imbalance is observed between the treatment arms. The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic and/or predictive factors. Only HR estimate along with 95% CI will be presented. No adjustment to the significance level for testing of the subgroup and sensitivity analyses will be made since all these analyses will be considered supportive of the analysis of OS.

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 the primary endpoint (OS) will not be formally analysed. In this case, only descriptive summaries will be provided.

A forest plot will be presented comparing OS for MEDI4736 + tremelimumab combination therapy vs. SoC (in all patients regardless of PD-L1 status and in PD-L1 negative patients) and MEDI4736 monotherapy vs. SoC (in all patients regardless of PD-L1 status) within each subgroup.

Analysis of OS12, OS18, and OS24

OS12, OS18 and OS24 will be summarized (using the Kaplan-Meier curve) and presented by treatment arm.

Assumptions of Proportionality

For OS presented by the overall population and by PD-L1 status, the assumption of proportionality will be assessed. Proportional hazards will be tested first by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time-dependent covariate to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect will be described by presenting piecewise HR calculated over distinct time periods. In such circumstances, the HR can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found, this may be a result of treatment-by-covariate interactions, which will be investigated.

Sensitivity Analysis

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

The number of patients prematurely censored will be summarized by treatment arm. A patient would be defined as prematurely censored if their survival status was not defined at the data cut-off. In addition, duration of follow-up will be summarized using medians:

- In censored patients who are alive at data cut-off only: Time from randomization to date of censoring (date last known to be alive) by treatment arm.
- In all patients: Time from randomization to the date of death (i.e. overall survival) or to the date of censoring for censored patients regardless of treatment arm.

Effect of covariates on the HR estimate (Cox Proportional Hazards model)

An unstratified Cox proportional hazards modeling will be employed to assess the effect of covariates on the HR estimate for the overall population. The following covariates will be included in the statistical model: treatment and the stratification factors as main effects (PD-L1 status, Tumor location/HPV Status (oropharyngeal cancer with HPV positive status, oropharyngeal cancer with HPV negative status, and non-oropharyngeal cancer regardless of HPV status), and smoking status).

This model will be done to ensure that any output from the Cox modeling is likely to be consistent with the results of the stratified log-rank test for the overall population.

Moreover, a sensitive analysis will be performed to evaluate the treatment effect, adjusted for pre-specified baseline prognostic factors. For this analysis, a stratified Cox adjusted for the following covariates will be used. They are:

- PD-L1 status
- Tumor location/HPV Status (oropharyngeal cancer with HPV status positive, oropharyngeal cancer with HPV status negative and non-oropharyngeal cancer regardless of HPV status)
- Smoking Status (>10 and ≤10 pack-years)
- Time to recurrence from multimodality therapy (<6 months versus ≥6 months or presentation with metastatic disease)
- Sex at randomization
- Age at randomization

The two models described above 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.

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

Impact of changing (crossover outside of this study) to immunotherapies (or other potentially active investigational agents) on overall survival analyses

Exploratory analyses of OS adjusting for the impact of subsequent immunotherapy or other investigational treatment may be performed if a meaningful proportion of patients change therapy (>15%). Methods such as Rank Preserving Structural Failure Time (Robins and Tsiatis 1991), Inverse Probability of Censoring Weighting (Robins 1993), and other methods in development will be explored. The plausibility of the underlying assumptions and impact of model selection will be investigated as sensitivities to the default methods. Baseline and time-dependent characteristics will be explored, and summaries of baseline characteristics will be generated by treatment arm, designating between those that have and haven't changed immunotherapies at the time of the analyses.

Further detail will be provided in the Payer Analysis Plan. These analyses are intended to support reimbursement appraisals. 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 according to line of subsequent therapy, i.e. immediately after immunotherapy or as a later line.

5.2.2 Progression Free Survival (PFS)

PFS analyses will be based on the programmatically derived RECIST 1.1 data using the Investigator tumor assessments. The data will be analyzed from all patients, regardless of PD-L1 status, using stratified log-rank test as described for OS analyses in section 5.2.1.

The effect of treatment will be estimated by the HR together with its corresponding CI and p-value. Kaplan-Meier plots of PFS 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) will be provided along with median PFS for each treatment.

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

The analysis will be repeated in the PD-L1 positive and negative groups adjusting for tumor location/HPV status, and smoking status.

5.2.3 Objective Response Rate (ORR)

The ORR will be based on the programmatically derived RECIST 1.1 using the site Investigator data. The ORR will be compared between MEDI4736 + tremelimumab combination therapy versus SoC and MEDI4736 monotherapy versus SoC using logistic regression models adjusting for the same factors as the co-primary variables. The results of the analysis will be presented in terms of an odds ratio together with its associated profile likelihood 95% CI and p-value ((e.g. using the option 'LRCI' in SAS procedure GENMOD) and p-values (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model).

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 and summarized over time for all patients (i.e., the FAS).

The analysis will be repeated for the PD-L1 positive and negative groups adjusting for tumor location/HPV status, and smoking status.

5.2.4 Best Objective Response (BoR)

BoR, based on site investigator data (RECIST 1.1), will be summarized by n (%) for each category (CR, PR, SD, PD, and NE), per treatment arm. No formal statistical analyses are planned for BoR.

5.2.5 Disease Control Rate (DCR)

The DCR as per site investigator data (RECIST 1.1) will be analyzed using logistic regressions as described in section 5.2.3 for ORR.

5.2.6 Duration of Response (DoR)

Descriptive data will be provided for the DoR in responding patients by treatment arm, including the Kaplan-Meier curves (without any formal comparison of treatment arms or p-value attached). Swimmer plots will be produced. This depicts each patient's nature and duration of response as a separate bar (horizontally) over time.

5.2.7 Proportion of Patients Alive and Progression Free at 6 and 12 Months

The APF6 and APF12 will be summarized (using the Kaplan-Meier curve) and presented by treatment arm.

5.2.8 Time to Response (TTR)

The TTR, based upon the site investigator data, will be summarized (i.e., number of patients [%] based upon the number of responders) by the scheduled assessment timepoint that the response was first observed. Additionally, descriptive summary statistics (i.e., minimum, maximum, median, Q1 and Q3) will also be presented.

5.2.9 Time from Randomization to Second Progression (PFS2)

Second progression (PFS2) will be analyzed using identical methods as outlined for the analysis of PFS and adjusting for the same set of covariates. Medians and Kaplan–Meier plots will be presented to support the analysis.

5.2.10 Time from Randomization to First and Second Subsequent Therapy or Death

For supportive purposes, the time to the start of first (TFST) and the second (TSST) subsequent therapy or death will be analyzed using the same methodology and model as that used for the analysis of PFS. The HR for the treatment effect together with its 95% CI will be presented. In addition, a Kaplan-Meier plot of the time to the start of first and the second subsequent therapy will be presented by treatment arm, and the time between progression and start of first and the second subsequent therapy will be assessed. This interval will be summarized per treatment arm, but no formal comparisons will be made. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

In patients who received subsequent anticancer therapy, a summary table of first and the second subsequent anticancer therapies by treatment arm will be provided, as well as response to first and the second subsequent anticancer therapy by treatment arm (if available).

The number of patients prematurely censored will also be summarized.

5.2.11 Change in Tumor Size

The absolute values and percentage change in target lesion tumor size from baseline will be summarized using descriptive statistics and presented at each timepoint per 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.

Tumor size will also be presented graphically using waterfall plots, 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. Reference lines at the +20% and –30% change in tumor size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively. Additionally, 'spider' plots will be produced. This depicts each patient's percentage change in tumor size as a line over time. 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 site data based upon RECIST 1.1 assessments.

5.2.12 Patient Reported Outcome (PRO)

The PRO endpoints that have been identified as secondary are EORTC QLQ-C30 time to HRQoL deterioration for global health status, time to symptom deterioration for fatigue, time to symptom deterioration for functional deterioration for physical domain and QLQ-H&N35 time to symptom deterioration for these 2 symptoms; pain and swallowing. These are not part of the main multiple testing procedures and as supportive endpoints will need a Bonferroni adjustment to the significance level to aid interpretation. Therefore, these 5 endpoints will be tested at a 1.0% significance level and 99% CIs will be produced. Also, 95% CIs will be produced.

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

5.2.12.1 EORTC QLQ-C30

Time to symptom deterioration will be analyzed for each of the 3 symptom scales (fatigue, pain, and nausea/vomiting). Time to HRQoL/function deterioration will be analyzed for the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/QoL.

This will be achieved by comparing treatment arms using a stratified log-rank test and stratified Cox proportional hazards model (to get The HR and its CI) as described for the analysis of the co-primary variables of OS.

Time to deterioration will be presented using a Kaplan-Meier plot for each of the 3 symptom scales (fatigue, pain, and nausea/vomiting), 5 functional scales (physical, role, emotional, cognitive, and social), and global health status/QoL.

A summary of the symptom improvement rate for each of the 3 symptom scales (fatigue, pain, and nausea/vomiting) items will be produced. Similarly, a summary of HRQoL/function improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/QoL will be produced. Symptom improvement rate and HRQoL/function improvement rate will be analyzed by comparing between treatment arms

using a logistic regression model as described for the analysis of ORR. If there are too few responses for a meaningful analysis formal analysis will not be performed and only summaries presented (it is not considered appropriate to perform formal tests if there are fewer than 10 responses in a subgroup/scale).

Summaries of absolute and change from baseline values for each of the 3 symptom scales/items (fatigue, pain, and nausea/vomiting), 5 individual symptom items (dyspnoea, insomnia, appetite loss, constipation, and diarrhoea), 5 functional scales (physical, role, emotional, cognitive, and social), and the global health status\QoL score will be reported by visit for each treatment arm. Graphical presentations may also be produced as appropriate.

5.2.12.2 EORTC QLQ-H&N35

Time to symptom deterioration for each of the 4 symptom scales/ items in the QLQ-H&N35 (pain, swallowing, senses and speech) will be compared between treatment arms using a stratified log-rank test and stratified Cox proportional hazards model (to get The HR and its CI) as described for the analysis of the co-primary variables of OS.

For each of these 4 symptom scales/ items in the QLQ-H&N35 above, time to deterioration in symptoms will be presented using a Kaplan-Meier plot.

A summary of the symptom improvement rate for each of the 4 symptom scales/items (pain, swallowing, senses and speech) will be produced. The symptom improvement rate will be compared between treatment groups using a logistic regression model as described for ORR. If there are too few responses for a meaningful analysis formal analysis will not be performed and only summaries presented (it is not considered appropriate to perform formal tests if there are fewer than 10 responses in a subgroup/scale).

Summaries of absolute and change from baseline values for each of the 7 symptom scale/item (pain, swallowing, senses, speech, social eating, social contact and sexuality) and 11 single-item measures (teeth, problems with mouth opening, dry mouth, sticky saliva, coughing, feeling ill, use of analgesics, use of nutritional supplements, use of a feeding tube, weight gain, and weight loss) will be reported by visit for each treatment group. Graphical presentations may also be produced as appropriate.

[redacted]

[redacted]

5.2.12.7 PRO Compliance

Overall compliance and compliance over time will be summarized for each PRO scale and subscales of interest (EORTC QLQ-C30, QLQ-H&N35, [redacted].

5.2.13 Safety

Safety data will be summarized. No formal statistical analyses will be performed on the safety data. All safety and tolerability data will be using the safety population.

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

5.2.13.1 Adverse Events

All AEs, both in terms of current Medical Dictionary for Regulatory Activities (MedDRA) preferred term and Common Toxicity Criteria for Adverse Events (CTCAE) grade, will be listed and summarized descriptively by count (n) and percentage (%). The current MedDRA dictionary will be used for coding. Any AE occurring before treatment with IP will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'. However, any AE occurring before the

administration of the first infusion 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 IP or until the initiation of the first subsequent therapy (excluding palliative radiation as a subsequent therapy) following discontinuation of IP (whichever occurs first) will be used for reporting of all of the AE summary tables. This will more accurately depict AEs attributable to IP only as opposed to presenting all AEs reported up to 90 days following discontinuation of IP. This is due to the fact that a number of AEs up to 90 days following discontinuation are likely to be attributable to subsequent therapy. However, to assess the long term toxicity profile, a small selection of the AE summaries may be repeated containing AEs observed up until 90 days following discontinuation of IP (i.e. without taking subsequent therapy into account). A small selection of AE summaries may also be produced containing AEs observed from initiation of the first subsequent therapy following discontinuation of IP until 90 days following discontinuation of IP treatment (i.e. summarising those AEs experienced by patients taking subsequent therapy during the 90 days AE collection follow-up window post discontinuation of IP). 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 will be listed only apart from a separate summary that presents any events that occur prior to dosing or starting more than 90 days after discontinuing IP.

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

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

- All AEs
- All AEs causally related to study medication
- AEs with CTCAE grade 3 or 4
- AEs with CTCAE grade 3 or 4, causally related to study medication
- AEs with outcome of death
- AEs with outcome of death causally related to study medication
- All SAEs
- All SAEs causally related to study medication

- AEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication, causally related to study medication
- Immune mediated AEs based on pre-defined criteria presented in the immune mediated AE charter and/or per eCRF (representing physician's evaluation). Footnotes will be added in the summary tables to specify how the immune mediated AEs were identified (i.e. by immune mediated AE charter definition or by eCRF/physician's evaluation).
- Infusion reaction AEs

An overall summary of the number and percentage of patients in each category will be presented, as will an overall summary of the number of episodes in each category. In addition, a truncated AE table of most common AEs, showing all events that occur in at least 5% of patients overall will be summarized by preferred term, by decreasing frequency in the MEDI4736 + tremelimumab arm. This cut-off may be modified after review of the data. When applying a cut-off (e.g., 5%), 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%).

Each AE event rate (per 1000 patient years) will also be summarized by preferred term within each system organ class. For each preferred term, the event rate (defined as the number of patients with that AE divided by the total duration of follow-up across all patients in each group multiplied by 1000) will be presented.

AEs will be assigned CTCAE grades (National Cancer Institute (NCI) CTCAE version 4.03) and summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, system organ class, preferred term.

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

Summaries of the number and percentage of patients with AEs leading to dose interruptions of IP will be presented by preferred term.

In addition, AEs with outcome of death, SAEs, AEs leading to discontinuation of treatment, and AEs causally related to IP will be listed.

A summary of deaths will be provided with number and percentage of patients, categorized as:

- Related to disease under investigation,
- AE outcome = death,
- Both related to disease under investigation and with AE outcome=death,

- Patients with unknown reason for death and
- Other deaths.

A corresponding listing will also be produced.

Adverse Events of Special Interest

Preferred terms used to identify adverse events of special interest will be listed before database lock (DBL) and documented in the Study Master File. Grouped summary tables of certain MedDRA preferred terms will be produced. For each 'grouped' term, the number (%) of patients experiencing any of the specified terms will be presented by maximum CTCAE grade. Additional summaries will include Time to Onset of first CTCAE grade 3 or higher. Time to onset of first AE for each grouped term and preferred term within it will also be produced. Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

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

- At least one adverse event of special interest presented by outcome
- At least one adverse event of special interest by CTCAE grade
- At least one adverse event of special interest causally related to study medication
- At least one adverse event of special interest leading to discontinuation of IP

A summary of total duration (days) including median duration 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, as well as a summary of time to resolution to grade 1 or less and time to resolution to grade 2 or less.

Additionally, there will be several summaries of AESIs requiring concomitant treatment, and particularly the relationship of AESIs to the use of immunosuppressive agents (i.e., depicting which AESI triggered immunosuppressive use) and, separately, to the use of immunosuppressive agents at high doses (High dose is defined as >= 40mg of prednisone or equivalent per day). Furthermore, median treatment duration of steroids/immunosuppressant will be provided.

Haemorrhages adverse events

Key summary tables will also be produced considering selected hemorrhage adverse events. Summary information (the number and percent of patients by system organ class and preferred term) will be tabulated for:

- All Haemorrhage AEs, including event rate

- Time to onset
- All Haemorrhage AEs causally related to study medication
- All Haemorrhage AEs with maximum CTCAE grade
- Haemorrhage AEs with CTCAE grade 3 or 4
- Haemorrhage AEs with CTCAE grade 3 or 4, causally related to study medication
- All deaths
- Haemorrhage AEs with outcome of death
- AEs with outcome of death causally related to study medication
- All SAEs
- All SAEs causally related to study medication
- AEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication, causally related to study medication
- Impact of concomitant medication with potential risk of bleeding on the occurrence of the most severe bleed within each patient
- Medical history of bleeding in patients who experience bleeding AEs

Additionally, a listing of all Haemorrhage SAEs and Haemorrhage adverse event patients who experience death and will be provided.

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 categorized by each day after dosing. The prevalence will be plotted over time presented. 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. 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, as well as the median time to resolution to grade 1 or less and time to resolution to grade 2 or less.

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 this will be 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 M 2006). These plots will only be produced for AESIs that have ≥10 events.

5.2.13.2 Laboratory Assessments

Data obtained up until the 90 days following discontinuation of IP or until the initiation of the first subsequent therapy following discontinuation of IP (whichever occurs first) will be used for reporting. This will more accurately depict laboratory toxicities attributable to IP only as a number of toxicities up to 90 days following discontinuation of IP are likely to be attributable to subsequent therapy. However, to assess the long term toxicity profile, a small selection of the summaries of laboratory data will be repeated containing data collected up until 90 days following discontinuation of IP (i.e., without taking subsequent therapy into account). A small selection of summaries of laboratory data will also be produced containing data from initiation of the first subsequent therapy following discontinuation of IP until 90 days following discontinuation of IP (i.e., summarising the laboratory data collected on patients taking subsequent therapy during the 90 day follow-up window post discontinuation of IP). 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 will be listed only.

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

All laboratory data will be listed. Flags will be applied to values falling outside - reference ranges (which will be explicitly noted on these listings where applicable), and to values for which CTC grading applies.

Scatter plots (shift plots) of baseline to maximum value on treatment as well as plots for baseline to minimum value on treatment (on treatment is defined as data collected between the start of treatment and the relevant follow-up period following the last dose of IP) will be produced for: clinical chemistry (ALT, AST, ALP, total bilirubin, corrected calcium, magnesium, sodium, potassium, creatinine, albumin, corrected calcium, magnesium, sodium, potassium, gamma glutamyl transferase and glucose) and haematology (haemoglobin, lymphocyte (count, absolute), neutrophils (count, absolute), platelet count and leukocytes).

If clinically indicated; box-plots of absolute values by week, and box-plots of change from baseline by week, may be presented for haemoglobin; neutrophil count, absolute; lymphocyte count, absolute; platelet count; AST; ALT; ALP; Total bilirubin; albumin; total protein; corrected calcium; phosphate; sodium; potassium; creatinine and urea nitrogen.

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

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

- Clinical chemistry: ALT, AST, Alkaline Phosphatase (ALP), Total bilirubin, Albumin, Magnesium – hypo and – hyper, Phosphate –hypo and -hyper, Sodium – hypo and – hyper, Potassium – hypo and – hyper, Corrected calcium – hypo and – hyper, Glucose – hypo and – hyper, GGT, Bicarbonate, Creatinine

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

Note that for leukocyte differentials (lymphocytes, monocytes, neutrophils), SI units are considered absolute counts, not % values, so should all be converted to 10**9/L. This conversion to absolute counts is achieved by multiplying the % value by the leukocyte count at the current timepoint.

Hy's law

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

- Elevated ALT, AST, and Total bilirubin during the study

    - ALT ≥ 3x –<= 5x, > 5x – <=8x, > 8x - < 10x, > 10x - <= 20x, and > 20x Upper Limit of Normal (ULN) during the study

    - AST ≥ 3x – <=5x, > 5x – <=8x, > 8x - <= 10x, > 10x - <= 20x. and > 20x ULN during the study

    - Total bilirubin ≥2x - <= 3x, >3x - <=5x, and >5x ULN during the study

- Narratives will be provided for potential Hy's law cases defined as an increase in serum AST or ALT ≥ 3x ULN preceding, or coincident with, Total Bilirubin ≥ 2xULN, irrespective of serum Alkaline Phosphatase (ALP), at any point during the study following the start of randomised therapy.

Liver biochemistry test results over time for patients with elevated ALT or AST, and elevated total bilirubin (at any time) will be plotted. Individual patient data where ALT or AST plus Total bilirubin are elevated at any time will be listed also.

Plots of ALT and AST vs. Total bilirubin will also be produced with reference lines at 3×ULN for ALT, AST, and 2×ULN for Total bilirubin. In each plot, Total bilirubin will be in the vertical axis.

Abnormal Thyroid function

Elevated TSH will be summarized per treatment group in terms of number (%) of patients with elevated TSH (higher than the upper normal range), low TSH (lower than lower normal range), elevated TSH post-dose and within normal range at baseline, low TSH post-dose and within normal range at baseline.

5.2.13.3 ECGs

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

Overall evaluation of ECG is collected at screening and as clinically indicated in terms of normal or abnormal, and the relevance of the abnormality is termed as "clinically significant" or "not clinically significant".

5.2.13.4 Vital Signs

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

5.2.14 WHO/ECOG Performance Status

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

[redacted]

[redacted]

5.2.17 Demographic, Initial Diagnostics and Baseline Characteristics Data

The following will be listed and summarized for all patients in the FAS (unless otherwise specified):

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations specified in this SAP
- Inclusion in analysis populations
- Demographics (age, age group [<65, ≥65 – <75, and ≥75 years], sex, race and ethnicity)
- Patient characteristics at baseline weight, weight group [<70, ≥70 – <90, and ≥ 90 kg], Body Mass Index (BMI) [<18.5, ≥18.5 – 25.0, ≥25.0 – <30.0, ≥30.0 kg/m^2].
- Patient recruitment by country and center
- Previous disease-related treatment modalities
- Number of regimens of previous chemotherapy for SCCHN at baseline
- Previous head and neck cancer therapy
- Disease characteristics at baseline (WHO/ECOG performance status, best response to previous therapy)

- Disease characteristics at initial diagnosis (primary tumor location, histology type, tumor grade, time from diagnosis to first dose, overall disease classification, AJCC staging)
- Time to recurrence from last dose of platinum in a platinum-containing multimodality therapy (<6 months, ≥6 months)
- Extent of disease at baseline
- Disease related medical history
- Time from most recent disease progression to randomization
- Post-discontinuation cancer therapy
- Smoking status (>10 versus ≤10 pack-years)
- PD-L1 status (positive, negative) for all patients treated and also from those who have screened.
- Tumor location/HPV Status (oropharyngeal cancer with HPV positive status, oropharyngeal cancer with HPV negative status, and non-oropharyngeal cancer regardless of HPV status)
- Stratification factors by IVRS and CRF

The following will also be listed for all patients in the FAS (unless otherwise specified) per ICH guidelines:

- Important protocol deviations specified in this SAP
- Subject excluded from analysis populations
- Demographics (age, age group [<65, ≥65 – <75, and ≥75 years], sex, race and ethnicity)

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

5.2.18 Treatment Exposure

The following summaries related to MEDI4736, tremelimumab, and SoC will be produced for the safety analysis set:

- Total exposure.
- Actual exposure (Actual exposure will not be provided for SoC patients).
- Number of dose interruptions.
- RDI (entire intended treatment period).

For patients on study treatment at the time of the ORR and OS analysis, the DCO date will be used to calculate exposure. Summaries of exposure will also be presented for the subgroup of discontinued patients.

All treatment information data (study drug administration, MEDI4736, tremelimumab and SoC) will be listed for the safety analysis set.

5.2.19 Subsequent Therapy

Subsequent therapies received after discontinuation of study treatment will have summaries produced, together with number of regimens received. Moreover, a descriptive summary will be produced for time to subsequent therapy from discontinuation of study drug treatment.

6. INTERIM ANALYSIS

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

In addition, 1 interim analysis for OS will be performed for superiority when a total of approximately 300 death events (80% of required events) have been accumulated across the MEDI4736 + tremelimumab combination therapy and SoC arms in all patients, regardless of PD-L1 status; it is expected that approximately 300 death events have accumulated across the MEDI4736 monotherapy and SoC arms at this time. The Lan-DeMets spending function that approximates an O'Brien Fleming approach will be used to account for the multiplicity introduced by including the interim analysis for superiority (Lan and DeMets 1983).

If exactly 80% of target death information is available at the time of the interim analysis, that is, 300/375 death events have been accumulated across MEDI4736 + tremelimumab combination therapy and SoC arms; and approximately 300/375 death events have been accumulated across MEDI4736 monotherapy and SoC arms, in all patients, regardless of PD-L1 status, the 2-sided significance level to be applied for the interim and final analyses would be 1.0% and 2.2%, respectively for each co-primary objective. If the interim analyses indicate superiority in both of the co-primary objectives, then subsequent analysis of the secondary objective will be performed in accordance with the multiple testing strategy. A separate Lan-DeMets (O'Brien Fleming) spending function will be used to determine the alpha levels at the interim and final analyses for the secondary objective, as applicable. If the interim analysis results do not meet the criterion for stopping for superiority in either of the co-primary objective, then follow-up will continue until approximately 375 death events have been accumulated across MEDI4736 +tremelimumab combination therapy and SoC arms regardless of PD-L1 status; and approximately 375 death events have accumulated across MEDI4736 monotherapy and SoC arms in all patients, regardless of PD-L1 status. OS will be analyzed in all patients, regardless of PD-L1 status and in PD-L1-negative populations at the final analysis.

7. CHANGES OF ANALYSIS FROM PROTOCOL

Table 13 Changes of Analysis from Protocol

Section of SAP Affected (If applicable)	Change	Rationale
Section 1.1.2	TTR, PFS2, TFST and TSST were added in the secondary endpoints	TTR is supportive of ORR. Time to first and second subsequent therapy (TFST, TSST) are supportive of anti-tumor activity. PFS2 analysis was mentioned in the protocol, but PFS2 was not identified in the secondary endpoints. It was added for consistency.
Section 3	The analysis sets table 12 in the CSP indicates that DoR and DCR will be analyzed in both the PDL1 negative and positive analysis sets. These subset analysis won't be performed.	The inclusion of DoR and DCR subset analyses were not intended.
Section 4.1	The following '(based on observed information fraction in this subgroup)' was added: A separate spending function (based on observed information fraction in this subgroup) will be used to adjust the alpha levels at the interim and final analyses in the PD-L1-negative population.	Clarification of text, no change to methodology.
Section 5	The following sentence was updated from either to both: If the interim analyses indicate superiority in both of the co-primary objectives, then subsequent analysis of the secondary objective will be performed in accordance with the multiple testing strategy.	This was a mistake in the protocol text.
Section 5	The following sentence was removed: Similarly, if the criterion for stopping for superiority is met at the interim in either of the co-primary objective, but is not met in the secondary objective (PD-L1-negative population) at that time, then follow-up will continue until the final target number of deaths in the PD-L1-negative population has been observed. The final analysis in the PD-L1- negative population will then be conducted.	Section was not updated when protocol was amended. Now consistent with protocol.
[redacted]	[redacted]	[redacted]

Table 13 Changes of Analysis from Protocol

Section of SAP Affected (If applicable)	Change	Rationale
Section 5	Landmark analyses (OS12, OS18, OS24, PFS6, PFS12) will not be formally analyzed (using the methods of Klein et al).	Formal analysis was not considered necessary to support interpretation beyond the Kaplan Meier estimates of the median event times.

8. REFERENCES

Bjordal et al 2000

Bjordal K, de Graeff A, Fayers PM, Hammerlid E, van Pottelsberghe C, Curran D, et al, for the EORTC Quality of Life Group. A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients. Eur J Cancer 2000;36(14):1796-807.

Breslow 1974

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

Burman et al 2009

Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Statistics in Medicine 2009;28:739-61.

Cox 1972

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

Fayers et al 2001

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.

Gail and Simon 1985

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

[redacted]

Lan and DeMets 1983

Lan KKG and DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70(3):659-63.

Osoba et al 1998

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

Pardoll 2012

Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nature reviews: cancer 2012;12:252.

Pintilie M. 2006

Competing risks: A practical perspective. Wiley.

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.

Shin and Khuri 2013

Shin DM, Khuri FR. Advances in the management of recurrent or metastatic squamous cell carcinoma of the head and neck. Head Neck 2013;35(3):443-53.

Vermorken et al 2008

Vermorken JB, Mesia R, Rivera F, Remenar E, Kawecki A, Rottey S, et al. Platinum-based chemotherapy plus cetuximab in head and neck cancer. N Engl J Med 2008;359(11):1116-27.

9. APPENDIX (NOT APPLICABLE)