AstraZeneca

**Statistical Analysis Plan**

| | |
|---|---|
| Drug Substance | MEDI4736 and Tremelimumab |
| Study Code | D419LC00001 |
| Edition Number | 07 |
| Date | 16 DEC 2020 |

# A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 Alone or in Combination with Tremelimumab versus Standard of Care (EXTREME) in the Treatment of First-line Recurrent or Metastatic Squamous Cell Head and neck Cancer patients

# A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 Alone or in Combination with Tremelimumab versus Standard of Care (EXTREME) in the Treatment of First-line Recurrent or Metastatic Squamous Cell Head and neck Cancer patients

**Study Statistician**

PPD

Date

# A Phase III Randomized, Open-Label, Multi-Center, Global Study of MEDI4736 Alone or in Combination with Tremelimumab versus Standard of Care (EXTREME) in the Treatment of First-line Recurrent or Metastatic Squamous Cell Head and neck Cancer patients

**Global Product Statistician**

PPD

Date

# TABLE OF CONTENTS PAGE

## LIST OF TABLES

## LIST OF FIGURES

# LIST OF ABBREVIATIONS

| Abbreviation or special term | Explanation |
|---|---|
| ADA | Antidrug antibody |
| AE | Adverse event |
| AEPI | Adverse events of possible interest |
| AESI | Adverse events of special interest |
| ALT | Alanine aminotransferase |
| ALQ | Above Limit of Quantification of PK assay |
| 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 enrollment. |
| BICR | Blinded Independent Central Review |
| BLQ | Below Limit of Quantification of PK assay |
| BMI | Body Mass Index |
| BoR | Best objective response |
| CI | Confidence interval |
| COVID-19 | Coronavirus Disease 2019 |
| 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 |
| ctDNA | Circulating tumor DNA |
| CTLA-4 | Cytotoxic T–lymphocyte-associated antigen 4 |
| CTM | Clinical team manager |
| CTMS | Clinical trial management system |
| DCO | Data cut-off |
| DCR | Disease control rate |
| DoR | Duration of response |

| Abbreviation or special term | Explanation |
|---|---|
| ECG | Electrocardiogram |
| ECOG | Eastern Cooperative Oncology Group |
| EDoR | Expected Duration of Response |
| EM | Early mortality |
| EORTC | European Organisation for Research and Treatment of Cancer |
| CCI | CCI |
| FAS | Full analysis set |
| HLT | High level term |
| HLGT | High level group term |
| HR | Hazard ratio |
| CCI | CCI |
| HRQoL | Health-related quality of life |
| IC | Tumor-associated immune cell |
| IDMC | Independent Data Monitoring Committee |
| ICR | Independent Central Reviewer |
| imAE | Immune-mediated Adverse Events |
| IMT | Immunomodulatory therapy |
| IP | Investigational product |
| irRECIST | Immune-related response criteria |
| ITT | Intent-to-treat |
| IV | Intravenous |
| KM | Kaplan-Meier |
| LLOQ | Lower Limit of Quantification of PK assay |
| MD | Medical doctor |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MEDI4736 | Immune-mediated therapy |
| mg | Milli-gram |
| MMA | Medical monitoring associate |
| MRI | Magnetic resonance imaging |
| MMRM | Mixed effect model repeat measurement |
| NA | Not applicable |
| NCI | National Cancer Institute |

| Abbreviation or special term | Explanation |
|---|---|
| NE | Not evaluable |
| NED | No evidence of disease |
| CCI | CCI |
| NRR | No Reportable Result |
| NTL | Non-target lesions |
| ORR | Objective response rate |
| OS | Overall survival |
| 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) |
| PFS | Progression free survival |
| CCI | CCI |
| PK | Pharmacokinetic(s) |
| 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 |
| RPD | Relative Percent Difference |
| RR | Response rate |
| SAE | Serious adverse event |
| SAP | Statistical Analysis Plan |

| Abbreviation or special term | Explanation |
|---|---|
| SCCHN | Squamous cell carcinoma of the head and neck |
| SD | Stable disease |
| SDV | Source data verify |
| SBP | Systolic blood pressure |
| TC | Tumor cell |
| TL | Target lesions |
| TMB | Tumor mutation burden |
| TFST | Time to first subsequent therapy |
| TTR | Time to Response |
| TSST | Time to second subsequent therapy |
| ULN | Upper limit of normal |
| WHO | World Health Organization |

# AMENDMENT HISTORY

| Date | Brief description of change |
|---|---|
| **May 2017** | In line with the Clinical Study Protocol (CSP) Amendment 6<br>• A primary objective has been added for OS in a PD-L1 TC/IC-selected subgroup.<br>• Furthermore, secondary objectives assessing the MEDI4736+Tremelimumab combination therapy versus standard of care in the PD-L1 TC/IC-selected subgroup have been included in addition to the all-comers.<br>• The timing of the PFS analysis has been moved to coincide with the same time as the 80% of death events interim analysis of OS, thereby removing the first interim analysis of OS from the protocol.<br>• All relevant sections throughout the SAP have been updated to reflect these changes.<br>• Sections 3.1, 3.1.3, 3.2.1 and all relevant sections have been updated to indicate that the sensitivity analysis of PFS based on RECIST 1.1 modified for confirmation of progression will be performed using site Investigator assessments, rather than BICR assessments.<br>• Section 3.5.2 – Calculation or derivation of pharmacokinetic variables**.** This section has been updated to reflect that PK parameters, such as peak and trough concentration of MEDI4736 and tremelimumab, will be derived from raw data measurements and that no formal non-compartmental PK analysis methods will be used as only sparse PK is being collected in this study.<br><br>In line with project wide developments<br>• Clarification that the progression date for BICR assessment will be based on progression scan dates (PROGSCDT) instead of the earliest scan dates<br>• Clarification that the death summary table will now be broken into 3 separate tables<br>• Stratification factors: Only the analysis for primary objectives will use IVRS based stratification factors; all other analyses, including subgroup analysis, baseline characteristics summaries, etc, as well as all other factors, will be based on the values recorded on the eCRF<br><br>Other Changes<br>• Addition of BoR (best objective response)<br>• Addition of BMI and BMI group at the patient characteristics summary<br>• Added a text explaining how to impute missing dates<br>• Addition of Mixed effect model repeated measurement (MMRM) analysis for EORTC QLQ-C30 and EORTC QLQ-H&N 35<br>• Minor clarifications |

| | |
|---|---|
| **January 2018** | **Brief description of change** |

In line with the Clinical Study Protocol (CSP) Amendment 7

- Change of PFS from primary objective to secondary objective
- Removal of interim analysis for overall survival (OS)
- Addition of secondary objectives, APF6, OS12 and OS18
- All analyses using tumor assessment data will use site investigator data as primary analysis instead BICR data, BICR will be used as either sensitivity CCI when applicable

In line with project wide developments

- Hazard ratio and confidence interval derived at landmarks as per Klein's method were removed.
- Clarifying the baseline definition for PRO variables
- Summarizing AEs up to 90 days after last dose regardless of re-treatment

Other changes

- Clarifying the approach to be used for deriving HR and its associated CI from Cox model
- Providing details and clarifications for the derivation of exposure for the Combination and SOC arms
- Additional minor clarifications such as attrition bias, two missed assessment for PFS, minimum duration needed for stable Disease, definition of potential Hy's law and many others

| | |
|---|---|
| **July 2018** | **Brief description of change** |

- The prior version of this SAP had only included the initial two levels of the MTP in the MTP figure. Hence, the full MTP figure has now been included for accuracy.
- The important protocol deviation list has been updated as PFS is no longer a primary endpoint.
- Text has now been added to detail how partial death dates will be handled.
- Minor text corrections and clarification have also been made.
- The AESIs have been updated to be aligned with the CSP.
- Text has now been added to detail how missing CRF stratification factors for sensitivity analysis of the primary objectives will be handled.
- Text has been added to clarify the rules for handling BLQ data when a certain percentage of the serum concentrations are not quantifiable; and that ALQ results of the PK assay will be excluded from summary statistics calculations.

| | |
|---|---|
| **March 2019** | **Brief description of change** |

In line with the Clinical Study Protocol (CSP) Amendment 8, CSP edition 9 and Amendment 9, CSP edition 10 includes changes as follows:

- The primary objective has changed from statistical testing for overall survival of MED4736 +tremelimumab combination therapy compared to SoC in the all-comer population to MED4736 monotherapy compared to SoC in the all-comer population
- The MTP has been updated to reflect the change primary objective and prioritization of a secondary objective
- The analysis of ctDNA TMB has been updated from CCI secondary objective
- All relevant sections throughout the SAP have been updated to reflect the two changes above
- DoR, BoR, TFST, TSST, APF6 and APF12 have been included as secondary objectives to assess the efficacy of MEDI4736 monotherapy compared to SoC
- The definition for PD-L1 high and PD-L1 low were clarified

Other changes

- Clarifying that PRO analyses will include retreatment data
- CCI
- Minor administrative corrections
- Addition of ADA analysis set
- Updating the derivation of the two missed visit rule for RECIST 1.1 assessment
- Addition of infection AEs to the safety outputs

| | |
|---|---|
| **June 2020** | **Brief description of change** |

In line with the CSP edition 12, the following changes have been made:

- Change of primary objective from MEDI4736 monotherapy versus SoC in all randomized patients (all-comers) to MEDI4736 monotherapy versus SoC in PD-L1 TC/IC high subgroup in terms of OS. All relevant sections throughout the SAP have been updated to reflect these changes.
- The MTP has been updated to reflect the change in primary objective and prioritization of secondary objectives
- The low risk of early mortality (EM) analysis set has been added and the secondary objectives have been updated to include this additional subgroup population

Other changes:

- All COVID-19 related non-important PDs and issues will now be summarized and listed
- Addition of AEPIs and imAEs to the safety outputs
- ECG and vital signs data will now be reported up to 90 days after date of last dose of study treatment or until initiation of subsequent therapy
- A further sensitivity analysis for OS i.e. stratified max-combo test, has been included
- CCI [redacted]
- In addition to the changes listed above, minor administrative corrections and clarifications have been made throughout the SAP and are not detailed here further

**December 2020** **Brief description of change**

- Further details have been added to the max-combo sensitivity analysis of OS, CCI [redacted]
- The MMRM analysis has been updated to include a baseline by visit interaction
- An approach for handling duplicate PK data has been added

# 1. STUDY DETAILS

The target population for this study is male and female patients aged 18 and over with histologically or cytologically confirmed PD-L1-positive or -negative, recurrent or metastatic SCCHN (oral cavity, oropharynx, hypopharynx, or larynx) who are not amenable to local curative therapy with surgery or radiation and who have not received prior systemic therapy for recurrent/metastatic disease, unless it was given as part of multimodality treatment for locally advanced or locally recurrent disease.

Multiple lines of evidence suggest that SCCHN tumors create a highly immunosuppressive environment and that the PD-1/PD-L1 axis and inhibition of the activation of T cells play an important role and may be amenable to therapeutic intervention with immune-modulating agents (see CSP for detailed description).

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 other mAbs targeting the PD-L1/PD-1 pathway, which provide early evidence of clinical activity and a manageable safety profile in multiple tumor types (Brahmer et al 2012, Topalian et al 2012). Currently available data from the first-time-in-human, single-agent study (Study CD-ON-MEDI4736-1108; referred to hereafter as Study 1108) in patients with advanced solid tumors using MEDI4736 monotherapy indicates encouraging response rates (RRs) and duration of response (DoR) with a manageable safety profile in patients with a variety of solid malignancies, including patients with SCCHN. Preliminary data in the SCCHN expansion cohort demonstrated a confirmed ORR of 11% (7 out of 62 partial response [PR]) and a disease control rate (DCR) at 24 weeks of 15% (9 out of 62 patients) based on Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) (Segal et al 2015).

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 for this combination has been observed in Study D4190C00006.

## 1.1 Study objectives

### 1.1.1 Primary Objective

**Table 1 Primary Objectives**

| **Primary Objective:** | **Outcome Measure:** |
|---|---|
| To assess the efficacy of MEDI4736 monotherapy compared to SoC (EXTREME) in terms of OS | OS in the PD-L1 TC/IC high subgroup |

OS Overall survival;

### 1.1.2 Secondary Objectives

**Table 2 Secondary Objectives**

| Secondary Objective: | Outcome Measures: |
|---|---|
| To further assess the efficacy of MEDI4736 monotherapy compared to SoC (EXTREME) in terms of OS, PFS, ORR, DoR, BoR, TTR, TFST, TSST, APF6, APF12, PFS2, OS12, OS18 and OS24 | OS in the low risk of early mortality (EM) subgroup, ctDNA TMB high subgroup (≥16 mut/Mb) and all-comers<br><br>OS12, OS18, OS24 in the PD-L1 TC/IC high subgroup, low risk of early mortality (EM) subgroup, ctDNA TMB high subgroup (≥16 mut/Mb) and all-comers<br><br>PFS, ORR, APF6 and APF12 using site investigator assessments according to RECIST 1.1 in the PD-L1 TC/IC high subgroup, low risk of early mortality (EM) subgroup, ctDNA TMB high subgroup (≥16 mut/Mb) and all-comers<br><br>DoR, BoR and TTR using site investigator assessments according to RECIST 1.1 in the PD-L1 TC/IC high subgroup and all-comers<br><br>PFS2 using local standard clinical practice in the PD-L1 TC/IC high subgroup and all-comers<br><br>TFST and TSST in the PD-L1 TC/IC high subgroup and all-comers |

| | |
|---|---|
| To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC (EXTREME) in terms of OS, PFS, ORR, DoR, BoR, TTR, APF6, APF12, PFS2, TFST, TSST, OS12, OS18 and OS24 | OS, OS12, OS18, OS24 in the PD-L1 TC/IC high subgroup, low risk of early mortality (EM) subgroup, ctDNA TMB high subgroup (≥16 mut/Mb) and all-comers<br><br>PFS, ORR, APF6 and APF12 using site investigator assessments according to RECIST 1.1 in the PD-L1 TC/IC high subgroup, low risk of early mortality (EM) subgroup, ctDNA TMB high subgroup (≥16 mut/Mb) and all-comers<br><br>DoR, BoR and TTR using site investigator assessments according to RECIST 1.1 in the PD-L1 TC/IC high subgroup and all-comers<br><br>PFS2 using local standard clinical practice in the PD-L1 TC/IC subgroup and all-comers<br><br>TFST and TSST in the PD-L1 TC/IC high subgroup and all-comers |
| To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to MEDI4736 monotherapy in terms of PFS, ORR, and OS | PFS and ORR using site investigator assessments according to RECIST 1.1 in the PD-L1 TC/IC high subgroup, low risk of early mortality (EM) subgroup, ctDNA TMB high subgroup (≥16 mut/Mb) and all-comers<br><br>OS in the PD-L1 TC/IC high subgroup, low risk of early mortality (EM) subgroup, ctDNA TMB high subgroup (≥16 mut/Mb) and all-comers |
| To assess disease-related symptoms and health-related quality of life in patients treated in patients treated with MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy compared to SoC (EXTREME) using the EORTC QLQ-C30 v3 and the QLQ-H&N35 module | EORTC QLQ-C30: global health QoL, functioning (physical) and symptoms (fatigue in the PD-L1 TC/IC high subgroup and all-comers<br><br>EORTC QLQ-H&N35: symptoms (pain, swallowing) in the PD-L1 TC/IC high subgroup and all-comers<br><br>Changes in WHO/ECOG performance status in the PD-L1 TC/IC high subgroup and all-comers |
| To assess the PK of MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy | Concentration of MEDI4736 and tremelimumab in blood and PK parameters, such as peak concentration and trough (as data allow; sparse sampling) |
| To investigate the immunogenicity of MEDI4736 and tremelimumab | Presence of ADAs for MEDI4736 and tremelimumab |

ADA Anti-drug antibody; APF6 Proportion of patients alive and progression free at 6 months; APF12 Proportion of patients alive and progression free at 12 months; DoR Duration of response; ECOG Eastern Cooperative Oncology Group; EM Early mortality; EORTC European Organisation for Research and Treatment of Cancer; HRQoL Health-related quality of life; ORR Objective response rate; OS Overall survival; OS12 Proportion of patients alive at 12 months after randomization; OS18 Proportion of patients alive at 18 months after randomization; OS24 Proportion of patients alive at 24 months after randomization; PFS Progression free survival; PFS2 Time from randomization to second PFS; TTR Time to response; TFST Time from randomization to first subsequent therapy or death; TSST Time from randomization to second subsequent therapy or death; IC Tumor associated immune cells; TC tumor cells; BoR Best objective response; QLQ-C30 v3 30-item core quality of life questionnaire, version 3; QLQH& N35 35-item head and neck quality of life questionnaire; QoL Quality of life; RECIST 1.1 Response Evaluation Criteria in Solid Tumors version 1.1; SCCHN Squamous cell carcinoma of the head and neck; SoC Standard of Care; ctDNA Circulating tumor DNA; DNA Deoxyribonucleic acid; TMB Tumor mutation burden. The ctDNA TMB cut-point of 16 mut/MB for MEDI4736 + tremelimumab and MEDI4736 monotherapy was derived from the EAGLE study (Li et al, 2020).

### 1.1.3 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 (EXTREME) in the first-line setting for treatment of SCCHN | AEs, physical examinations, laboratory findings (including clinical chemistry, hematology, and urinalysis), vital signs (including blood pressure and pulse), and ECGs in the PD-L1 TC/IC high subgroup, low risk of EM subgroup and all-comers |

AE Adverse Event; ECG Electrocardiogram; EM Early mortality; SoC Standard of Care; IC Tumor-associated immune cells; TC Tumor cells

### 1.1.4 Exploratory Objectives

**Table 4 Exploratory Objectives**

CCI

## 1.2 Study design

This is a randomized, open-label, multi-center, 3-arm, global Phase III study to determine the efficacy and safety of MEDI4736 monotherapy compared to SoC in the treatment of patients with recurrent or metastatic SCCHN who are not amenable to local curative therapy with surgery or radiation and who have not received prior systemic therapy for recurrent/metastatic disease, unless it was given as part of multimodality treatment for locally advanced or locally recurrent disease. Note that MEDI4736 + tremelimumab combination therapy compared to SoC is a secondary objective.

A schematic diagram of overall study design is shown in Figure 1, a flow chart of the study design treatment periods is presented in

Figure 2, and a flow chart of the optional retreatment period for the MEDI4736 + tremelimumab combination therapy arm is in Figure 3.

Patients will undergo an assessment on their tumor tissue sample to determine PD-L1 status. Patients with tumoral PD-L1 expression above or below a pre-specified cut-off level of ≥25%, as determined by an IHC assay (referred to hereafter as PD-L1–positive or -negative tumors, respectively), will be enrolled in the study. If the patient has already been tested as a part of the screening process for any AstraZeneca study using the VENTANA assay, this test result can be used for the determination of eligibility. The specified expression cut-off level of ≥25% in the tumor cells will be used for the purpose of stratification and therefore included in the stratified log rank tests for OS and PFS.

HPV status will be assessed in patients with oropharynx tumors according to local standards or by the p16 IHC assay.

Data presented by AstraZeneca at the 2015 ASCO meeting (**Error! Reference source not found.**) demonstrate that treatment with MEDI4736 resulted in an ORR of 18% and 8% in patients with PD-L1-positive and -negative SCCHN tumors, respectively. Therefore, it appears that the selection of patients based on PD-L1 expression levels within the tumor microenvironment may improve the probability and/or quality of responses to PD-1 pathway-targeting agents and, therefore, may have merit as a patient enrichment tool. Competitors have presented similar data with this class of agents. Specifically, the phase 3 KEYNOTE-048 study, which demonstrated that pembrolizumab monotherapy improved OS when compared with the EXTREME regimen in patients whose tumors had PD-L1 expression ≥1 and ≥20 by Combined Positive Score (CPS) [hazard ratio (HR) 0.78 (0.64–0.96), P=0.0086 and HR 0.61 (0.45– 0.83), P=0007, respectively] (Burtness et al., 2019). CCI

[redacted]

- PD-L1 TC/IC high is defined as either ≥ 50% of the tumor cells (TC) or ≥ 25% of the immune cells (IC) staining for PD-L1 at any intensity if >1% of the tumor area contains IC, or ≥ 50% of TC or 100% of IC staining for PD-L1 at any intensity if 1% of the tumor area contains IC.
- PD-L1 low is defined as not meeting any of the criteria for PD-L1 high

Patients will be randomized in a 2:1:1 manner to either the MEDI4736 + tremelimumab combination arm, the MEDI4736 monotherapy arm, or the SoC arm respectively. Patients randomized to the MEDI4736 + tremelimumab combination arm will receive tremelimumab 75 mg intravenous (IV) q4w for 4 doses and MEDI4736 1500 mg IV q4w until PD, those in the MEDI4736 monotherapy arm will receive MEDI4736 1500 mg IV q4w until PD, while patients in the SoC arm will receive up to six 3-week cycles consisting of a platinum [cisplatin 100 mg/m$_2$ or carboplatin area under the curve (AUC) of 5 mg/mL/min IV] on day 1 of each cycle, 5FU 1000 mg/m$_2$/day on days 1-4 of each cycle, and cetuximab weekly; cetuximab will be administered at 400 mg/m$_2$ on Day 1 of cycle 1, and then 250 mg/m$_2$ weekly for up to six 3-week cycles and 250 mg/m$_2$ IV weekly for maintenance until disease progression, initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or another discontinuation criterion is met (see Section 3.9 of CSP).

Patients in all arms with confirmed PD by RECIST 1.1 who, in the Investigator's opinion, continue to receive benefit from their assigned treatment and who meet the criteria for treatment in the setting of PD may continue to receive their assigned treatment (see Section 7.2.3 of CSP). Since the phenomenon of pseudoprogression in patients treated with immunotherapy is well documented and the limitations of RECIST 1.1 in identifying those who are achieving clinical benefit with immunotherapies are known, patients enrolled in the IMT arms will be permitted to continue therapy even in the presence of PD as per RECIST 1.1 (Weber et al 2012). However, patients whose disease progresses in target lesions which had previously shown an objective response as defined by RECIST 1.1 will not be permitted to continue immunotherapy. Similarly, patients in the SoC arm may continue therapy in the setting of PD if they remain clinically stable, meet criteria for treatment in the setting of PD, and continue to receive some clinical benefit in the Investigator's opinion.

Patients in all arms with a symptomatic solitary lesion, a brain lesion, or a lesion of clinical importance (e.g., impending fracture, etc) may be treated with radiation after approval from the Sponsor. Non-irradiated lesions need to continue to be measurable by RECIST 1.1. The time interval between the last systemic treatment and radiation should be 10 days, and systemic treatment can be resumed 1week post radiation.

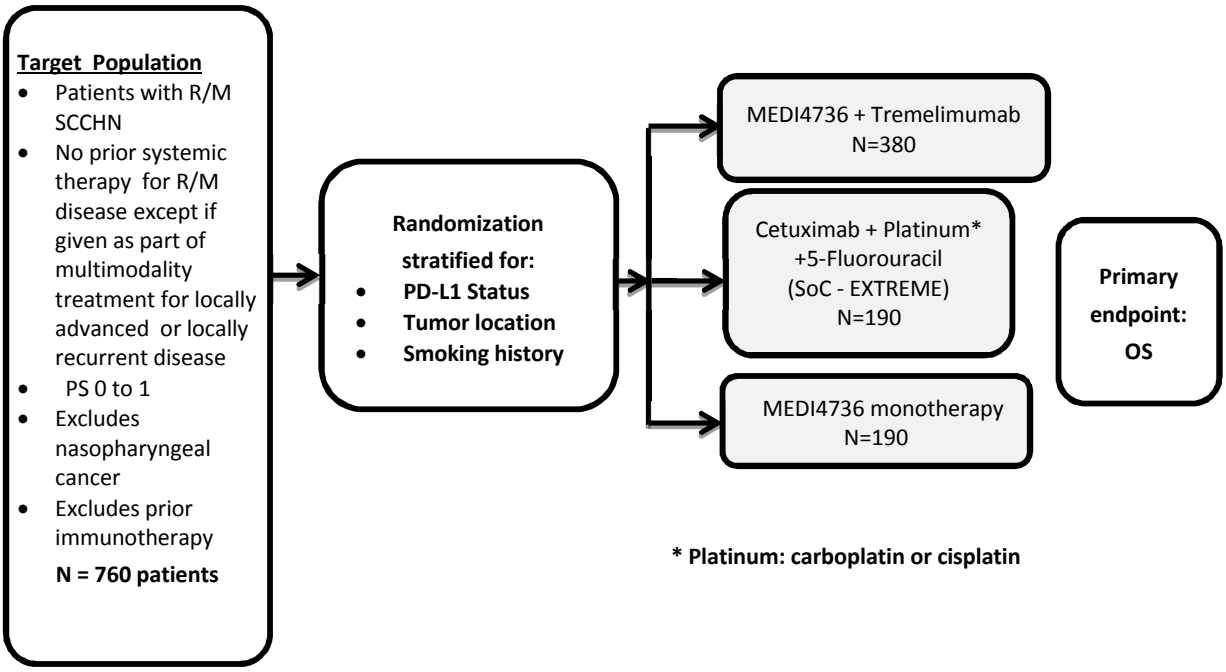
Patients who have discontinued therapy due to disease progression will enter follow-up until death (see Table 4 CSP). Patients who have discontinued treatment due to toxicity or any reason other than disease progression will be followed up until confirmed disease progression and for survival (see Table 4 CSP).

Tumor assessments will be performed on images from computed tomography (CT) or magnetic resonance imaging (MRI) scans, preferably with IV contrast, at the times specified in Table 2, Table 3, and Table 4 of CSP. RECIST 1.1 measurements as given by site investigator assessments will be used to derive the secondary variables PFS, ORR, DoR, BoR, TTR, and proportion of patients alive and progression free at 6 and 12 months (APF6 and APF12).

This study will enroll approximately 1016 patients with PD-L1-positive and -negative disease at sites globally to obtain 760 patients who are likely to be evaluable for the primary endpoint.

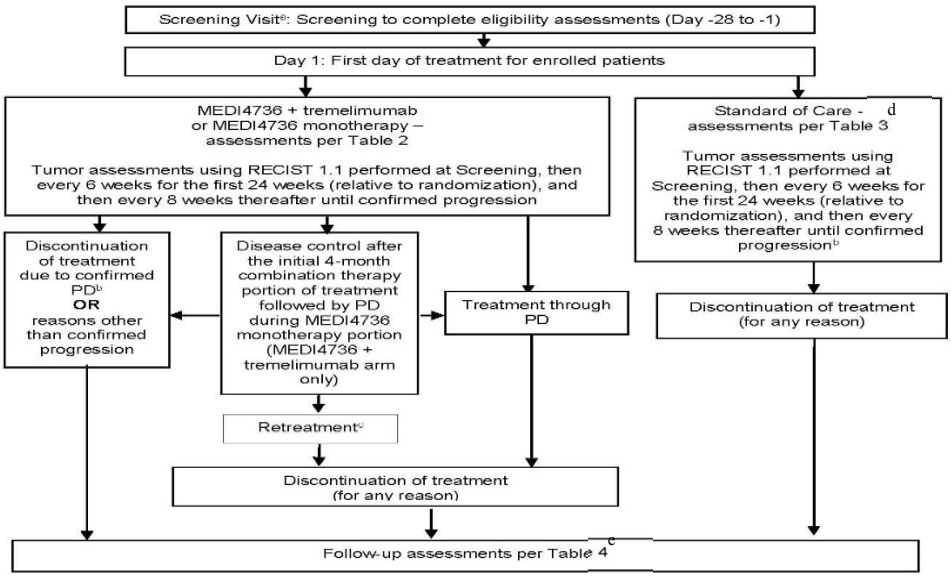
Below is a schematic diagram of the overall study design (Figure 1). A flow chart for all treatment groups (MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy, and SoC) is presented in Figure 2. A flow chart for patients on the MEDI4736 + tremelimumab combination therapy arm who enter the optional re-treatment phase is presented in Figure 3.

## Figure 1 Overall study design




OS Overall survival; PD-L1 Programmed cell death ligand 1; PS Eastern Cooperative Oncology Group performance status; SCCHN Squamous cell cancer of the head and neck; SoC Standard of Care.

## Figure 2 Study flow chart-treatment period

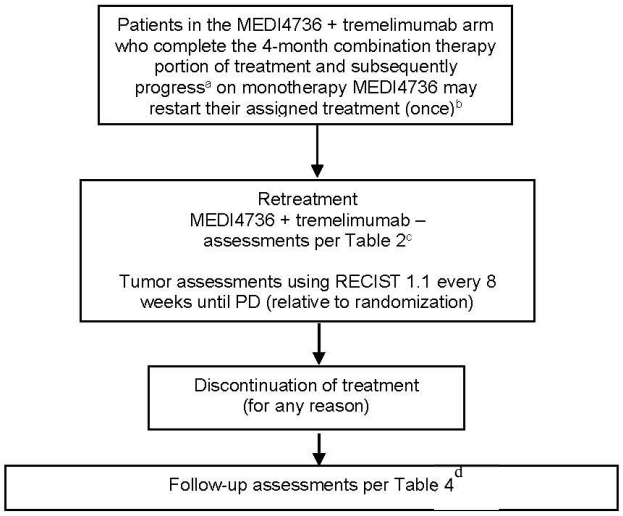



[a] Informed consent of study procedures and tumor biopsy sample may be obtained prior to the 28-day screening window, if necessary, in order to permit tumor biopsy sample acquisition and analysis prior to randomization

[b] Disease progression requires confirmation for patients receiving IMT. Disease progression in patients in the SoC arm should be confirmed if clinically feasible (see Section 5.1 of the CSP for more information).

[c] Patients in the MEDI4736 + tremelimumab combination therapy arm who are eligible for retreatment will be treated according to Figure 3
[d] This refers to Table 3 in the CSP
[e] This refers to Table 4 in the CSP
IMT Immunomodulatory therapy; PD Progressive disease; RECIST 1.1 Response Evaluation Criteria in Solid Tumors version 1.1.

**Figure 3 Study flow chart - optional retreatment for MEDI4736 + tremelimumab arm**




[a] With or Without confirmation
[b] Before restarting MEDI4736 + tremelimumab combination therapy, the Investigator should ensure that the patient does 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 until PD (relative to the date of restarting study treatment).
[c] This refers to Table 2 in the CSP. PK, ADA, and MDSC assessments do not need to be collected during retreatment.
[d] This refers to Table 4 in the CSP
ADA Anti-drug antibody; MDSC Myeloid-derived suppressor cell; PK Pharmacokinetics; q8w Every 8 weeks; RECIST 1.1 Response Evaluation Criteria in Solid Tumors version 1.1.

## 1.3 Number of subjects

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 1016 patients in order to randomize 760 eligible patients in a 2:1:1 ratio (380:190:190 patients) to MEDI4736 + tremelimumab combination therapy, MEDI4736 monotherapy, or SoC respectively.

The study was originally sized to characterize the OS benefit of MEDI4736 + tremelimumab combination therapy versus SoC in all-comers (i.e., regardless of PD-L1 tumor expression status) and in PD-L1 TC/IC subgroup population. The sizing assumes a 3-month delay in separation of the OS curves between each arm, hence the use of average HRs.

With amendment 11 (CSP version 12) the primary objective was changed to characterize the OS benefit of MEDI4736 monotherapy versus SoC in the PD-L1 TC/IC high subgroup. Due to this change the primary analysis of OS will be performed when approximately 147 death events have occurred in approximately 172 patients (85% maturity) across the MEDI4736 monotherapy and SoC treatment groups in the PD-L1 TC/IC high subgroup population. The number of patients (172) assumes that 45% of patients randomized are within this subgroup. No interim efficacy analyses will be performed in this study.

If superiority over SoC is found, a hierarchical multiple testing procedure (MTP) will be employed to strongly control for the overall type I error at the 5% alpha level (two-sided), as shown in Section 4, Figure 4.

**MEDI4736 Monotherapy versus SoC in the PD-L1 TC/IC high subgroup (OS)**

If OS at 24 months in the PD-L1 TC/IC high subgroup was 30% with MEDI4736 monotherapy and 12% with SoC (with a 10.1-month median OS), and assuming the true average OS HR is 0.59; the trial will have approximately 90% power to demonstrate statistical significance at the 5% level (using a 2-sided test) for the comparison of MEDI4736 monotherapy versus SoC in the PD-L1 TC/IC high subgroup, with approximately 147 events. With a 17-month recruitment period and a minimum follow-up period of 23.6 months from "last patient in" assumed, it is anticipated that the analysis will be performed approximately 40.6 months after the first patient has been recruited.

**MEDI4736 Monotherapy versus SoC in the low risk of EM subgroup (OS)**

Assuming approximately 80% of patients are in the low risk of EM subgroup and assuming the true average OS HR is 0.70, the trial will have approximately 81% power to demonstrate statistical significance at the 5% level (using a 2-sided test) for the comparison of MEDI4736 monotherapy versus SoC in the low risk of EM subgroup, with approximately 254 events.

**MEDI4736 Monotherapy versus SoC in the ctDNA TMB high (≥16 mut/Mb) subgroup (OS)**

Assuming approximately 16% of patients are in the ctDNA TMB high (≥16 mut/Mb) subgroup and assuming the true OS HR is 0.40, the trial will have approximately 82% power to demonstrate statistical significance at the 5% level (using a 2-sided test) for the comparison of MEDI4736 monotherapy versus SoC in the ctDNA TMB high (≥16 mut/Mb) subgroup, with approximately 48 events.

**MEDI4736 + tremelimumab combination therapy versus SoC in the ctDNA TMB high (≥16 mut/Mb) subgroup (OS)**

Assuming the true OS HR is 0.40, the trial will have approximately 92% power to demonstrate statistical significance at the 5% level (using a 2-sided test) for the comparison of MEDI4736 + tremelimumab combination therapy versus SoC in the ctDNA TMB high (≥16 mut/Mb) subgroup, with approximately 68 events.

### MEDI4736 monotherapy versus SoC in all-comers (OS)

If OS at 24 months in the all-comers population was 24% with MEDI4736 monotherapy and 12% with SoC (with a 10.1-month median) and assuming the true average OS HR is 0.70, the trial will have approximately 91% power to demonstrate statistical significance at the 5% level (using a 2-sided test) for the comparison of MEDI4736 monotherapy versus SoC in all-comers, with approximately 336 events. .

### MEDI4736 + tremelimumab combination therapy versus SoC-all comers (OS)

With approximately 570 patients randomized across the MEDI4736 + tremelimumab combination therapy and SoC treatment groups and a true average OS HR of 0.70, an estimated 495 death events (87% maturity) are expected to have occurred at 40.6 months from "first patient in," providing 97% power to demonstrate statistical significance at the 5% level (using a 2-sided test), for the comparison of MEDI4736 + tremelimumab combination therapy versus SoC in all-comers.

Table 5 provides a summary of these statistical assumptions and calculations.

**Table 5 Summary of Statistical Assumptions and Calculations**

| | N ratio | Overall HR | Landmarks at 24 months | Events (maturity) | Power | Critical values HR (landmarks) |
|---|---|---|---|---|---|---|
| **Primary Objectives** | | | | | | |
| OS: Mono versus SoC; PD-L1 TC/IC high subgroup | 86:86 | 0.59 | 30% vs 12% | 147 (85%) | 90% | 0.72 (22%) |
| **Secondary objectives:** | | | | | | |
| OS: Mono versus SoC; low risk of EM subgroup | 152:152 | 0.70 | 24% vs 12% | 254 (84%) | 81% | 0.78 (19%) |
| OS: Mono versus SoC; ctDNA TMB high (≥16 mut/Mb) subgroup | 30:30 | 0.40 | 43% vs 12% | 48 (80%) | 88% | 0.56 (29%) |
| OS: Combo versus SoC; ctDNA TMB high (≥16 mut/Mb) subgroup | 60:30 | 0.40 | 43% vs 12% | 68 (76%) | 96% | 0.62 (26%) |
| OS: Mono versus SoC; all-comers | 190:190 | 0.70 | 24% vs 12% | 336 (88%) | 91% | 0.81 (18%) |
| OS: Combo versus SoC; all-comers | 380:190 | 0.70 | 24% vs 12% | 495 (87%) | 97% | 0.83 (17%) |

Note: The sample size estimates in the PD-L1 TC/IC high subgroup comparisons assume that 45% of the patients enrolled are within this subgroup. The sample size estimates in the low risk of EM subgroup assume that 80% of the patients randomized are within this subgroup. The sample size estimates in the ctDNA TMB high (≥16 mut/Mb) subgroup assume that 16% of the patients randomized are within this subgroup. A 3-month delayed effect is assumed for all hypotheses except for ctDNA TMB high.

Combo MEDI4736 + tremelimumab combination therapy; HR Hazard ratio; IC tumor-associated immune cells; Mono MEDI4736 monotherapy; OS Overall survival; SoC Standard of Care; PD-L1 Programmed cell death ligand 1; PFS Progression-free survival; TC Tumor cells; TMB Tumor Mutational Burden.

# 2. ANALYSIS SETS

## 2.1 Definition of analysis sets

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

### 2.1.1 Full analysis set (FAS)

The full analysis set (FAS) will include all randomized patients (all-comers) (i.e., the Intent-to-Treat [ITT] population). 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.

### 2.1.2 PD-L1 TC/IC analysis set

The PD-L1 TC/IC analysis set will include the subset of patients in the FAS with PD-L1evaluable samples and available results, as determined by the analytically validated VENTANA PD-L1 (SP263) Assay. Subgroups are defined as PD-LI TC/IC high, low or unknown.

- PD-L1 TC/IC high defined as either ≥ 50% of the tumor cells (TC) or ≥ 25% of the immune cells (IC) staining for PD-L1 at any intensity if >1% of the tumor area contains IC, or ≥ 50% of TC or 100% of IC staining for PD-L1 at any intensity if 1% of the tumor area contains IC.
- PD-L1 TC/IC low is defined as not meeting any of the criteria for PD-L1 high.

The scoring algorithm was derived from data from two phase 2 studies, D4193C00001 (HAWK, DCO September 2016) and D4193C00003 (CONDOR, DCO March 2017).

### 2.1.3 Safety analysis set

All patients who received at least 1 dose of study treatment will be included in the safety analysis set. The patients will be classified on the basis of the treatment actually received. When assessing safety and tolerability, summaries will be produced based on the safety analysis set. Such summaries will be produced for all-comers and also for those in the PD-L1 TC/IC high and low risk of EM subgroups separately.

### 2.1.4 PK analysis set

All patients who received at least 1 dose of IP per the protocol for whom any post-dose PK data are available.

### 2.1.5 ctDNA TMB analysis set

The ctDNA TMB analysis set will include the subset of patients in the FAS with ctDNA TMB evaluable samples and available results. Subgroups are defined as ctDNA TMB high (≥16 mut/Mb), low (<16 mut/Mb) and unknown.
The "unknown" category includes the following ctDNA TMB outcomes: Failed, Not Done, Not Provided, and No Call/not evaluable.

### 2.1.6 Low risk of early mortality (EM) analysis set

The low risk of EM analysis set consists of patients identified as having low risk of early mortality based on laboratory values according to a specific prognostic model [redacted]

The detailed model specifications and description are documented in the SAP Appendix.

### 2.1.7 ADA analysis set

Patients in the safety analysis set with non-missing baseline ADA result and at least 1 post-baseline ADA result.

**Table 6 Summary of Outcome Variables and Analysis Populations**

| Outcome variable | Population |
|---|---|
| **Efficacy data** | |
| OS, PFS, ORR, APF6 APF12, OS12, OS18, OS24 | PD-L1 TC/IC analysis set, low risk of EM analysis set, ctDNA TMB analysis set and all-comers (FAS) |
| DoR, BoR, TTR, TFST, TSST, PFS2, PROs and symptom endpoints | PD-L1 TC/IC analysis set and all-comers (FAS) |
| Study population and demography | PD-L1 TC/IC analysis set and all-comers (FAS) |
| PK data | PK analysis set |
| WHO/ECOG performance status | PD-L1 TC/IC analysis set and all-comers (FAS) |

| Outcome variable | Population |
|---|---|
| **Safety Data** | |
| Exposure | Safety analysis set |
| AEs | Safety analysis set |
| Laboratory measurements | Safety analysis set |
| Vital signs | Safety analysis set |

AE Adverse event; APF12 Proportion of patients alive and progression free at 12 months; DoR Duration of response; EM Early mortality; IC tumor-associated immune cell; ITT Intent-to-Treat; ORR Objective response rate; OS Overall survival; OS12 Proportion of patients alive at 12 months; OS18 Proportion of patients alive at 18 months; OS24 Proportion of patients alive at 24 months after randomization; PD-L1 Programmed cell death ligand 1; PFS Progression-free survival; PFS2 Second progression; PK Pharmacokinetic; PRO Patient-reported outcome; TC tumor cell; TTR Time to response; TFST Time to first subsequent therapy; TSST Time to second subsequent therapy; TMB Tumor Mutation Burden. The ctDNA TMB cut-point of 16 mut/MB for MEDI4736 + tremelimumab and MEDI4736 monotherapy was derived from the EAGLE study (Li et al 2020).

## 2.2 Violations and deviations

The important protocol deviations (PDs) specified below will be summarized and listed. None of the deviations listed below 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 where feasible.

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). The key inclusion/exclusion criteria identified as the following:
    - (Inclusion 3): Histologically or cytologically confirmed recurrent or metastatic SCCHN (oral cavity, oropharynx, hypopharynx, or larynx) not amenable to local curative therapy with surgery or radiation therapy
    - (Inclusion 4): No prior systemic therapy for recurrent/metastatic disease. Systemic therapy given as part of multimodality treatment for locally advanced or locally recurrent disease is allowed
    - (Inclusion 7): Confirmed PD-L1–positive or –negative SCCHN by the Ventana 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 analytically validated Ventana assay as a part of the screening process for another AstraZeneca/MedImmune study, this test result can be used for the determination of eligibility

- 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 determined by 0% to 24% of tumor cells with membrane staining for PD-L1

- (Inclusion 8): World Health Organization (WHO)/ECOG performance status of 0 or 1

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

- (Exclusion 2): Tumor progression or recurrence within 6 months of last dose of platinum therapy given as part of multimodality treatment for locally advanced or locally recurrent disease

- (Exclusion 13) Patients with a history of brain metastases, spinal cord compression, or leptomeningeal carcinomatosis, or involvement of any other anatomic area that, in the opinion of the Investigator, may cause significant symptoms if an inflammatory reaction occurs

- Received prohibited concomitant medications (Deviation 2). Please refer to the CSP Section 7.7 for those medications that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to database lock. The output of important PDs due to use of prohibited concomitant medications will be based on Sponsor's manual review.

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.

In addition, all COVID-19 related non-important PDs and issues will be summarized and listed and included in the CSR. Treatment cycle and study visit delays resulting from COVID-19 are captured in Veeva Clinical Vault (VCV) as protocol deviations with the PD code COVID19. These PDs will be extracted from VCV, reviewed by the study team and categorised accordingly for analysis and reporting prior to database lock.

# 3. PRIMARY AND SECONDARY VARIABLES

## 3.1 Derivation of RECIST Visit Responses

The analyses of the RECIST 1.1 based secondary endpoints, PFS, ORR, DoR, TTR, BoR, APF6 and APF12, will be based on site investigator assessments using RECIST 1.1. In addition, PFS2 will be defined by local clinical practice.

For all patients, the RECIST version 1.1 (see further Appendix E 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) though 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 randomization and ideally should be performed as close as possible to the date of randomization (Tables 2 and 3 of CSP). Follow-up assessments will be performed every 6 weeks for the first 24 weeks (relative to the date of randomization) and then every 8 weeks as indicated in the schedule of procedures 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). 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 in 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 clinically 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.

RECIST 1.1 will be regarded as primary in terms of the efficacy analyses and RECIST 1.1 modified for confirmation of progression assessment is supportive (Section 3.1.3).

In the absence of significant clinical deterioration, the investigator should continue study treatment until progression is confirmed.

If progression is not confirmed, then the patient should continue on study 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, if clinically feasible, 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.

**Criteria for Retreatment for Patients in the MEDI4736 + tremelimumab arm**

Patients who complete the 4 dosing cycles of the MEDI4736 + tremelimumab combination therapy portion of the regimen (with clinical benefit per Investigator's judgment) but subsequently have evidence of PD during the MEDI4736 monotherapy portion of the combination regimen, with or without confirmation according to RECIST 1.1, may restart treatment with the entire combination regimen (including monotherapy maintenance).

For patients in the MEDI4736 + tremelimumab combination therapy arm, before a patient restarts treatment, the Investigator should ensure that the patient:

1. Does not have any significant, unacceptable, or irreversible toxicities that indicate continuing treatment will not further benefit the patient

2. Still fulfils the eligibility criteria for this study, including re-consenting to restart MEDI4736 + tremelimumab combination therapy

3. Has not have received an intervening systemic anticancer therapy after their assigned treatment discontinuation

4. Has had a baseline tumor assessment within 28 days of restarting their assigned treatment; all further scans should occur every 8 weeks relative to the date of restarting treatment until study treatment is stopped

5. Undergoes a tumor biopsy as described in Section 5.5.1 of the CSP

RECIST outcomes will be calculated using a computer program for both the BICR and site investigator data.

### 3.1.1 Blinded Independent Central Review (BICR) Assessment Using RECIST 1.1

A BICR of radiological imaging data will be conducted according to RECIST 1.1 and used for a sensitivity analysis for the secondary objective of PFS. This will use all radiological scans for all patients (including those at unscheduled visits or outside visit windows) obtained on or before Jan 8th, 2018 which was the data cut-off for the last BICR. (Cessation of the BICR at this data cut-off occurred due to PFS being changed to a secondary objective as documented in version 8 of the clinical study protocol so that the primary analysis will be conducted using investigator assessments instead of BICR).

Prior radiotherapy will also be provided to the BICR to allow the selection of appropriate target lesions.

The imaging scans will be reviewed by 2 independent radiologists using RECIST 1.1 and will be adjudicated, if required. For each patient, the BICR will define the visit response data (CR, PR, SD, PD, or not evaluable [NE]) and the relevant scan dates for each timepoint (i.e., for visits where response or progression is/is not identified). If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is evidence of progression, in which case the response will be assigned as PD). Sensitivity analysis of PFS will be derived from the visit response data.

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

### 3.1.2 Site Investigator Assessment Using RECIST 1.1

All RECIST 1.1 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 NE (unless there is evidence of progression in which case the response will be assigned as PD).

#### 3.1.2.1 Target lesions (TLs)

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

A patient can have a maximum of 5 measurable lesions recorded at baseline with a maximum of 2 lesions per organ (representative of all lesions involved suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). If more than one baseline scan is recorded, then measurements from the one that is closest to the date of randomization will

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

All other lesions (or sites of disease) not recorded as TL should be identified as 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 3.1.2.2 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 nadir (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.
- An 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.

**Irradiated lesions/lesion intervention**

Previously irradiated lesions (i.e. lesion irradiated prior to entry into the study) should be recorded as NTLs and should not form part of the TL assessment. Lesions in a previously irradiated field can be used as measurable disease provided that there has been demonstrated progression in the lesion.

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 irradiated lesions/lesion intervention)**

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

#### 3.1.2.2 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 assessment of NTLs 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. |
| 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.<br>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.

#### 3.1.2.3 Site Investigator Assessment Using RECIST 1.1: Overall visit response

Table 9 defines 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 |

**Table 9 Overall Visit Responses**

| Target Lesions | Non-target lesions | New Lesions | Overall Response |
|---|---|---|---|
| PD | Any | Any | PD |
| Any | PD | Any | PD |
| Any | Any | Yes | PD |
| NA | NA | No | NED |

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), NED No evidence of disease.

### 3.1.3 Site Investigator Assessment Using RECIST 1.1 Modified for Confirmation of Progression

PFS as determined by RECIST 1.1 modified for confirmation of progression will be calculated using investigator assessments as detailed in Sections 3.1. This means that the visit response of PD must be confirmed by another visit response of PD at the next scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinical deterioration and if clinically feasible. Confirmation of progression needs to be programmatically derived.

## 3.2 Outcome Variables

All RECIST assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues investigational product or receives another anti-cancer therapy. All efficacy and PRO data will be summarized and analysed using the full analysis set (FAS), refer to Table 6 for the analysis set for these variables.

### 3.2.1 Primary endpoint (OS)

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

Note: Survival calls will be made following the date of data cut-off for the analysis (these contacts should generally occur within 7 days of the data cut-off). 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 (as applicable under local laws).

#### 3.2.1.1 Progression free survival (PFS)

PFS (per RECIST 1.1 as assessed by 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 discontinues 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 consecutive missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the 2 missed visits (Note: NE visit is not considered a missed visit).

Given the scheduled visit assessment scheme (i.e., six-weekly for the first 24 weeks then eight-weekly thereafter) the definition of 2 missed visits will be as follows:

- If the previous RECIST assessment is <= study day 35 (i.e. week 5) then two missing visits will equate to 13 weeks since the previous RECIST assessment, allowing for a late visit (i.e. 2 x 6 weeks + 1 week for a late assessment = 13 weeks).
- If the previous RECIST assessment is >35 and < study day 120 (i.e. week 17) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks).
- If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to eight-weekly this will equate to 16 weeks (i.e. take the average of 6 and 8 weeks which gives 7 weeks and then apply same rationale, hence 2 x 7 weeks + 1 week for an early assessment + 1 week for a late assessment = 16 weeks). The time period for the previous RECIST assessment will be from study days 120 to 161 (i.e. week 17 to week 23).
- From week 23 (day 162) onwards (when the scheduling changes to eight-weekly assessments), two missing visits will equate to 18 weeks (i.e. 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks).

If the patient has no evaluable visits or does not have baseline data, they will be censored at Day 1 unless they die within 2 consecutive visits of baseline, then they will be treated as an event with date of death as the event date.

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

RECIST 1.1 assessments/scans contributing toward 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 progression.

- For BICR assessments, date of progression will be determined from the progression date (PROGSCDT) provided by the BICR progression scan dates (PROGSCDT) dates.

- 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.

PFS based on RECIST 1.1 modified for confirmation of progression may be performed for sensitivity analysis using the algorithm described above for the RECIST 1.1, but following a modification whereby any objective disease progression must be confirmed by the next scheduled scan. The confirmatory scan must be no sooner than 4 weeks after the initial documented progression. If disease progression is confirmed (or disease progression occurs and no further scans are recorded), then the date of progression will be when it was originally observed. Patients with a single disease progression and no further tumor assessment scans will be treated as PD in the analysis. In the absence of significant clinical deterioration, the investigational site is advised to continue the patient on their randomized MEDI4736 + tremelimumab combination therapy or MEDI4736 monotherapy until progression has been confirmed. 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 confirmatory scan is required for all patients in the SoC group, if clinically feasible, even if a subsequent treatment is started.

### 3.2.2 Objective response rate

ORR (per RECIST 1.1 using investigator assessments) is defined as the number (%) of patients with at least 1 visit response of CR or PR. Data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue study treatment without progression, receive a subsequent therapy, and then respond will not be included as responders in the ORR.

Patients who continue study treatment through progression and subsequently respond with further treatment would not be included as responders in ORR assessment. Likewise, patients on the MEDI4736 + tremelimumab combination therapy arm who progress during the monotherapy portion of the regimen and subsequently respond with the optional retreatment regimen would not be included as responders in the ORR assessment.

### 3.2.3 Duration of response

DoR (per RECIST 1.1 as assessed by 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 (i.e., date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the RECIST 1.1 PFS endpoint.

The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of CR or PR. If a patient does not progress following a response, then their DoR will be censored at the PFS censoring time. DoR will not be defined for those patients who do not have documented response.

### 3.2.4 Time to response

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

### 3.2.5 Proportion of patients alive and progression free at 6 and 12 months

The APF6 and APF12 will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed by investigator) at 6 months and 12 months respectively.

### 3.2.6 Proportion of patients alive at 12, 18 and 24 months

The OS12, OS18 and OS24 will be defined as the Kaplan-Meier estimate of OS at 12, 18 and 24 months respectively after randomization.

### 3.2.7 Time from randomization to the first subsequent therapy or death (TFST)

Time to the first subsequent therapy (TFST) 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.

### 3.2.8 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.

### 3.2.9 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 (TFST) or death. Note the patient does not need to remain on their second treatment for the progression event to occur. The date of second progression will be recorded by the Investigator in the eCRF (PFS2 CRF) 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 later of the date of assessment on the PFS2 form indicating no progression (PFSRESP=2) or at the date of the last evaluable RECIST assessment (or day 1 if no post-baseline RECIST data is available; 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).

### 3.2.10 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 site investigator data based upon RECIST assessments. Target lesions are measurable tumor lesions. 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 3.1.

### 3.2.11 Best objective response (BoR)

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST 1.1 assessment, as described in Appendix F. It is the best response a patient has had during their time in the study up until RECIST 1.1 progression or the last evaluable assessment in the absence of RECIST 1.1 progression.

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

BoR will be determined programmatically based on RECIST 1.1 using all site investigator assessments up until the first progression event. For patients whose progression event is death, BoR will be calculated based upon all evaluable RECIST 1.1 assessments prior to death.

For determination of a best response of CR/PR, at least one determination of CR/PR must be observed before progression disease (PD)

For determination of a best response of SD, at least one SD should be recorded at 5 weeks (i.e., 6 weeks minus 1 week), i.e. at least 35 days (to allow for an early assessment within the assessment window), after date of randomization (and not qualifying for CR or PR).

For PD, patients who die with no evaluable RECIST 1.1 assessments, if the death occurs ≤98 days (i.e., 2 × (6 weeks ±1week)) after randomization, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST 1.1 assessments, if the death occurs >98 days (i.e., 2 × (6 weeks ±7 days)) after the date of randomization, then BoR will be assigned to the NE category.

Progression events that have been censored due to them being >90 days because of 2 or more consecutive missed visits will not contribute to the BoR derivation.

## 3.3 Safety

Safety and tolerability will be assessed in terms of adverse events (AEs) (including serious adverse events [SAEs]), deaths, laboratory data, vital signs, electrocardiograms (ECGs) and exposure. These will be collected for all patients. 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) regardless of re-treatment period. For AEs, on treatment (or treatment-emergent AEs) will be defined as any AEs that started after dosing or prior to dosing and which worsens following exposure to the treatment. If an AE is not worse than the baseline (pre-dose) severity, then it will not be classified as TEAE.

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

### 3.3.1 Adverse events (AEs)

AEs and SAEs will be collected from the time of 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) regardless of whether the AE occurred during re-treatment with tremelimumab or not. 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.

The Medical Dictionary for Regulatory Activities (MedDRA) dictionary 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.

Currently, these AESI's have been identified in the CSP as Diarrhea/Colitis and intestinal perforation, Pneumonitis/ILD, hepatitis/transaminase increases, Endocrinopathy (i.e. Events of hypophysitis/hypopituitarism, adrenal insufficiency, and hyper- and hypothyroidism and type I diabetes mellitus), Neuropathy/neuromuscular toxicities (eg Guillain-Barre and myasthenia gravis), Rash/Dermatitis, Nephritis/blood creatinine increases, Pancreatitis/serum lipase or amylase increases, myocarditis, myositis/polymyositis and other inflammatory responses that are rare/less frequent with a potential immune-mediated aetiology including, but not limited to, pericarditis, sarcoidosis, uveitis and other events involving the eye, skin, haematological and rheumatological events. .

An AstraZeneca medically qualified expert, after consultation with the Global Patient Safety Physician, has reviewed the AEs of interest and identified which preferred terms contribute to each AESI. A further review will take place prior to database lock 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.

**AEs of possible interest (AEPIs)**

Adverse events of possible interest (AEPIs) are AEs that also have a potential inflammatory or immune-mediated pathophysiological basis resulting from the mechanism of action of durvalumab and/or tremelimumab but are more likely to have occurred due to other pathophysiological mechanisms and, thus, the likelihood the event is inflammatory or immune-mediated in nature is not high and/or is most often or usually explained by the other causes. These AEs not routinely arising from an inflammatory or immune-mediated mechanism of action – typically quite general clinical terms that usually present from a multitude of other causes –are classified as AEPIs.

A framework of questions was developed by the Sponsor to identify event terms that may have an inflammatory or immune-mediated mechanism of action but implicitly have a lower likelihood of that than other causes and, thus, should be included as an AEPI not an AESI term.

**Immune-mediated Adverse Events (imAE)**

Immune-mediated Adverse Events (imAE) will be identified from both AESIs and AEPIs based on programmatic rules that consider interventions involving systemic steroid therapy, immunosuppressant use, and/or endocrine therapy (which, in the case of AEPIs, occurs after first considering an Investigator's causality assessment and/or an Investigator's designation of an event as immune-mediated). Further details are provided in an imAE Charter v6.

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

**Infection Adverse Events**

Infection is an identified risk for another anti-PDL1 drug atezolizumab and listed in the Warnings and Precautions of its US label. Infection AEs will be presented in the CSR.

### 3.3.2 Treatment exposure

Exposure will be defined separately for MEDI4736 monotherapy, MEDI4736 on the MEDI4736+ tremelimumab combination arm, tremelimumab on the MEDI4736+tremelimumab combination arm, and also separately for cetuximab, a platinum [carboplatin or cisplatin], and 5FU on the standard of care arm for the initial period of treatment and for the retreatment period as follows:

Total (or intended) exposure of MEDI4736 (monotherapy)

- Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to earliest of "last dose date of study drug + 27 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 + 27 days" or death date or DCO regardless of tremelimumab being re-started or not.

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. This will include the sum of the duration in the 2 tremelimumab periods (i.e., initial period and re-treatment) but excluding the count of time between the 2 periods (for example, the duration would sum up to 6 months if they did 3 months in each case even if the 2 periods were separated by several months).

Actual exposure of MEDI4736/tremelimumab

- Actual exposure is defined as above, but excluding total duration of dose delays. Actual treatment duration for tremelimumab will be derived similarly as was done by total exposure, that by combining the periods of tremelimumab and summing up the delays in each period.

Total (or intended) exposure for the SOC regimens (cetuximab, a platinum [carboplatin or cisplatin], and 5FU)

- The total (or intended) exposure for the individual SOC regimens will be calculated according to the dose schedule required for each SOC regimen as follows:

  (a) Cetuximab: Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of "last dose date of study drug + 6 days" or death date or DCO

  (b) Platinum and 5FU: Total (or intended) exposure is defined as the total treatment period from first dose date of study drug to the earliest of "last dose date of study drug + 20 days" or death date or DCO

The total (or intended) exposure will also be summarised by combining the SOC treatments together. The following rules will be applied for SOC total

- 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 YY SoC drug + xx days" or death date or DCO, xx depends on whether the patient was still receiving platinum or 5FU, or only on cetuximab:
  - if last dose of cetuximab is $\geq$14 days after last dose of platinum and 5FU, then YY=cetuximab and xx= 6 days
  - if last dose of cetuximab is < 14 days after last dose of platinum and 5FU, then YY=platinum and xx= 20 days

Actual exposure will not be calculated for SOC.

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

Exposure will also be measured by the number of cycles received. For the immunotherapy arms, a cycle corresponds to a 4-week treatment; for SOC, a cycle corresponds to a 3-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 for the calculation of exposure.

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

MEDI4736 (monotherapy):

- For all dosing dates:
  Total duration of dose delays= Sum of (Date of the dose - Date of previous dose – 28 days)
  Thus, if no delays were encountered, the duration would sum up to 0, since infusions were done every four weeks.

MEDI4736 (given in combination):

- Duration of dose delays= Sum of (Date of the dose - Date of previous dose – 28 days); this will include re-treatment period as was done for the exposure derivation

Tremelimumab (given in combination):

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

### 3.3.3 Dose intensity

Dose intensity will be derived for the initial treatment period and the re-treatment period for the immunotherapy agents. It will also be derived for each of the SOC agents (cetuximab, a platinum [carboplatin or cisplatin], and 5FU). 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.

For patients who started cisplatin and later switched to carboplatin, the last day of dosing for both the intended and actual dosing for cisplatin will be the last day prior to switching to carboplatin, and for carboplatin, day 1 will be the first day of dosing of carboplatin in these patients (e.g., if the first day of carboplatin was given on study day 22, then day1 for carboplatin=study day 22, etc). If patients switch from carboplatin to cisplatin in any way other than once from cisplatin to carboplatin, no RDI will be calculated for them.

In cases where the units for the recorded dose is different from the planned dose, the following conversion should be applied when feasible:

1. 100mg/m$^2$ of body surface area: If the recorded dose is in mg instead of mg/m$^2$, height and weight (use most current weight) will be used to derive body surface area using Du Bois formula)

2. AUC of 5 mg/mL/min: If the recorded dose unit is in mg instead of AUC then the following algorithm should be used to convert doses in mg to AUC

    1. $GFR = Sex * \left(\frac{(140 - Age(years))}{(Creatinine(\text{mg/dL}))}\right) * (Weight(kg)/72)$, values for sex: Male=1, Female=0.85

    2. $TargetAUC = Carboplatin\ Dose\ (mg)/(GFR + 25)$

    3. $CarboplatinDose = TargetAUC * (GFR + 25)$

Note: For MEDI4736 monotherapy in the combination arm and tremelimumab, RDI would use the intended dose for the 2 periods (initial and re-treatment periods) combined.

### 3.3.4 Laboratory data

Laboratory data will be collected throughout the study, from screening to the follow-up visits as described in Tables 2, 3 and 4 of the CSP. Blood and urine samples for determination of hematology, clinical chemistry, and urinalysis will be collected as described in Section 5.2.1

of the CSP. For derivation of baseline and post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 3.3.7 below will be used.

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

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

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

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

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

Local reference ranges will be used for the primary interpretation of laboratory data. The denominator used in laboratory summaries of CTCAE 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.

### 3.3.5 ECGs

ECG data obtained up until 90 days after date of last dose of study treatment, or until initiation of another therapy, will be used for reporting. For derivation of post baseline visit values considering visit window and to handle multiple records present in any visit window, derivation rules as described in Section 3.3.7 below will be used.

Investigator's assessment of the ECG will be collected locally.

### 3.3.6 Vital signs

Vital signs (blood pressure [BP], pulse, respiratory rate, and temperature) will be evaluated according to the assessment schedules (Tables 2, 3, and 4 from CSP).

Vital signs data obtained up until 90 days after date of last dose of study treatment, or until initiation of another therapy, will be used for reporting. Change from baseline in vital signs variables will be calculated for each post-dose visit on treatment. For derivation of post baseline visit values considering visit window and to handle multiple records, derivation rules as described in Section 3.3.7 below will be used.

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

### 3.3.7 General considerations for safety assessments

Time windows will need defining for any presentations that summarize values by visit. (Note: this also applies to the PRO data, as well as safety analyses) 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 0). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day.

- For 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. Furthermore:-

- 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 start AE dates, the following will be applied

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

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

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

- 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 last dose of study drug then impute last dose date.

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

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

   a. For Missing day only – using the 1st of the month

   b. For Missing day and month -using the 1st of January.

### 3.4 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).

#### 3.4.1 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 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.

**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.

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.

### Time to symptom deterioration

For each of the symptom 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.

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.

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.

### 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 ≤90.

### 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
- 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.
- Compliance rate = Evaluable/Expected ×100
- Evaluability rate = Evaluable/Received ×100

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.

### 3.4.2 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. CCI [redacted]

**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 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 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 has 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. The denominator will consist of a subset of the FAS who have a baseline symptom score ≥10.

### 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

## 3.4.3 CCI

CCI

CCI [redacted]

### 3.4.4 CCI [redacted]

CCI [redacted]

### 3.4.5 CCI [redacted]

CCI [redacted]

### 3.4.6 CCI [redacted]

CCI [redacted]

## 3.5 Pharmacokinetics and Immunogenicity Variables

Analyses to evaluate the pharmacokinetics of MEDI4736 monotherapy, and MEDI4736 + tremelimumab combination therapy and to investigate the immunogenicity of MEDI4736 and tremelimumab will be performed by a designated third party on behalf of AstraZeneca.

### 3.5.1 Population of pharmacokinetics and exposure-response/safety analysis

A population PK model may be developed using a non-linear mixed-effects modelling approach for MEDI4736 monotherapy and for MEDI4736 in combination with tremelimumab. The impact of physiologically-relevant patient characteristics (covariates) and disease on PK of both agents may be evaluated. The relationship between MEDI4736 PK exposure or tremelimumab PK exposure and the effect on safety and efficacy endpoints may be evaluated. The results of such an analyses, if conducted, will be reported outside of the CSR.

### 3.5.2 Pharmacokinetic analysis

Pharmacokinetic data analyses of MEDI4736 and tremelimumab will be performed by the MedImmune Global Pharmacokinetics Pharmacodynamics (PK-PD) & Bioanalysis group or designee. No formal non-compartmental (NCA) PK analysis will be conducted due to the sparse PK sampling scheme for MEDI4736 and tremelimumab, which does not allow for meaningful determination of PK parameters.

MEDI4736 and tremelimumab concentration data and summary statistics will be tabulated by treatment arm and visit. PK parameters, such as peak and trough concentration of MEDI4736 and tremelimumab, will be derived from raw data measurements as data allow.

For data below limit of quantification (BLQ) the following rules will apply:

- If, at a given time point, 50% or less of the serum concentrations are not quantifiable (NQ), the geometric mean, CV%, geometric CV%, mean and SD will be calculated by substituting the limit of quantification (LOQ) for values which are NQ.
- If more than 50%, but not all, of the concentrations are NQ, the geometric mean, CV%, geometric CV%, mean and SD will be reported as not calculable (NC).
- If all the concentrations are NQ, the geometric mean and mean will be reported as NQ and the CV%, geometric CV% and SD as NC.

ALQ results of the PK assay will be excluded from summary statistics calculations.

If two results are obtained for one PK sample, the following rules will be applied:

- If the results are ≤ +/- 30% of each other (by calculating the relative percent difference (RPD)), then the mean of the two results will be reported

- If the results are > +/- 30% of each other (RPD), then this will be reported as NRR (No Reportable Result)

- If one result is BLQ and one > lower limit of quantification (LLOQ), this will be reported as NRR

The RPD will be calculated using the following formula:

$$RPD\ (\%) = \frac{Result\ 1 - Result\ 2}{Mean\ of\ Results\ 1\ and\ 2}\ x\ 100$$

### 3.5.3 Immunogenicity analysis

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

The number and percentage of patients who develop detectable ADA to MEDI4736 or tremelimumab will be determined for each of the following categories for each treatment group. The denominator for percentage calculations is the number of ADA-evaluable patients in the treatment group.

- ADA positive at any visit, baseline and/or post-baseline. The percentage is known as ADA prevalence.
- ADA positive post-baseline and positive at baseline
- ADA positive post-baseline and not detected at baseline (treatment-induced ADA positive)
- ADA not detected post-baseline and positive at baseline
- Baseline ADA titer that was boosted by >=4-fold following drug administration (treatment-boosted ADA positive)
- Treatment-emergent ADA positive, defined as the sum of treatment-induced ADA positive and treatment-boosted ADA positive. The percentage is known as ADA incidence.

- Persistently positive ADA, defined as having at least 2 post-baseline ADA positive measurements with at least 16 weeks (112 days) between the first and last positive measurements, or an ADA positive result at the last available assessment
- Transiently positive ADA, defined as having at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive
- nAb positive at baseline and/or post-baseline

### 3.6 Biomarker variables

PD-L1 expression status used as stratification factor for randomization 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.

## 4. ANALYSIS METHODS

The main hypothesis to be tested is improved OS in PD-L1 TC/IC high subgroup population treated with MEDI4736 monotherapy and SoC.

The study will be considered positive (a success) if this hypothesis is statistically significant. A multiple testing procedure has been implemented to strongly control the type I error rate at 5%

Additional hypotheses will be tested using a multiple testing procedure, which is outlined in Figure 4. The procedure will define the significance levels that should be applied to the interpretation of the nominal p-values.

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

- $H_0$: No difference between MEDI4736 monotherapy and SoC
- $H_1$: Difference between MEDI4736 monotherapy and SoC

The primary objective is OS in the PD-L1 TC/IC high subgroup for MEDI4736 monotherapy versus SoC. The study was originally sized to characterize the OS benefit of MEDI4736 + tremelimumab combination therapy versus SoC.

With CSP version 12, the primary objective was changed to characterize the OS benefit of MEDI4736 monotherapy versus SoC in the PD-L1 TC/IC high subgroup, due to this change the analysis of OS will be performed when approximately 147 death events from the PD-L1 TC/IC high subgroup population have occurred in approximately 172 patients (85% maturity) across the MEDI4736 monotherapy therapy and SoC treatment groups. All other endpoints will be analysed at this time. No interim analyses will be performed for any efficacy endpoint

**Multiple testing strategy**

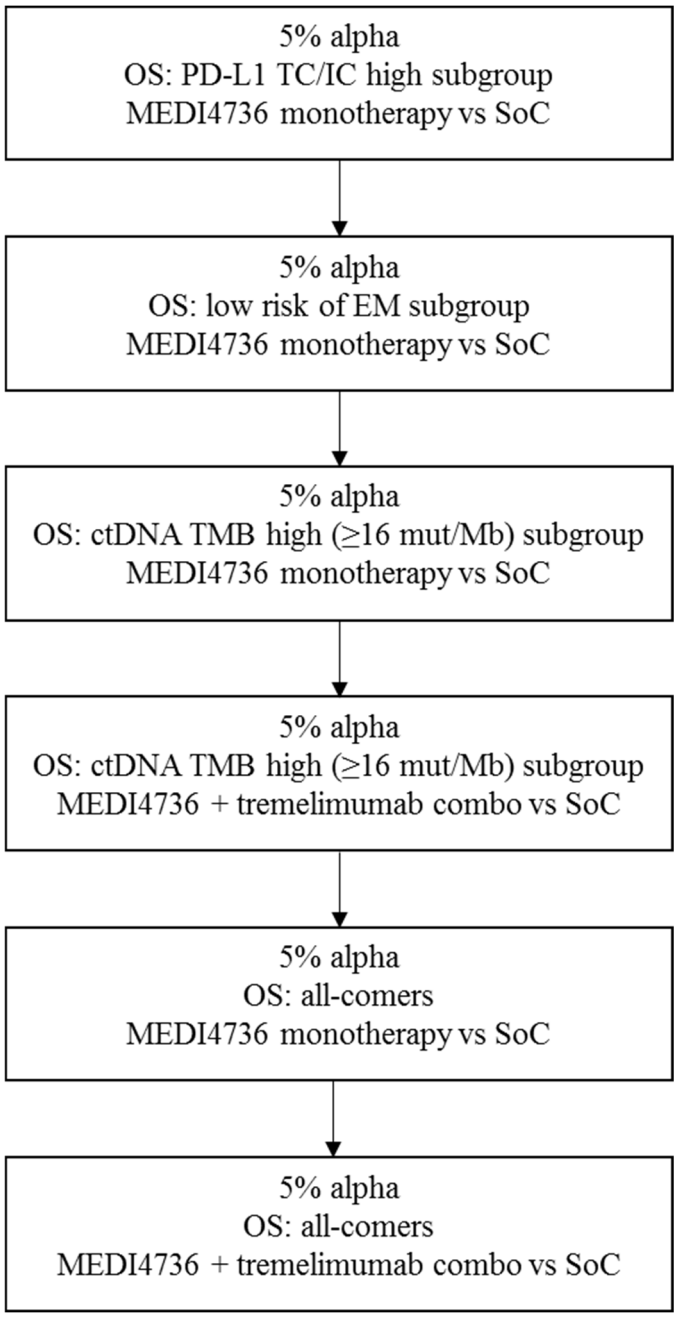
In order to strongly control the type I error at 5% 2-sided, a multiple testing procedure (MTP) will be used across the primary endpoint (OS), treatment regimens (MEDI4736 monotherapy, MEDI4736 + tremelimumab, and SoC), and across the analysis populations (PD-L1 TC/IC

high subgroup, low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers). If the highest-level hypothesis in the MTP is statistically significant, the next hypothesis in the hierarchy will then be tested as shown in Figure 4).

Hypotheses will be tested using a hierarchical multiple testing procedure to control for the overall type I error at the 0.05 level (two-sided). With this approach, hypotheses will be tested in a predefined order (sequentially rejective manner at α=0.05) as outline in Figure 4. Hypotheses in the hierarchy will be to test MEDI4736 monotherapy versus SoC in the PD-L1 TC/IC high subgroup population followed by MEDI4736 monotherapy versus SoC in the low risk of EM subgroup population, then MEDI4736 monotherapy versus SoC in the ctDNA TMB high (≥16 mut/Mb) subgroup, then MEDI4736 + tremelimumab combination therapy versus SoC in the ctDNA TMB high (≥16 mut/Mb) subgroup, then MEDI4736 monotherapy versus SoC in all-comers and finally MEDI4736 + tremelimumab combination therapy versus SoC in all-comers. All tests will be at the two-sided 5% alpha level and the procedure will stop at the first failure to find statistically significant evidence of superiority or if tests of all hypotheses show statistically significant evidence of superiority over SoC.

Note: the comparisons of MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy and the comparison of MEDI4736 + tremelimumab combination therapy versus SoC (in the PD-L1 TC/IC high subgroup and low risk of EM subgroup) will not be included in the MTP, and therefore will not be conducted under strict alpha control.

**Figure 4 Multiple testing procedure for controlling the type 1 error rate**




Combo MEDI4736 + tremelimumab combination therapy; EM Early mortality; IC tumor associated immune cell; Mono MEDI4736 monotherapy; OS Overall survival; PD-L1 Programmed death ligand 1; SoC Standard of Care; TC tumor cell; TMB Tumor Mutational Burden.

## 4.1 General principles

IVRS based stratification factors will be used for the analysis of primary objectives as well as the analyses for other secondary objectives namely; PFS, ORR, PFS2, TFST, TSST, Time to

deterioration of EORTC QLQ-C30 and EORTC QLQ-H&N35, improvement rate of EORTC QLQ-C30 and EORTC QLQ-H&N35.

The subgroup analyses for the stratification factors as well as all other factors will be based on values recorded on the CRF. 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 objectives, will be performed using CRF based stratification factors. For this sensitivity analysis if there are missing baseline covariate values, patients will be retained in the analysis by including the missing data as another stratification level.

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.

- All models will include 3 treatment groups and any pairwise comparisons will then be extracted.

- SAS® version 9.4 or higher will be used for all analyses.

- Palliative radiation will not be considered as a subsequent therapy.

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

Outcomes variables will be summarized and analyzed based on the analysis populations detailed in Tables 6 and 12. All outputs will be summarized by treatment arm.

In terms of retreatment, individual efficacy response data from the re-treatment phase will only be listed, however data from retreatment will be included in all PRO analyses. Any

derivations relative to baseline (e.g., day, RECIST derivations) in the retreatment period will be relative to the baseline scan prior to the retreatment period.

Note that the AEs, SAEs and AESI/AEPI events will be reported regardless of whether the event occurred during re-treatment period or not. Additionally, if there are sufficient numbers of retreated patients to warrant separate summaries following combinations of summaries would also be produced:

The period for assessing safety data from the retreatment phase will start at the date of first dose in the retreatment phase

- A small set of headline summaries of AE data from the retreatment phase only.

Table 12 details which efficacy endpoints are to be analyzed, together with pre-planned sensitivity analyses indicating which analysis is regarded as primary for that endpoint. Formal statistical analysis will be done for only OS endpoint. Nominal p-values will be provided for all other endpoint and sensitivity analyses. No adjustments for multiple testing of endpoints or subgroups will be made.

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

| Endpoints Analyzed | Notes |
|---|---|
| OS | Stratified log-rank tests and Cox proportional method for:<br>Primary objective<br>– MEDI4736 monotherapy versus SoC in the PD-L1 TC/IC high subgroup (stratified by PD-L1 status, tumor location, and smoking history)<br>Secondary objectives<br>– MEDI4736 monotherapy versus SoC in the low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers; (stratified by PD-L1 status, tumor location, and smoking history)<br>– MEDI4736 + tremelimumab combination therapy versus SoC in the PD-L1 TC/IC high subgroup, low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers (stratified by PD-L1 status, tumor location, and smoking history)<br>– MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy in the PD-L1 TC/IC high subgroup, low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers; (stratified by PD-L1 status, tumor location and smoking history)<br>Sensitivity analysis:<br>– Sensitivity analysis using the stratified Max-combo test |

| Endpoints Analyzed | Notes |
|---|---|
| PFS | Stratified log-rank tests and Cox proportional method for:<br>Secondary objectives using investigator assessments (RECIST 1.1):<br>– MEDI4736 monotherapy versus SoC for the PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers; (stratified by PD-L1 status, tumor location and smoking history)<br>– MEDI4736 + tremelimumab combination therapy versus SoC for PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers; (stratified by PD-L1 status, tumor location, and smoking history)<br>– MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy for the PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers; (stratified by PD-L1 status, tumor location and smoking history)<br>Sensitivity Analyses:<br>– Sensitivity analysis based on RECIST 1.1 modified for confirmation of progression using site investigator assessments<br>– Sensitivity analyses using BICR assessments (RECIST 1.1) |
| OS12, OS18 and OS24 | Secondary analysis:<br>Kaplan-Meier estimates at 12, 18 and 24 months for:<br>– MEDI4736 monotherapy therapy versus SoC for the PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers<br>– MEDI4736 + tremelimumab combination therapy versus SoC for the PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers |
| ORR | Logistic regression for:<br>Secondary objectives<br>– MEDI4736 monotherapy therapy versus SoC for the PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers using site investigator RECIST 1.1 assessments<br>– MEDI4736 + tremelimumab combination therapy versus SoC for the PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers using site investigator RECIST 1.1 assessments<br>– MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy for the PD-L1 TC/IC high subgroup; low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers using site investigator RECIST 1.1 assessments |
| BoR | Secondary objective<br>Descriptive statistic: N (%) using site Investigator data (RECIST 1.1) for all-comers and the PD-L1 TC/IC high subgroup populations |
| DoR | Secondary objective<br>Descriptive statistics and Kaplan-Meier plot for all-comers and the PD-L1 TC/IC high subgroup populations |

| Endpoints Analyzed | Notes |
|---|---|
| TTR | Secondary objective:<br>Kaplan-Meier plots and estimates using site investigator data (RECIST 1.1) for all-comers and the PD-L1 TC/IC high subgroup populations |
| APF6 and APF12 | Secondary objective:<br>Kaplan-Meier estimate of PFS at 6 and 12 months for all-comers and the PD-L1 TC/IC high subgroup populations |
| PFS2 | Secondary objective:<br>Stratified log-rank test for all-comers and the PD-L1 TC/IC high subgroup populations |
| TFST and TSST | Secondary objective:<br>Stratified log-rank test for all-comers and the PD-L1 TC/IC high subgroup populations |
| Time to deterioration EORTC QLQ-C30 and EORTC QLQ-H&N35 endpoints) | Stratified log-rank test for all-comers and the PD-L1 TC/IC high subgroup populations |
| Improvement rate (EORTC QLQ-C30 and EORTC QLQ-H&N35 endpoints) | Logistic regression<br>For all-comers and the PD-L1 TC/IC high subgroup populations |
| Average change from baseline (EORTC QLQ-C30 and EORTC QLQ-H&N35 endpoints) | Mixed effect model repeated measurement (MMRM) analysis. Descriptive statistics including change from baseline<br>for all-comers and the PD-L1 TC/IC high subgroup populations |

APF12 Proportion of patients alive and progression free at 12 months; BICR Blinded Independent Central Review; DoR Duration of response; EM Early mortality; EORTC European Organisation for Research and Treatment of Cancer; EQ-5D-5L EuroQol 5-Dimension, 5-Level health state utility index; HR Hazard ratio; ITT Intent-to-Treat; ORR Objective response rate; OS Overall survival; OS12 Proportion of patients alive at 12 months after randomization; OS18 Proportion of patients alive at 18 months after randomization; OS24 Proportion of patients alive at 24 months after randomization; ctDNA circulating tumor DNA; PD-L1 Programmed death ligand 1; PFS Progression-free survival; QLQ-C30 30-item core quality of life questionnaire; QLQ-H&N35 35-item head and neck quality of life questionnaire; RECIST 1.1 Response Evaluation Criteria in Solid Tumors version 1.1; TTR Time to response, MMRM Mixed effect model repeat measurement TMB Tumor Mutation Burden;

## 4.2 Analysis methods

### 4.2.1 Analysis of primary variable (OS)

The analysis of OS will be done using stratified log-rank test stratified by PD-L1 status (positive versus negative), tumor location (OPC versus non-OPC, with a subsequent adjustment for HPV status in patients with OPC), and smoking history (>10 versus ≤10 pack-years) for generation of the p-value.

The stratification factors will be based on the values entered into IVRS at randomization even if it is subsequently discovered that these values were incorrect. Sensitivity analysis using the

correct values will be performed if more than 10% of the subjects were incorrectly categorized in the wrong strata.

The effect of MEDI4736 monotherapy versus SoC treatment in the relevant analysis populations (Table 12) will be estimated by the HR together with its corresponding 95% CI. The hazard ratio (HR) of OS, along with its 95% confidence interval using the same strata information as above will be estimated using a stratified Cox proportional hazards model with treatment as the only covariate (with ties= Efron and any stratification variables included in the strata statement) and the CI calculated using a profile likelihood approach.

Kaplan-Meier plots of OS will be presented by treatment arm. 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.

An analysis of OS will again be performed to compare MEDI4736 + tremelimumab combination therapy versus SoC and MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy for the analysis populations described in Table 12.

**Impact of switching treatment post study treatment discontinuation to immunotherapies (or other potentially active investigational agents) on OS analyses**

CCI

[redacted]

The following treatment comparisons will be performed, based on the primary analysis model:

- PD-L1 TC/IC high subgroup – MEDI4736 monotherapy versus SoC
- All-comers - MEDI4736 monotherapy versus SoC
- All-comers – MEDI4736 + tremelimumab combination therapy versus SoC

**Sensitivity Analysis for OS variable**

A sensitivity analysis for OS will examine the censoring patterns to rule out attrition bias with regards to the primary treatment comparison. This analysis will be supported by Kaplan-Meier plot of time to censoring where the censoring indicator of OS is reversed.

In addition, to support OS analyses, 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)**

As a sensitivity analysis, an unstratified Cox-proportional hazards modeling will be employed to assess the effect of covariates on the HR estimate for the primary analysis, i.e., the MODEL statement will include the treatment group variable and also the stratification factors (PD-L1 Status, smoking status and tumor location) as main effects. Hazard ratio with associated 95% CI will be based the profile likelihood approach.

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.

Moreover, another sensitivity analysis will be performed to evaluate the treatment effect, adjusted for pre-specified baseline prognostic factors. This analysis will be performed for the PD-L1 TC/IC high subgroup, low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers. For this analysis, a stratified Cox model adjusted for the following covariates will be used. The covariates are:

- Sex at randomization (Male, Female)
- Age at randomization (<65, ≥65<75, ≥75 years of age)
- Race (Asian, non-Asian)
- WHO/ECOG performance status (0, ≥1)
- Primary Tumor location/HPV (OPC/Positive, OPC/Negative, non-OPC/Any HPV)

- Extent of Disease (only locoregionally recurrent, metastatic with or without locoregional recurrence)

The Hazard ratio with two-sided 95% confidence interval will be based on profile likelihood approach.

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.

**Assumptions of Proportionality**

For OS, the assumption of proportionality will be assessed. Proportional hazards will be checked 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 may be investigated.

The Grambsch-Therneau non-proportionality test may also be used to check violation of the proportional hazards assumption (Grambsch and Therneau 1994). If a lack of proportionality is evident, the stratified max-combo test (Karrison 2016) will be conducted as a sensitivity analysis for OS to test for treatment differences in the case of non-proportional hazards. The max-combo test is based on an adaptive procedure optimizing test statistics among the log-rank test ($FH^{0,0}$) and the Fleming-Harrington (FH) test ($FH^{0,1}$; $FH^{1,0}$) with alpha correction, and it is recommended by the Cross-Pharma Non-Proportional Hazards Working Group in the presence of non-proportional hazards (Lin et al 2020). This analysis will be produced by the Sponsor.

### 4.2.2 Subgroup analysis OS

Subgroup analyses will be conducted comparing OS between MEDI4736 + tremelimumab combination therapy versus SoC, MEDI4736 monotherapy versus SoC and MEDI4736+ tremelimumab combination therapy versus MEDI4736 monotherapy (in the PD-L1 TC/IC high subgroup, low risk of EM subgroup, ctDNA TMB high (≥16 mut/Mb) subgroup and all-comers) for the following subgroups (but not limited to):

- Sex (male, female)
- Age at randomization (<65, ≥65<75, ≥75 years of age)

- PD-L1 TC status for stratification (positive, negative)
- CCI [redacted]
- PD-L1 TC/IC status (TC≥50 or IC≥25 (High), TC<50 and IC<25(Low))
- ctDNA TMB (high (≥16 mut/Mb), low (<16 mut/Mb), unknown)
- Tumor location/HPV (OPC/Positive, OPC/Negative, non-OPC/Any HPV),
- Primary tumor site (Oral cavity, Oropharynx, Larynx, Hypopharnyx)
- Smoking history (>10, ≤10 pack-years)
- Race (Asian, non-Asian)
- ECOG Performance status (0,1)
- Extent of Disease (only locoregionally recurrent, metastatic with or without locoregional recurrence)

For these subgroup analyses any patient with missing values will be excluded from that particular subgroup.

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. Hazard ratio with two-sided 95% confidence interval will be based on the profile likelihood approach

A forest plot will be presented for the subgroups showing the comparison of MEDI4736+ tremelimumab combination therapy versus SoC, MEDI4736 monotherapy versus SoC and MEDI4736+ tremelimumab combination therapy versus MEDI4736 monotherapy for the analysis populations described above.

The subgroup analyses for the stratification factors will be based on the values recorded on the eCRF as indicated above. 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 the 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 comparison (it is not considered appropriate to present analyses where there are less than 20 events within a subgroup category (i.e., when the events in the treatment comparison does not add up to 20) in a subgroup), the relationship between that subgroup and the primary endpoint (OS) will not be formally analyzed. In this case, only descriptive summaries will be provided.

### 4.2.3 Analysis of the secondary variable (PFS)

The analysis of PFS will be based on the programmatically derived RECIST 1.1 using the site investigator assessments. The analysis will be performed using a stratified log-rank test stratified by PD-L1 status (positive versus negative), tumor location (OPC versus non-OPC, with a subsequent adjustment for HPV status in patients with OPC), and smoking history (>10 versus ≤10 pack-years) for generation of the p-value.

The stratification factors will be based on the values entered into IVRS at randomization even if it is subsequently discovered that these values were incorrect.

The effect of MEDI4736 monotherapy therapy versus SoC treatment in the relevant analysis populations (Table 12), will be estimated by the HR together with its corresponding 95% CI. The hazard ratio (HR) of PFS, along with its 95% confidence interval using the same strata information as above will be estimated using a stratified Cox proportional hazards model with treatment a as the only covariate (with ties= Efron and any stratification variables included in the strata statement) and the CI calculated using a profile likelihood approach

PFS analysis will also be performed to compare MEDI4736 + tremelimumab combination therapy versus SoC as well as to compare MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy for the analysis populations detailed in Table 12.

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

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.

**Sensitivity analyses for PFS variable:**

A sensitivity analysis of PFS using the BICR tumor data will be performed.

An sensitivity analysis of PFS using site investigator assessment based on RECIST 1.1 modified for confirmation of progression will be performed. The stratified log-rank test used for the secondary analysis of PFS will be repeated.

Furthermore, disagreements between the investigator and BICR assessment will also be summarised. Since the data cut-off for the last BICR is Jan 8$^{th}$, 2018, this data cut-off will also be used for the investigator assessment in this table to allow for a fair comparison.

**Subgroup analysis for PFS**

Subgroup analyses will be conducted comparing PFS between MEDI4736 ± tremelimumab combination therapy versus SoC. This subgroup analysis will be performed using the same subgroups and methodology from the OS subgroup analyses described in Section 4.2.2.

### 4.2.4 Objective response rate

The ORR will be based on the programmatically derived RECIST 1.1 using site investigator tumor data. The ORR will be compared between MEDI4736 monotherapy versus SoC using logistic regression models adjusting for the same factors as the primary endpoint (PD-L1 status, tumor location, and smoking history). The results of the analysis will be presented in terms of an odds ratio together with its associated profile likelihood 95% CI (e.g. using the option 'LRCI' in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model).

A similar analysis will be performed for the comparison between MEDI4736 + tremelimumab combination versus SoC and between MEDI4736 + tremelimumab combination versus MEDI4736 monotherapy in for the analysis populations detailed in Table 12.

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).

**Subgroup analysis for ORR**

The analysis using logistic regression to compare ORR between MEDI4736 ± tremelimumab combination therapy versus SoC will also be presented by subgroup. This subgroup analysis will be performed using the same subgroups from the OS subgroup analysis described in Section 4.2.2.

#### 4.2.4.1 Best Objective Response (BoR)

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

### 4.2.5 Duration of response

Descriptive data will be provided for the DoR in responding patients (i.e. median duration of response and 95% CIs) by treatment arm, including the associated Kaplan-Meier curves (without any formal comparison of treatment arms or p-value attached). Swimmer plots will be produced. This depicts each patient's nature and duration of response as a separate bar (horizontally) over time, ordered from the longest time in study to the shortest time in study.

### 4.2.6 Time to response

The TTR, based upon the site investigator assessment of RECIST 1.1, will be summarised (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.

As a sensitivity analysis, time to response will also be summarised based upon the site investigator tumor assessment data according to RECIST 1.1.

This analysis will be performed on all-comers and the PD-L1 TC/IC subgroup population

### 4.2.7 Proportion of patients alive and progression free at 6 and 12 months (APF12)

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

### 4.2.8 Time from randomization to PFS2

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

### 4.2.9 Time from randomization to first and second subsequent therapy or death

For supportive purposes, the time to the start of first and the second 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.

### 4.2.10 Analysis of OS12, OS18 & OS24

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

### 4.2.11 Change in tumor size

Tumor size will 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. 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.

Additionally, 'spider' plots will be produced. This depicts each patient's percentage change in tumor size at each visit as a line over time.

The above outputs will be programmed for the site investigator data based upon RECIST 1.1 assessments.

### 4.2.12 Patient reported outcome

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 for the comparison of MEDI4736 versus SoC.

The other time to symptom deterioration endpoints will be tested at a 5% significance level and 95% CIs will be produced. Compliance rates summarizing questionnaire completion at each visit will be tabulated.

If missing data is substantial, multiple imputation approach and pattern-mixture model may be explored where appropriate.

#### 4.2.12.1 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 as described for the PFS and OS analysis. The effect of MEDI4736 monotherapy versus SoC and MEDI4736 + tremelimumab combination therapy versus SoC treatment will be estimated by the HR together with its corresponding 95% CI using stratified Cox model with the same approach as described for PFS and OS analysis.

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 for each of the 3 symptom scales (fatigue, pain, and nausea/vomiting) and HRQoL/function

improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/QoL will be analyzed by comparing between treatment arms using a logistic regression model as described for the analysis of ORR.

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 (dyspnea, insomnia, appetite loss, constipation, and diarrhea), 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 provided at least one treatment arm has 20 or more subjects with data at a given visit. Graphical presentations may also be produced as appropriate.

Analysis of HRQoL or symptom will also be carried out by comparing mean change from baseline in the global health status/QoL, functions (physical, role, cognitive, social and emotional) scores (from the EORTC QLQ-C30 questionnaire) between MEDI4736 + tremelimumab combination therapy versus SoC and MEDI4736 monotherapy compared versus SoC. The analysis population for mean change in HRQoL or symptoms data will be the FAS (ITT) set and will include all randomised patients with an evaluable baseline assessment and at least one evaluable post baseline assessment.

Change from baseline will be derived using a MMRM analysis of all the post-baseline scores for each visit. The model will include treatment, visit, and treatment by visit interaction as explanatory variables and the baseline score and baseline score by visit interaction as covariates. Adjusted mean estimates per treatment group and corresponding 95% CIs will be presented along with an overall estimate of the treatment difference, 95% CI and p-value. An overall adjusted mean estimate will be derived that will estimate the average treatment effect over visits giving each visit equal weight using least squares means.

An unstructured matrix for the within-subject error variance-covariance will be used. If the model fit fails to converge, more parsimonious covariance structures will be considered to model the correlation between time-points from same patient. The following covariance structures will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, autoregressive and compound symmetry. If either the autoregressive with heterogeneity or autoregressive structure is used, patient will be included as random effect.

The Kenward-Roger approximation will be used to estimate denominator degrees of freedom (DDFM). Restricted maximum likelihood (REML) estimation will be used.

The analysis will be performed on all post-baseline visits up to the latest scheduled visit where 20 or more (≥20) patients on each treatment arm have a score.

Summary measures of overall compliance and compliance over time will be derived for the EORTC QLQ-C30. These will be based upon the compliance derivation described in Section 3.4.1

#### 4.2.12.2 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 as described for the primary analysis of OS. The effect of MEDI4736 monotherapy vs SoC and MEDI4736 + tremelimumab combination therapy versus SoC treatment will be estimated by the HR together with its corresponding 95% CI using stratified Cox model with the same approach as described for PFS and OS analysis.

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 for each of the 4 symptom scales/items (pain, swallowing, senses and speech) will be compared between treatment groups using a logistic regression model as described for ORR.

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) and11 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 provided at least one treatment arm has 20 or more subjects with data at a given visit . Graphical presentations may also be produced as appropriate.

The assessment of symptoms comparing mean change from baseline using the MMRM as described for EORTC QLQ-C30 will be repeated for pain, swallowing, senses, and speech of the EORTC QLQ-H&N35. All assumptions and outputs as described for the EORTC QLQ-C30 are applicable and the same approach from EORTC QLQ-C30 will be applied.

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

#### 4.2.12.3 CCI [redacted]

CCI [redacted]

CCI [redacted]

#### 4.2.12.4 CCI [redacted]

CCI [redacted]

#### 4.2.12.5 CCI [redacted]

CCI [redacted]

### 4.2.13 Safety

Safety data will be summarized. No formal statistical analyses are planned for safety data. All safety and tolerability data will be presented for all-comers and also for those in the PD-L1 TC/IC subgroup using the safety population. The following sections describe the planned safety summaries for AEs, vital signs, laboratory parameters, ECG and WHO/ECOG performance status. However, additional safety tables (not specified in this SAP) may need to be produced to aid interpretation of the safety data. This may include repeating the planned safety summaries for the low risk of EM subgroup. Any safety summaries examining retreatment with MEDI4736 + tremelimumab will be produced separately.

#### 4.2.13.1 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 and which did not worsen during the course of the study 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 worsens after Study Day 1 will be regarded as treatment emergent and thus will be included in the summary tables. Note: If an AE is not worse than the baseline (pre-dose) severity then it will not be classified as TEAE.

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) regardless of whether the AE occurred during re-treatment with tremelimumab or not 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 longer 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 summary will also be produced containing all AEs (by system organ class and preferred term) observed from the initiation of the first subsequent therapy following discontinuation of IP until 90 days following discontinuation of IP treatment (i.e. summarizing those AEs experienced by patients taking subsequent therapy during the 90 days AE collection follow-up window post discontinuation of IP). 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 any episode level summaries which may be produced).

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

- SAEs leading to discontinuation of study medication
- SAEs 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.
- Infusion reaction AEs

An overall summary of the number and percentage of patients in each category will be presented. In addition, a truncated AE table of most common AEs, showing all events that occur in at least 5% of patients in any treatment group will be summarized by preferred term, by decreasing frequency. This cut-off may be modified after review of the data. When applying a cut-off (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 100 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 drug exposure of patients and then multiplied by 365.25 x 100 to present in terms of 100 patient years) will be presented. For the MEDI4736 + tremelimumab arm, event rates will be based upon total combination duration including any tremelimumab re-treatment period (i.e. the very last dose date of study drug (MEDI4736 or tremelimumab) regardless of it being after tremelimumab re-started should be used as last dose in the calculation)

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.

For each AE, time to first onset of the AE from the date of first dose will be presented in patients in the safety analysis.

A summary of 3 deaths tables

1. All deaths (Full analysis set)
2. Deaths within 90 days of last dose (Safety analysis set)
3. Death within 90 days of last dose or prior to start of subsequent therapy (Safety analysis set)

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

- Related to disease under investigation,
- AE with 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 and possible interest**

Preferred terms used to identify adverse events of special interest (AESIs) and adverse events of possible interest (AEPIs) 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 and may also show the individual preferred terms which constitute each AESI/AEPI grouping. 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 4. 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 AESI/AEPI presented by outcome
- At least one AESI/AEPI causally related to study medication (as determined by the reporting investigator)
- At least one AESI/AEPI leading to discontinuation of IP

A summary of time to resolution to grade 1 or less and time to resolution to grade 2 or less will be provided.

Additionally, there will be several summaries of AESIs/AEPIs requiring concomitant treatment, and particularly the relationship of AESIs/AEPIs to the use of immunosuppressive agents (i.e., depicting which AESI/AEPI 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).

**Immune-mediated Adverse events (imAEs)**

The imAEs (as classified by the Sponsor) will be summarized in the similar manner as for the summaries for AESI/AEPI described above. The Sponsor will be responsible for producing these summaries.

**Infection AEs**

These AEs will be summarized by pooled terms and PTs in two ways: (1) using MedDRA HLGT/HLT pooled terms (2) Custom pooled terms. The list of terms will be provided in a separate excel file. The following summaries will be reported for both HLGT/HLT pooled terms and custom pooled terms and PTs:

- Infection Adverse Events by pooled term and PT
- Causally-related Infection AEs by pooled term and PT
- Serious Infection AEs by pooled term and PT
- Causally-related Serious Infection AEs by pooled term and PT
- Infection AEs with maximum CTCAE grade by pooled term and PT
- Infection AEs with outcome of death by pooled term and PT
- Infection Adverse Events leading to treatment discontinuation by pooled term and PT
- Infection Adverse Events leading to dose delay or interruption by pooled term and preferred term
- Infection Adverse Events -List of preferred terms

**Haemorrhages-related adverse events**

Treatment-emergent haemorrhage events will be identified using the Haemorrhage (excluding laboratory terms) Standardized Medical Dictionary for Regulatory Activities (MedDRA) Query (SMQ) from the current MedDRA version.

The following key summary tables and listings will be produced for the subset of patients who are reported to have a treatment-emergent adverse event within the Hemorrhage SMQ:

- Haemorrhage SMQ patient disposition (all patients)
- Haemorrhage SMQ Demographic characteristics (Safety Analysis Set)
- Haemorrhage SMQ Disease characteristics at initial diagnosis or study entry (Safety Analysis Set)
- Haemorrhage SMQ Extent of disease at study entry (Safety Analysis Set)
- Haemorrhage SMQ Adverse events in any category - patient level (Safety analysis set)
- Haemorrhage SMQ Adverse events and event rate by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Time to onset of adverse event (days), arranged by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Causally related adverse events by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Adverse events by system organ class, preferred term, and maximum reported CTCAE grade (Safety analysis set)
- Haemorrhage SMQ Adverse events of CTCAE grade 3 or 4 by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Causally related adverse events of CTCAE grade 3 or 4 by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ All deaths (Safety Analysis Set)

- Adverse events with outcome of death, by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Adverse events with outcome of death, causally related to study treatment, by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Serious adverse events by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Serious adverse events, causally related to study treatment, by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Adverse events, leading to discontinuation of study treatment, by system organ class and preferred term (Safety analysis set)
- Haemorrhage SMQ Listing of Deaths (Safety Analysis Set)
- Haemorrhage SMQ Serious adverse events - Listing of key information (Safety analysis set)
- Haemorrhage SMQ Impact of concomitant medication with potential risk of bleeding on the occurrence of the most severe bleed within each patient (Safety analysis set)
- Haemorrhage SMQ Medical history of bleeding (Safety analysis set)

**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, that is 10 events in at least one treatment group.

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, that is 10 events in at least one treatment group.

For each AESI, median time to first onset of the AESI from the date of first dose will be presented in patients in the safety analysis. Patients who did not experience the AESI 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 and AESI will be presented in patients who experienced each, 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, that is 10 events in at least one treatment group.

A cumulative incidence plot of the raw cumulative incidence over time will be presented. 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. These plots will only be produced for AESIs that have ≥10 events, that is 10 events in at least one treatment group.

#### 4.2.13.2 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.

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 CTCAE grading applies.

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

Boxplots of absolute values by week, and boxplots of change from baseline by week, may be presented for certain parameters if warranted after data review up to Week 40.

Shift tables for laboratory values by worst common toxicity criteria (CTCAE) 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 CTCAE grade shift outputs will be produced are:

- Hematology: Hemoglobin, Leukocytes, Lymphocytes, absolute count, Neutrophils, absolute count, Platelets

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

For the parameters with no CTCAE grading that are listed in the CSP, shift tables from baseline to worst value on treatment will be provided. Additional summaries will include a

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

**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 ULN Upper Limit of Normal (ULN) during the study

  > 5x- ≤8x ULN

  > 8x -≤ 10x ULN

  >10x -≤20 ULN

  and > 20x ULN

  AST

  ≥ 3x –≤5x ULN

  > 5x -≤8x ULN

  > 8x -≤ 10x ULN

  >10-≤ 20 ULN

  and > 20x ULN

  Total bilirubin

  ≥2x -≤3x ULN,

  >3x-≤5x ULN,

  >5x ULN during the study

  ALT or AST

  ≥3x - ≤5x ULN,

  >5x - ≤8x ULN,

>8x-≤10x ULN,

>10 -≤ 20 ULN

and >20x ULN during the study

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

- Narratives will be provided in the CSR for patients who have ALT ≥ 3x ULN plus Total bilirubin ≥ 2x ULN or AST ≥ 3x ULN plus Total bilirubin ≥ 2x ULN at any visit.

Liver biochemistry test results over time for patients with elevated ALT (i.e. ≥ 3x ULN) or AST (i.e. ≥ 3x ULN), and elevated total bilirubin (i.e. ≥ 2x ULN) (at any time) will be plotted. Individual patient data where ALT or AST 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.

#### 4.2.13.3 ECGs

ECG data obtained up until the 90-day safety follow-up visit will be included in the summary tables.

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".

#### 4.2.13.4 Vital signs

Boxplots for absolute values and change from baseline by week may be presented for certain vital signs parameters if warranted after data review, up to Week 40.

### 4.2.14 WHO/ECOG performance status

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

### 4.2.15 PK Data

MEDI4736/tremelimumab concentration data will be listed for each patient and each dosing day, and a summary provided for all patients in the PK analysis set.

If the data are suitable, the relationship between PK exposure and efficacy/safety parameters may be investigated graphically or using an appropriate data modeling approach.

These outputs will be produced by a designated third party on behalf of AstraZeneca.

### 4.2.16 Immunogenicity analysis

A summary of the number and percentage of patients who develop detectable ADA to MEDI4736 or tremelimumab by ADA categories (Section 3.5.3) in MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy arms will be presented. The denominator for percentage calculations will use the number of ADA-evaluable patients. Immunogenicity results will be listed for all patients in the Safety Analysis Set regardless of ADA-evaluable status. ADA titer and nAb data may be presented for samples confirmed positive for the presence of ADA to durvalumab and/or tremelimumab. AEs in ADA positive patients by ADA positive category (ADA-positive at any visit, treatment-emergent ADA-positive, and nAb positive) may be listed.

The effect of immunogenicity as well as the effect of its neutralizing properties on PK, pharmacodynamics, efficacy, and safety will be evaluated, if the data allow.

### 4.2.17 Biomarker data

CCI [redacted]

PD-L1 tumor cell expression (used for purposes of stratification) and PD-L1 tumor and tumor-associated immune cell expression (used for purposes of identifying patients for the PD-L1 TC/IC-selected subgroup) determined by IHC will be reported in the CSR. CCI [redacted]

### 4.2.18 Demographic, initial diagnosis and baseline characteristics data

The following will be summarized for all patients in the FAS and PD-L1 TC/IC high subgroup analysis set (unless otherwise specified):

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations specified in this SAP
- COVID-19 related non-important PDs
- Inclusion in analysis populations

- Demographics (age, age group [< 65, ≥65<75, ≥ 75 years], sex, race and ethnicity)
- Patient characteristics at baseline (weight, weight group, BMI and BMI group)
- Patient recruitment by country and centre
- Previous disease-related treatment modalities
- Disease characteristics at baseline (WHO/ECOG performance status, best response to previous therapy, overall disease classification)
- Disease characteristics at initial diagnosis (primary tumor location, histology type, tumor grade, TNM classification, AJCC staging)
- Extent of disease at baseline (recurrent or metastatic)
- Disease related medical history
- Time from most recent disease progression to randomization
- Post-discontinuation cancer therapy
- Smoking status (>10, ≤10 pack-years)
- Nicotine use, categorized (current, former, never)
- Alcohol use (current, former, never)
- PD-L1 TC status for purposes of stratification (positive, negative) for all patients treated.
- Human papilloma virus (HPV) status（positive or negative, for patients with oropharyngeal cancer only)
- Stratification factors by IVRS and CRF
- Discrepancy between local and central review
- Concomitant medication during study treatment

Additional Demographic data will be summarized for the TMB high and low risk of EM subgroups.

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

- COVID-19 related non-important PDs
- Subject excluded from analysis populations
- Demographics (age, age group [< 65, >=65-< 75, ≥ 75 years], sex, race and ethnicity)

### 4.2.19 Treatment exposure

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

- Total exposure and a plot of exposure over time (See Section 3.3.2)
- Actual exposure. (Actual exposure will not be provided for SoC patients)
- Number of dose interruptions/delays and reasons for dose interruptions/delays (drug interruptions represent interruption of an infusion and therefore any interruptions recorded on dates when study treatment was not administered will not be counted, delays are defined as a dose delay that occurs between 2 administrations of study treatment)
- RDI (entire intended treatment period).
- Cumulative exposure over time (excludes 28-day window)

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.

### 4.2.20 Subsequent Therapy

Subsequent therapies received after discontinuation of IP will have summaries produced, together with number of regimens received. Moreover, a descriptive summary will be produced for time to first and second subsequent therapy from randomization.

## 5. INTERIM ANALYSES

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

No interim analyses will be performed for any efficacy endpoint.

# 6. CHANGES OF ANALYSIS FROM PROTOCOL

**Table 13 Changes of Analysis from Protocol**

| SAP edition# | Section of SAP Affected (If applicable) | Change | Rationale |
|---|---|---|---|
| 6 | CCI [redacted] | | |

# 7. REFERENCES

**Argiris et al 2013**
Argiris A Ghebremichael M, Gilbert J, Lee JW, Sachidanandam K,Kolesar JM,Burtness B,Forastiere AA.Phase III randomized, placebo-controlled trial of docetaxel with or without gefitinib in recurrent or metastatic head and neck cancer: An Eastern Cooperative Oncology Group trial J Clin Oncol 31: 1405– 1414,2013

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

**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.

**Brahmer et al 2012**
Brahmer JR, Tykodi SS, Chow LQM, Hwu WJ, Topalian SL, Hwu P, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med 2012;366(26):2455-65

**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

**Burtness et al., 2019**
Burtness B, Harrington KJ, Greil R, Soulières D, Tahara M, de Castro Jr G et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for

recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. Lancet 2019; 394: 1915–28. Published Online October 31, 2019 https://doi.org/10.1016/ S0140-6736(19)32591-7

**Cohen et al 2019**
Cohen EEW, Soulières D, Le Tourneau C, Dinis J, Licitra L, Ahn MJ, Soria A et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. Lancet 2019; 393:156.

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

**Duke-Margolis 2018**
Oncology Clinical Trials in the Presence of Non□Proportional Hazards, The Duke□Margolis Center for Health Policy Feb. 2018.

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

**EMEA Guideline 2005**
European Medicines Agency. Guideline on the choice of the non-inferiority margin. 27 July 2005. Doc.Ref: EMEA/CPMP/EWP/2158/99.

**ESMO Asia 2018**
Phase III Trial to Assess Efficacy and Safety of Cetuximab for the Treatment of Chinese Participants With Head and Neck Cancer. NCT02383966

**Fayers et al 1999**
Fayers P, Aaronson NK, Bjordal K, Curran D, Groenvold M on behalf of the EORTC Quality of Life Study Group. EORTC QLQ-C30 Scoring Manual: 2nd Edition 1999. Available on request.

**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.

**FDA Guidance 2016**
Non-Inferiority Clinical Trials to Establish Effectiveness. Available from:URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials

**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.

**Grambsch and Therneau 1994**
Grambsch, PM. and Therneau TM. 1994. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 81:515–26

**Grau et al 2009**
Grau JJ, Caballero M, Verger E, Monzo M, Blanch JL. Weekly paclitaxel for platin-resistant stage IV head and neck cancer patients. Acta Otolaryngol, 129 (11) (2009), pp. 1294-1299

**Guigay et al 2019**
Guigay J, Fayette, J, Mesia R, Lafond C, Saada-Bouzid E, Geoffrois L, et al. TPExtreme randomized trial: TPEx versus Extreme regimen in 1st line recurrent/metastatic head and neck squamous cell carcinoma (R/M HNSCC). J Clin Oncol 37, 2019 (suppl; abstr 6002)

**Guo et al 2015**
Guo Y, Shi M, Yang A, Feng JF,Zhu XD, Choi YJ et al. Platinum□based chemotherapy plus cetuximab first□line for Asian patients with recurrent and/or metastatic squamous cell carcinoma of the head and neck: Results of an open-label, single-arm, multicenter trial. Head Neck. 2015; 37: 1081-1087.

**Karrison 2016**
Karrison TG. Versatile tests for comparing survival curves based on weighted log-rank statistics. Stata J 2016; 16: 678–90.

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

**Latimer et al 2018**
Latimer NR, Abrams KR, Lambert PC, Morden JP, Crowther MJ. Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study. Stat Methods Med Res. 2018;27(3):765–84.

**Li et al 2020**
Weimin Li, Sophie Wildsmith, Jiabu Ye, Han Si, Nassim Morsli, Philip He et al. Plasma-based tumor mutational burden (bTMB) as predictor for survival in phase III EAGLE study: Durvalumab (D) ± tremelimumab (T) versus chemotherapy (CT) in recurrent/metastatic head and neck squamous cell carcinoma (R/M HNSCC) after platinum failure. Journal of Clinical Oncology 2020 38:15_suppl, 6511-6511

**Lin et al 2020**
Lin RS, Lin J, Roychoudhury S, et al. Alternative analysis methods for time to event endpoints under nonproportional hazards: a comparative analysis. Stat Biopharm Res 2020; 12: 187–98.

**Klein et al 2007**
Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. Stat Med 2007 Oct 30; 26(24): 4505–19.

**Machiels et al 2011**
Machiels JP, Subramanian S, Ruzsa A , Repassy G,Lifirenko I, Flygare A,et al. Zalutumumab plus best supportive care versus best supportive care alone in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck after failure of platinum-based chemotherapy: an open-label, randomised phase 3 trial. Lancet Oncol, 12 (2011), pp. 333-343

**Machiels et al 2015**
Machiels JP, Haddad RI, Fayette J, Licitra LF, Tahar M, Vermorken JB, et al. Afatinib versus methotrexate as second-line treatment in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck progressing on or after platinum-based therapy (LUX-Head & Neck 1): an open-label, randomised phase 3 trial. Lancet Oncol 2015;16:583-594

**Nishino et al 2013**
Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. Clin Cancer Res 2013;19(14):3936-43.

**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.**
Competing risks: A practical perspective. Wiley.

**Rischin et al 2019**
Rischin D, Harrinton KJ,Greil R,Soulieres D,Tahara M,De Castro G, Psyrri A, et al. Protocol-specified final analysis of the phase 3 keynote-048 trial of Pembrolizumab (PEMBRO) as first-line therapy for recurrent/metastatic head and neck squamous cell carcinoma(R/M HNSCC). J Clin Oncol 37,2019 (suppl;Abstr 6000)

**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.

**Segal et al 2015**
Segal H. Ou S-HI, Balmanoukian AS, Fury MG, Massarelli E, Brahmer JR,. Safety and efficacy of MEDI4736, an anti-PD-L1 antibody, in patients from a squamous cell carcinoma of the head and neck (SCCHN) expansion cohort [abstract 3011]. Poster presented at the American Society of Clinical Oncology (ASCO) Annual Meeting, Chicago, IL, 29 May to 2 Jun 2015

**Soulieres et al 2017**
Soulieres D, Faivre S, Mesia R, Remenar E, Li SH, Karpenko A, et al. Buparlisib and paclitaxel in patients with platinum-pretreated recurrent or metastatic squamous cell carcinoma of the head and neck (BERIL-1): a randomised, double-blind, placebo-controlled phase 2 trial. Lancet Oncol, 18 (2017), pp. 323-335

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

**Stone et al 2015**
Stone A, Macpherson E, Smith A, Jennison C. Model free audit methodology for bias evaluation of tumor progression in oncology submitted

**Sun and Chen 2010**
Sun X, Chen C. Comparison of Finkelstein's Method with the Conventional Approach for Interval-Censored Data Analysis. Stat Biopharm Res 2010;2(1):97-108.

**Topalian et al 2012**
Topalian, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012;366(26):2443-54

**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.

**Weber et al 2012**
Weber JS, Kähler KC, Hauschild A. Management of immune-related adverse events and kinetics of response with ipilimumab. J Clin Oncol 2012; 30:2691-7.

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

# 8. APPENDIX

## 8.1 Early risk of mortality prognostic model

CCI

[redacted]

# SIGNATURE PAGE

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

| **Document Name:** d419lc00001-sap-ed-7 | | |
|---|---|---|
| **Document Title:** | Statistical Analysis Plan Edition 7 | |
| **Document ID:** | Doc ID-004446239 | |
| **Version Label:** | 1.0 CURRENT LATEST APPROVED | |
| **Server Date** **(dd-MMM-yyyy HH:mm 'UTC'Z)** | **Signed by** | **Meaning of Signature** |
| 18-Dec-2020 15:00 UTC | PPD | Content Approval |
| 18-Dec-2020 08:14 UTC | PPD | Author Approval |

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