B9991016

A RANDOMIZED DOUBLE-BLIND PHASE 3 STUDY OF AVELUMAB IN COMBINATION WITH STANDARD OF CARE CHEMORADIOTHERAPY (CISPLATIN PLUS DEFINITIVE RADIATION THERAPY) VERSUS STANDARD OF CARE CHEMORADIOTHERAPY IN THE FRONT-LINE TREATMENT OF PATIENTS WITH LOCALLY ADVANCED SQUAMOUS CELL CARCINOMA OF THE HEAD AND NECK

JAVELIN HEAD AND NECK 100

STATISTICAL ANALYSIS PLAN – B9991016

Compounds:	MSB0010718C PF-06834635
Compound Name:	Avelumab
Version:	3.0
Date:	30-Aug-2019

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B9991016 is based on the protocol amendment 3 dated 2-Aug-2019.

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
3	30-Aug-2019	Updated the target number of events for PFS and OS and associated operating characteristics as per protocol amendment 3 in Section 2 "Introduction", Section 5.1 "Hypothesis and Decision Rules" and subsections and Section 7.2.2 "Interim analysis for OS". CCI [redacted] Section 3.4.1 "Study drug, study treatment and baseline definitions" – the definition of start dates of study treatment within each treatment phase was added to support the summaries of safety by treatment phase. Section 6.2.2.2 "Sensitivity analyses for progression-free survival" – per FDA request, sensitivity analyses of PFS by Blinded Independent Central Review (BICR) per standard RECIST v1.1 were added; early discrepancy rate and late discrepancy rate interpretations were corrected. Section 6.2.2.4 "Locoregional failure" – included an additional possible censoring reason associated with Distant Metastatic failure in Table 17 "Time to LRF Censoring Reasons and Hierarchy". Section 6.2.2.5 "Distant metastatic failure" – included an additional possible censoring reason associated with Locoregional failure in Table 19 "Time to DM Censoring Reasons and Hierarchy". Section 6.2.2.6 "Neck dissection" – the positive and negative predictive values definitions were corrected to refer to individual lesions rather than to patients as patients may have multiple lesions. Section 6.2.2.7 "Pathologic complete response" – pathologic complete responder definition was clarified for patients with more than one resected lesion. Section 6.2.6 "Biomarker endpoints" – biomarker analysis was simplified to focus on the analyses associated with the pre-specified objectives. Section 6.2.7.3"Analysis of PK and safety by immunogenicity status" – clarified that the analysis of IRRs by ADA status will be conducted for IRRs associated with avelumab. Section 6.5.1.3 "Disease characteristics" - "Time since diagnosis of local/regional recurrence of disease" – deleted as data not collected in the CRF. The smoking history analysis was simplified. Section 6.5.2.1 "Patient disposition" – summaries for CRT Phase were added; frequency of patients completing each study phase was added; frequency of patients entering CRT Phase and frequency of patients entering Maintenance Phase were added. Section 6.5.3.1 "Exposure to avelumab or matching placebo" – duration of exposure to avelumab was changed from days to weeks; duration of exposure and total number of infusions to avelumab across all phases was added. Section 6.5.3.2 "Exposure to cisplatin" – considerations regarding split dose were added. Section 6.5.3.3 "Exposure to IMRT" – analysis of compliance was added. Section 6.6.1 "Adverse events" – by treatment phase summaries of TEAEs, Serious TEAEs, TEAEs leading to death, Deaths, irAEs, and IRRs were added;

		clarified in the IRR summaries that IRRs can be associated with avelumab or cisplatin and which IRR summaries will be performed for IRRs associated with avelumab, associated with cisplatin or associated with either avelumab or cisplatin; summary of IRRs leading to discontinuation of IMRT was deleted as not IMRT is not an infusion; clarified that summaries associated with discontinuation of any study drug or all study drugs is applicable to the CRT Phase only; deleted the summary of TEAEs excluding SAEs. Section 6.6.5 "Laboratory Data" – by treatment phase summaries of hematology and chemistry parameters by CTCAE grade were added; listing of liver function test elevations modified to include only patients to be reviewed for assessment of potential Hy's law. Section 6.6.7 "Electrocardiogram" – per protocol, only approximately 30 patients are expected to have scheduled on-treatment ECGs and the study-specific QT correction for heart rate (QTcP) has been deleted from the SAP. Minor editorial and consistency changes.
2	30-Apr-2019	CCI [redacted] Section 2.2 "Study Design' updated as per protocol amendment 1. Section 3.2.6 "Biomarker endpoints" - Method of determination was changed from Daco pharmDx kit to VENTANA PD-L1 (SP263) assay Section 3.4.1 "Study drug, study treatment and baseline definitions" - definition of baseline for PRO analyses was clarified. Definition of baseline for immunogenicity analyses was added. Section 3.5 "Adverse Events" - the definition of Treatment-emergent adverse events (TEAEs) was modified to include all AEs with onset date during the on-treatment period. Section 4 "Analysis Sets" - the per protocol analysis of the secondary efficacy endpoint of OS was added to Table 5. Section 5.2.5 "Definition of start of new anti-cancer drug therapy" – randomization date was replaced by date of first dose of study treatment since used in the definition of on-treatment period. Section 5.2.6 "Definition of start of new anti-cancer therapy" - first dose of study treatment was replaced by date of randomization. Section 5.2.8 "Standard derivations and reporting conventions" - formula for BSA was added. Section 5.2.10 "Adequate baseline tumor assessment" - the criteria were updated to include only unequivocal criteria. Section 5.3.3.1 "Date of last contact" – added withdrawal of consent date in the derivation of date of last contact. Section 5.3.3.2 "Death date" imputations – deleted the last step in the imputation "if the day is missing from the date of last contact it will be imputed to 1st day of the month and year of last contact only if derived from the 'Survival Follow-up' eCRF page" since not needed. Section 5.3.3.4 "Date of start of new anti-cancer therapy" was added to provide details of imputation rules for incomplete dates for start date of new anti-cancer therapy (drug therapy, radiation, surgery). Section 6.1.1.1 "Primary analysis" – added the details of the derivation f progressive disease per modified RECIST as outlined in Appendix 3 of the B9991016 protocol; added that RCIs will be calculated for the analyses of PFS; provided more details on analysis of Time of Follow-Up for PFS.

		Section 6.2.2.1 “Overall survival” - added that RCIs will be calculated for the analyses of OS; provided more details on analysis of Time of Follow-Up for OS. Section 6.2.2.2 “Sensitivity analyses for progression-free survival” - the plot of RMST versus tau was deleted as detailed summaries for the 3 different tau values will be provided. PD was appropriately replaced with PFS in the analyses of concordance/discordance for PFS Section 6.2.2.9 “Duration of response” – added details regarding censoring for duration of response since “no adequate baseline assessment” which is used in censoring for PFS analyses is not applicable to analyses of duration of response for patients with objective response. Section 6.2.4 “Pharmacokinetic endpoint” – additional parameters included, if data permit. Section 6.2.7 “Endpoints for immunogenicity data of avelumab” - added details for the analysis of ADA and nAb. Section 6.4 “Subset Analyses” – pooled geographical region updated to consider Western Europe and Eastern Europe separately (instead of Europe combined); updated the p-value for the interaction test to be based on Wald’s test; removed forest plots for median DR by subgroup. Section 6.5.1.1 “Demographic characteristics” - pooled geographical region updated to consider Western Europe and Eastern Europe separately (instead of Europe combined); BSA added; randomization stratification factors and subsite of primary diagnosis removed since described in Section 6.4. Section 6.5.3 “Study treatment compliance and exposure” and subsections - by-cycle and week-based analyses of exposure were removed; formulas for actual dose intensity for avelumab and cisplatin were corrected to refer to intended duration of exposure; summaries of dose omissions were removed since the non-adherence to the treatment schedule will be described with dose delays and interruptions; added details for derivation and summary of dose delays. Section 6.6.1 “Adverse events” and subsections – updated definition of treatment-emergent adverse event. Given the nature of immune-related AEs (irAEs) and infusion related reactions (IRRs), removed the summaries of treatment-related irAEs and treatment-related IRRs; removed the requirement that all summaries by SOC and PT will be replicated by PT only. Added summaries of AEs leading to dose reduction, AEs leading to interruption of study treatment and AEs leading to both dose reduction and interruption of study treatment. Section 6.6.1.2 “Adverse events leading to dose reduction” and Section 6.6.1.3 “Adverse events leading to interruption of study treatment” were added. Section 6.6.2 “Death” – deleted references to primary reason for death since not collected in the eCRF. Section 6.6.4 “Other significant adverse events” – nominal p-values for risk difference for Tier-1 events will not be calculated. Section 6.6.5 “Laboratory data” and subsections –added summary for patients with newly occurring or worsening laboratory abnormities; updated lab parameters that will be listed. Section 6.6.7 “Electrocardiogram” – Deleted the summary for patients with ECG abnormalities based on morphology since collected in a free text field.. Section for “Physical Examination” was removed as data not collected in the eCRF (abnormal findings are reported as AEs). Section 8 ”References” – references updated. Appendix 1 “Immune-related Adverse Events” – updated the ATC codes for concomitant medications in Step 4 of the case definition for irAEs and added a

		statement regarding medical review if needed to refine the programmatic derivation. Minor editorial and consistency changes throughout the document.
1	21-Oct-2016	Not applicable (N/A)

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in study B9991016. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

A separate SAP will cover the interim analyses for periodic safety review by the External Data Monitoring Committee (E-DMC).

Statistical analyses will be performed using cleaned eCRF data as well as non-CRF data (ie, intensity-modulated radiation therapy [IMRT] data, immunogenicity data, pharmacokinetic [PK] concentration data).

The primary analysis will include all data up to a cut-off date which is determined by the number of progression-free survival (PFS) events by modified RECIST v1.1 and minimum follow-up of 24 months (104 weeks) after the last randomized patient required for the analysis of PFS. The cut-off date is determined once a data extract (before database lock) is available which indicates that the required number of events for PFS (289) and minimum follow-up of 24 months (104 weeks) is expected to occur by the cut-off date.

Due to cleaning activities, the final number of events might deviate from the planned number. The data cut-off date will not be adjusted retrospectively in this case.

2.1. Study Objectives

Primary Objectives

- To demonstrate that treatment with avelumab in combination with standard of care chemoradiotherapy (SOC CRT) is superior to SOC CRT alone in prolonging PFS in front-line patients with high risk, locally advanced squamous cell carcinoma of the head and neck (SCCHN) who are candidates for definitive CRT with cisplatin.

Secondary Objectives

- To compare the overall survival (OS) of avelumab in combination with SOC CRT vs SOC CRT alone;
- To evaluate the anti-tumor activity of avelumab in combination with SOC CRT and SOC CRT alone;
- To evaluate the overall safety and tolerability profile of avelumab in combination with SOC CRT and SOC CRT alone;
- To evaluate the PK of avelumab;

- To evaluate the PK of cisplatin (total and free);
- To assess avelumab ADAs;
- To evaluate the effect of avelumab in combination with SOC CRT and SOC CRT alone on patient-reported outcomes (PROs) of disease-related symptoms and health related quality of life (HRQoL);
- To evaluate candidate immune-related predictive biomarkers of sensitivity or insensitivity to treatment with avelumab in combination with SOC CRT in pretreatment tumor samples (eg, PD-L1 expression);
- To evaluate candidate immune-related predictive biomarkers of sensitivity or insensitivity to treatment with avelumab in combination with SOC CRT in tumor samples (eg, PD-L1 expression) after one dose of avelumab in patients who provide this optional biopsy.

CCI [redacted]

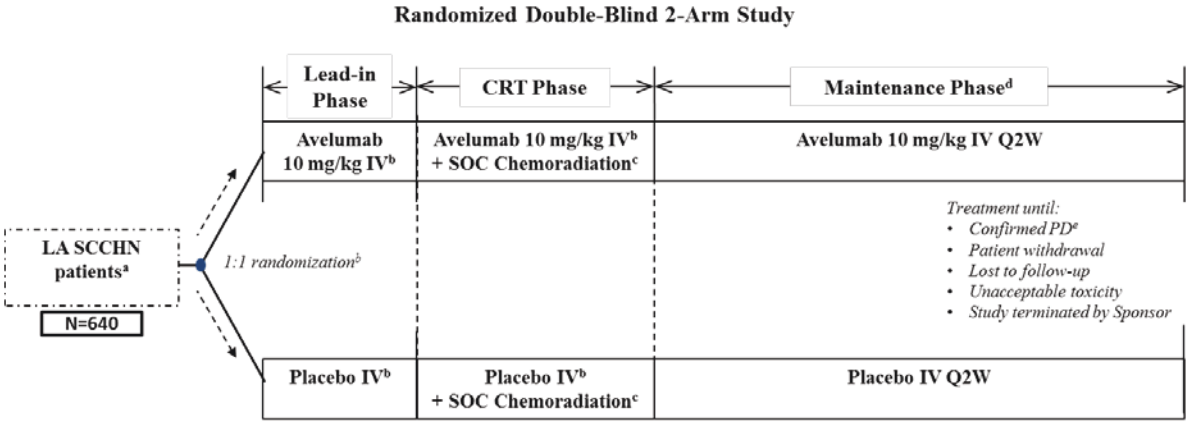
2.2. Study Design

This is a Phase 3, international, multicenter, randomized, double-blind, parallel, 2-arm study in patients with previously untreated, histologically confirmed locally advanced SCCHN (oral cavity, oropharynx, larynx, or hypopharynx) who are candidates for definitive chemoradiotherapy (CRT) with cisplatin.

A total of approximately 640 patients will be randomized in a 1:1 ratio to either Arm A (avelumab + SOC CRT) or Arm B (placebo + SOC CRT). Randomization will be stratified by tumor (T) stage (<T4 vs T4), nodal (N) stage (N0/N1/N2a/N2b vs N2c/N3), and HPV status (positive vs negative) as measured by p16 expression by immunohistochemistry (IHC).

The study schematic is provided in Figure 1.

Figure 1. Study Design Schema



a. Patients with LA SCCHN (oral cavity, oropharynx, larynx, or hypopharynx) eligible for front-line treatment: HPV negative disease stage III, IVa, IVb; non-oropharyngeal HPV positive disease stage III, IVa, IVb; HPV positive oropharyngeal disease T4 OR N2c OR N3 (staging per TNM [tumor, node, metastases] guidelines for head and neck sites AJCC 7th Edition). (Note: entry criteria for patients with HPV-positive tumors are different than for HPV-negative tumors).

b. Avelumab or placebo IV to be administered on Day 1 of the Lead-in Phase (1 week prior to the start of the CRT Phase), and on Days 8, 25, and 39 during the CRT Phase.

c. SOC CRT = IMRT (70 Gy/35 fractions/7weeks; 1 fraction/day, 5 fractions/week) for 7 weeks + cisplatin (100 mg/m^2 Days 1, 22, 43) during the CRT Phase.

d. Maintenance Phase to start after completion of the CRT Phase and continue for 12 months.

e. Patients will continue treatment until confirmed disease progression as assessed by Investigator per modified RECIST v1.1.

AJCC = American Joint Committee on Cancer; CRT = chemoradiotherapy; HPV = human papillomavirus; IMRT = intensity-modulated radiation therapy; LA = locally advanced; Q2W = every 2 weeks; RECIST = Response Evaluation Criteria in Solid Tumors; SCCHN = squamous cell carcinoma of the head and neck; SOC = standard of care.

There will be 3 treatment phases in this study.

- Lead-in Phase: On Day 1 of the Lead-in Phase of the study, patients will receive a single dose of avelumab or matching placebo, administered 7 days prior to initiation of the CRT Phase;
- CRT Phase: Avelumab or matching placebo will be administered on Days 8, 25, and 39 in conjunction with SOC CRT starting on Day 1 of the CRT Phase;
- Maintenance Phase: Following completion of the CRT Phase, avelumab or matching placebo will be administered Q2W for 12 months during the Maintenance Phase.

The treatment schedule is described in Table 2.

Table 2. Chemoradiation and Avelumab/Placebo Treatment Schedule

	Lead-in Phase (7 days)[a]	CRT Phase (63 days)								Maintenance Phase[b]
	-- Day --	---------- Day ----------								
	1	1	8	15	22	25	29	36	39	43	
IMRT (70 Gy/35 fractions/7 weeks; 1 fraction per day, 5 fractions/week [Monday – Friday])		X→	X→	X→	X	→	X→	X→	→	X→	
Cisplatin (100 mg/m^2 Q3W)		X			X					X	
Avelumab (10 mg/kg) or placebo	X		X		X[c]			X			Q2W

a. Lead-in Phase to start 7 days prior to initiation of CRT Phase.
b. Maintenance Phase to start after completion of the CRT Phase (ie, 2 weeks following completion of CRT) and be 12 months in duration. CRT phase may be extended as required for patient recovery from CRT; however, it is hypothesized that patients will receive maximal benefit from beginning maintenance as close to the scheduled time as possible.
c. Day 25 dose of avelumab/placebo during the CRT Phase may be administered between Day 24 and Day 29.

Cisplatin should be administered on a Monday or Tuesday if feasible to maximize overlap with IMRT.

CRT = Chemoradiotherapy; IMRT = intensity-modulated radiation therapy; Q2W = every 2 weeks; Q3W = every 3 weeks.

The investigator will assess antitumor activity based on radiological assessments and clinical evaluation of patients using modified RECIST v1.1 at baseline, 12 weeks after the completion of CRT (ie, 10 weeks following completion of the CRT Phase of the study), every 16 weeks thereafter for 48 months (208 weeks), and every 24 weeks thereafter until confirmed disease progression per modified RECIST v1.1 and regardless of discontinuation of study treatment or initiation of subsequent anti-cancer therapy.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

- PFS per modified RECIST v1.1 by investigator assessment.

  PFS is defined as the time from the date of randomization to the date of the first documentation of progressive disease (PD per modified RECIST v1.1) or death due to any cause, whichever occurs first.

3.2. Secondary Endpoints

3.2.1. Safety endpoints

- Adverse events (AEs) and laboratory abnormalities as graded by National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) v4.03.

AEs will be graded by the investigator according to CTCAE v4.03 and coded using the Medical Dictionary for Regulatory Activities (MedDRA).

- Vital signs (blood pressure, pulse rate).

3.2.2. Efficacy endpoints

- OS.

  OS is defined as the time from the date of randomization to the date of death due to any cause.

- Antitumor activity: Pathologic complete response in any resected specimens, neck dissection.

  Pathologic complete response is defined as the absence of histologically identifiable residual cancer in any resected specimen.

- Antitumor activity: Locoregional failure, objective response (OR), distant metastatic failure, and duration of response (DR), per modified RECIST v1.1 by investigator assessment.

  OR is defined as complete response (CR) or partial response (PR) according to RECIST v1.1 from the date of randomization until the date of the first documentation of PD per modified RECIST v1.1 or death due to any cause. Response does not need to be confirmed at a subsequent assessment.

  DR is defined, for patients with OR, as the time from the first documentation of objective response (CR or PR) to the date of first documentation of PD per modified RECIST v1.1 or death due to any cause.

  Time to locoregional failure is defined as the time from the date of randomization to the date of the first documentation of locoregional PD per RECIST v1.1 with pathologic confirmation, locoregional clinically detectable progression with pathologic confirmation, salvage of primary tumor with tumor present on final pathology, salvage neck dissection >20 weeks after completion of CRT with tumor present on final pathology or death due to any cause, whichever occurs first.

  Time to distant metastatic failure or distant metastasis is defined as the time from the date of randomization to the date of the first documentation of metastatic PD per RECIST v1.1 or death due to any cause, whichever occurs first.

3.2.3. Patient reported outcomes

- Disease-related symptoms and HRQoL as measured by the National Cancer Comprehensive Network (NCCN) Head and Neck Symptom Index-22 items (FHNSI-22), and the EuroQoL Group 5-Dimension 5- Level Self-Report Questionnaire (EQ-5D-5L).

  The NCCN FHNSI-22 was developed using methods consistent with the Food and Drug Administration (FDA) PRO guidance (FDA, 2009) and surveyed input from patients with advanced cancers and physician experts (Cella D, et al., 2011[6]). The FHNSI-22

questionnaire is specifically designed to be a stand-alone instrument to measure disease symptoms, treatment side effects and overall quality of life in patients with head and neck cancer.

Responses on the FHNSI-22 questionnaire are used to calculate a total score and scores for four subscales: TSE, DRS-P, DRS-E, and FWB. The questionnaire contains 22 items with 5-point Likert scales ranging from 'not at all' to 'very much'. Higher scores mean better symptomatology, quality of life or functioning.

The EuroQol EQ-5D-5L is a patient-completed questionnaire designed to assess health status in terms of a single index value or utility score (Herdman M, et al., 2011[10]). There are 2 components to the EuroQol EQ-5D-5L: a descriptive system in which individuals rate their level of problems (none, slight, moderate, severe, extreme/unable) in 5 areas (mobility, self- care, usual activities, pain/discomfort, and anxiety/depression) and a Visual Analogue Scale (VAS) in which patients rate their overall health status from 0 (worst imaginable) to 100 (best imaginable). Published weights are available that allow for the creation of a single summary score for the 5 Likert scale items.

3.2.4. Pharmacokinetic endpoints

- C_{max} and C_{trough} for avelumab
- Area under the concentration-time curve extrapolated to infinity (AUC_{inf}), C_{max}, clearance (CL), time to maximum plasma concentration (T_{max}), elimination half-life ($t_{1/2}$),and volume of distribution (V_z) for cisplatin (total and free), as data permit.

Table 3. PK Parameters to be Determined for Avelumab and Cisplatin

Parameter	Definition	Method of Determination
AUC_{inf}	Area under the plasma concentration-time profile from time zero extrapolated to infinite time	$AUC_{last} + (C_{last}/k_{el})$ where C_{last} is the predicted plasma concentration at the last quantifiable time point estimated from the log-linear regression analysis, and k_{el} is the terminal phase rate constant calculated by linear regression of the log-linear concentration-time curve. Only those data points judged to describe the terminal log-linear decline will be used in the regression.
C_{max}	Maximum observed plasma concentration	Observed directly from data
T_{max}	Time for C_{max}	Observed directly from data as time of first occurrence
$t_{½}$[a]	Terminal half-life	$Log_e(2)/k_{el}$, where k_{el} is the terminal phase rate constant calculated by a linear regression of the log-linear concentration-time curve. Only those data points judged to describe the terminal log-linear decline were used in the regression.

C_{trough}	Predose concentration during multiple dosing	Observed directly from data
CL	Clearance	Dose / AUC_{inf}
V_z[a]	Apparent volume of distribution	Dose / ($AUC_{inf} \cdot k_{el}$)
AUC_{inf} (dn)[a]	Dose normalized AUC_{inf}	AUC_{inf} / Dose
C_{max} (dn)	Dose normalized C_{max}	C_{max} / Dose

[a] If data permit

3.2.5. Immunogenicity endpoints

- ADA (neutralizing antibody) against avelumab

  Anti-Drug Antibody (ADA) / Neutralizing antibodies (nAb) titers for avelumab.

3.2.6. Biomarker endpoints

- Tumor tissue biomarkers including but not limited to, PD-L1 expression and tumor-infiltrating CD8+ T-lymphocytes.

Table 4. Biomarker Definition and Determination

Parameter	Definition	Method of Determination
PD-L1	Protein expression using immunohistochemistry assay	VENTANA PD-L1 (SP263) Assay
Tumor infiltrating CD8+ lymphocytes	The number of CD8+ cells per unit area and the percent of counted cells that are scored as CD8+	Pathologist, assisted by image analysis

CCI [redacted]

3.4. Baseline Variables

3.4.1. Study drug, study treatment and baseline definitions

In this study, '**study drug**' refers to avelumab, avelumab-matching placebo, cisplatin or IMRT and '**study treatment**' (or '**treatment arm**') refers to one of the following:

- Arm A = Avelumab + SOC CRT;
- Arm B = Placebo + SOC CRT.

Start and end dates of study treatment:

The date/time of first dose of study treatment is the earliest date/time of the first non-zero dose date/time for each of the study drugs.

The date/time of last dose of study treatment is the latest date/time of the last non-zero dose date/time for each of the study drugs.

Start dates of study treatment within each treatment phase:

- Lead-in Phase

  The start date/time of study treatment is the first dose date/time of avelumab or matching placebo across the Lead-in Phase visits.

- CRT Phase

  The start date/time of study treatment is the earliest dose date/time of avelumab or matching placebo, cisplatin or IMRT across the CRT Phase visits.

- Maintenance Phase

  The start date/time of study treatment is the earliest dose date/time of avelumab or matching placebo across the Maintenance Phase visits.

Definition of baseline:

Definition of baseline for efficacy and PRO analyses

The last measurement prior to randomization will serve as the baseline measurement for efficacy and PRO analyses. If such a value is missing (since per protocol the first PRO assessment is planned to occur prior to dosing on Day 1 of the Lead-in Phase), the last measurement prior to the first dose of study treatment will be used as the baseline measurement except for analyses of tumor assessments data where the baseline assessment would be considered as missing.

Definition of baseline for immunogenicity analyses

The last available assessment prior to the start of treatment with avelumab is defined as 'baseline' result or 'baseline' assessment. If an assessment is planned to be performed prior to the first dose of avelumab in the protocol and the assessment is performed on the same day as the first dose of avelumab, it will be assumed that it was performed prior to avelumab administration, if assessment time point is not collected or is missing.

Definition of baseline for safety analyses

The last available assessment prior to the start of study treatment is defined as 'baseline' value or 'baseline' assessment for safety analyses. If an assessment is planned to be performed prior to the first dose of study treatment in the protocol and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was

performed prior to study treatment administration, if assessment time point is not collected or is missing. If assessment time points are collected, the observed time point will be used to determine pre-dose on study day 1 for baseline calculation. Unscheduled assessments will be used in the determination of baseline. However, if time is missing, an unscheduled assessment on study day 1 will be considered to have been obtained after study treatment administration.

Patients who start treatment and discontinue from the study on the same day may have two different sets of data collected on study day 1 (one during study and one in the End of Treatment (EOT) visit. Data reported at the EOT visit are not eligible for baseline selection.

If a scheduled pre-dose measurement actually occurred post-dose, then the corresponding measurement will be treated and analyzed similar to an unscheduled post-dose measurement.

Baseline for RR and QT/QTc interval assessments will be derived from the visit where both RR and QT are not missing. Triplicate ECGs are collected in the study for approximately 30 patients in each treatment arm and the baseline for each ECG measurement is the average of the pre-dose replicate measurements on the baseline day. Unscheduled assessments will not be included in the calculation of the average. QTcB and QTcF will be derived based on RR and QT. The average of the replicate measurements will be determined after the derivation of the individual parameter at each time point.

Definition of baseline for biomarker analyses

Baseline biomarker is the last biomarker assessment prior to the start of study treatment. If an assessment is planned to be performed prior to the first dose of study treatment in the protocol and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was performed prior to study treatment administration, unless otherwise defined in the protocol.

3.4.2. Baseline Characteristics

Randomization is stratified by the following, as recorded in the Interactive Response Technology (IRT):

- Tumor (T) stage (<T4 vs T4);
- Nodal (N) stage (N0/N1/N2a/N2b vs N2c/N3);
- HPV status (positive vs negative) as measured by p16 expression by IHC.

The primary analyses of PFS and OS will be stratified by these randomization stratification factors.

Other baseline characteristics (including demographics, physical measurements, and disease history) are described in Section 6.5.1. These baseline characteristics are not planned to be included as stratification variables or covariates in statistical models unless otherwise specified in Section 6.

3.5. Safety Endpoints

3.5.1. Adverse events

Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period.

On-treatment period is defined as the time from the first dose of study treatment through minimum (30 days + last dose of study treatment, start day of new anti-cancer drug therapy – 1 day). The start day of new anti-cancer drug therapy after the first dose of study treatment is derived as outlined in Section 5.2.5.

Adverse Events of Special Interest (AESIs)

AESIs are immune-related adverse events (irAE) and infusion-related reactions (IRRs). The criteria for classification of an AE as an irAE or IRR are described in Appendix 1 and Appendix 2, respectively.

3-Tier Adverse Events

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analyses are generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates these exploratory analyses.

Adverse events and clusters of adverse events, of any causality and treatment-related, will also be summarized following a 3-tier approach. Under this approach, AEs are classified into 1 of 3 tiers.

Tier-1 events: These are pre-specified events or clusters of events of clinical importance and will be described in the Safety Review Plan (SRP).

Tier-2 events: These are events that are not Tier-1 but are “common”. A MedDRA PT is defined as a Tier-2 event if it is reported by

a) at least 10% of patients with any grade in any treatment arm, or

b) at least 5% of patients with Grade 3, 4 or 5 in any treatment arm.

Tier-3 events: All other AEs that are classified neither as Tier-1 nor Tier-2.

4. ANALYSIS SETS

Data for all patients will be assessed to determine if patients meet the criteria for inclusion in each analysis population prior to releasing the database and classifications will be documented per Pfizer's standard operating procedures.

Only patients who signed informed consent will be included in the analysis sets below.

Table 5 summarizes the use of the analysis sets for efficacy, safety, baseline characteristics and exposure.

Table 5. Statistical Analyses by Analysis Set

Endpoints	Full Analysis Set	Per Protocol Analysis Set	Safety Analysis Set
Baseline Characteristics	✓		✓
Prior and Concomitant Therapies	✓		✓
Exposure			✓
Efficacy: Primary	✓	✓	
Efficacy: Secondary	✓	✓ (OS only)	
CCI			
Safety			✓

4.1. Full Analysis Set

The full analysis set (FAS) will include all randomized patients. Patients will be classified according to the study treatment assigned at randomization.

4.2. Safety Analysis Set

The safety analysis set will include all patients who receive at least one dose of study drug. Patients will be classified according to the study treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case patients will be classified according to the first study treatment received.

4.3. Other Analysis Set

4.3.1. Per-protocol analysis set

Per protocol (PP) analysis set is a subset of the FAS and will include patients who do not meet any of the following criteria that could impact the key objectives of the study. Patients who meet any of the following criteria will be excluded from the PP analysis set.

- Patient did not receive at least one dose of the randomized study treatment;
- Baseline ECOG status ≥ 2;

- Patient did not meet inclusion criterion 5 or met any of exclusion criteria 1, 13 or 17; namely patient received any anti-cancer therapy for advanced stage SCCHN or immunotherapy with an anti-PD-1, anti-PD-L1, anti-PD-L2, anti-CD137, or anti-CTLA-4 antibody (including ipilimumab), or any other antibody or drug specifically targeting T cell co-stimulation or immune checkpoint pathways prior to randomization or patient received other non-protocol active anticancer therapy within 4 weeks prior to randomization.

The list of patients who are excluded from the PP analysis set will be documented with the reason for exclusion.

4.3.2. PK analysis sets

The PK concentration analysis set is a subset of the safety analysis set and will include patients who have at least one post-dose concentration measurement above the lower limit of quantitation (LLQ) for avelumab or cisplatin (total and free).

The PK parameter analysis set is a subset of the safety analysis set and will include patients who have at least one of the PK parameters of interest for avelumab or cisplatin (total and free).

4.3.3. Biomarker analysis set

The biomarker analysis set is a subset of the safety analysis set and will include patients who have at least one screening biomarker assessment. For the optional tumor biopsy assessment, a second analysis set will include patients who have at least one dose of avelumab and who have at least one on-treatment biomarker assessment. Analysis sets will be defined separately for blood-based and tumor tissue-based biomarkers.

4.3.4. Immunogenicity analysis set

The immunogenicity analysis set is a subset of the safety analysis set and will include patients who have at least one ADA/nAb sample collected for avelumab in Arm A.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and sample size determination

The study is designed to test the following hypotheses:

$$H_{01}: HR_{PFS(A/B)} \geq 1, \text{ versus } H_{a1}: HR_{PFS(A/B)} < 1$$

where $HR_{PFS(A/B)}$ represents the hazard ratio for PFS in Arm A vs Arm B.

Approximately 640 patients will be randomized to the treatment arms using a 1:1 randomization, stratified by tumor stage (<T4 vs T4), nodal stage (N0/N1/N2a/N2b vs N2c/N3), and HPV status (positive vs negative) as measured by p16 expression by IHC.

Two hundred eight-nine (289) PFS events per modified RECIST v1.1 will be required to have at least 90% power to detect a hazard ratio of 0.68 using a 1-sided log-rank test at a

significance level of 0.025, and a 2-look group-sequential design with Lan-DeMets (O'Brien-Fleming) α-spending function to determine the efficacy boundary and a Gamma Family (-7) β-spending function to determine the non-binding futility boundary.

The sample size for this study is determined based on the assumptions that the median PFS in Arm B is 33 months (Ang KK, et al., 2010[2]; Ang KK, et al., 2014[3]) and that combination treatment with avelumab + SOC CRT (Arm A) is expected to increase the median PFS to ≥ 48.5 months, corresponding to a hazard ratio of 0.68 under the exponential model assumption. The sample size further assumes a 15% drop-out rate within each treatment arm and a non-uniform patient accrual over a 22-month period. The data cut-off for the primary PFS analysis will occur after the target number of events has been reached and the last patient randomized in the study has been followed for at least 24 months after randomization.

The study will also include a formal comparison for OS. The following hypotheses will be tested according to a testing strategy that preserves the overall type I error in the study as described in Section 5.1.2:

H_{02}: $HR_{OS(A/B)} \geq 1$, versus H_{a2}: $HR_{OS(A/B)} < 1$

where $HR_{OS(A/B)}$ represents the hazard ratio for OS in Arm A vs Arm B.

With 392 deaths, the study will have 80% cumulative power (unadjusted for the pre-testing, in the hierarchical procedure, of PFS) to detect a hazard ratio of 0.75 using a 1-sided log-rank test at a significance level of 0.025 and a 5-look group-sequential design with Lan-DeMets (O'Brien-Fleming) α-spending function to determine the efficacy boundary. If the median OS in Arm B is 70 months (Ang KK, et al., 2010[2]; Ang KK, et al., 2014[3]), a hazard ratio of 0.75 corresponds to an increase in median OS from 70 months in Arm B to 93 months in Arm A under the exponential model assumption.

Table 6 summarizes the power to detect several hazard ratios and the estimated duration of the study based on the same number of patients that are planned to be enrolled in the study to meet the primary endpoint and assuming a 5% dropout rate for survival follow-up. These calculations are conditional on achieving statistical significance for the test of PFS.

Table 6. Power to Detect Specific Hazard Ratios for OS

HR under H_{a2}	Power	Calendar time (months) to 392 deaths under H_{a2}
0.75	80%	127
0.80	58%	124
0.70	93%	131

Simulations performed using EAST 6.4.1 with number of simulations = 10,000 and seed=7282016.

5.1.2. Decision rules

To protect the integrity of the study and to preserve the type I error rate, a fraction of alpha (0.0097) for PFS efficacy will be spent at the interim analysis and accounted for in the overall type I error rate if the interim analysis is performed exactly at the planned number of

PFS events. The nominal significance levels for the interim and final efficacy analyses of PFS will be determined by using the Lan-DeMets procedure with an O'Brien-Fleming stopping boundary. The overall significance level for the efficacy analysis of PFS will be preserved at 0.025 (1-sided).

Two analyses will be performed for PFS:

1) an interim analysis after all patients have been randomized in the study and at least 217 (75%) of the 289 PFS events have been documented per modified RECIST v1.1, and
2) the final analysis for PFS after all patients randomized in the study have been followed for 24 months and at least 289 PFS events have been documented per modified RECIST v1.1.

If the value of the test statistic for PFS exceeds the efficacy boundary for a comparison ($z < -2.338$, $p < 0.010$ if the IA is performed after exactly 217 events), then superiority of Arm A compared to Arm B could be declared. If the value of the test statistic for PFS exceeds the futility boundary ($z > -0.728$, $p > 0.233$ if the IA is performed after exactly 217 events), then the study may be stopped for futility.

Since the observed number of events at the interim analysis may not be exactly equal to the planned 217 PFS events, the efficacy and futility boundaries will be updated based on the actual number of observed events using the pre-specified α-and β-spending functions. Therefore, the observed Z-test statistic at the interim analysis will be compared with the updated efficacy and futility boundaries.

If the study continues to final analysis, the p-value that will be used to declare statistical significance at the final analysis will be based on the actual number of PFS events documented at the cut-off date for the final analysis and the α already spent at the interim analysis. Therefore, if the interim analysis occurs after exactly 217 events, and the study continues until the final analysis, the observed p-value for the comparison will have to be < 0.022 to declare statistical significance. If the number of events in the final analysis deviates from the expected number of events, the final analysis criterion will be determined so that the overall significance level across all analyses is maintained at 1-sided 0.025.

Based on the stopping boundaries defined above and the timing of interim analysis at 75% information fraction the design has the following operating characteristics.

Table 7. Simulated cumulative probabilities to stop for futility at the interim or final PFS analysis

Scenario	Look	Number of PFS events	Calendar Time (months)	P(Reject H_{01})	P(Reject H_{a1})
H_{01} is true (HR=1)	Interim	217	34	0.0093	0.7679
	Final	289	44	0.0246	0.9754
H_{a1} is true (HR=0.68)	Interim	217	38	0.6917	0.0197
	Final	289	51	0.9010	0.0990

Simulations performed in EAST 6.4.1 with number of simulations = 10,000 and seed=7282016.

The secondary OS endpoint will be analyzed using a hierarchical testing procedure, provided the primary endpoint PFS endpoint is statistically significant favoring Arm A. An α-spending function according to Lan-DeMets (O'Brien-Fleming) independent of the one used for the primary efficacy analysis will be used to preserve the 0.025 overall level of significance across the hypotheses and the repeated testing of the OS hypotheses in the interim and final analyses. The trial allows for the stopping of the study for a superior OS result, provided the primary PFS endpoint has already been shown to be statistically significant favoring Arm A.

A maximum of five analyses are planned for OS:

1) an interim analysis at the time of the interim analysis for PFS (provided PFS is significant);
2) an interim analysis at the projected time of the final analysis for PFS (provided PFS is significant);
3) an interim analysis when 270 deaths are observed;
4) an interim analysis when 345 deaths are observed;
5) a final analysis for OS when 392 deaths are observed.

The exact nominal p-values that will need to be observed to declare statistical significance at the time of these analyses for OS will depend on the number of OS events that have been observed at the time of these analyses and the α for OS already spent at the time of earlier analyses.

If OS is tested alone, independent of the testing strategy for PFS, the design concerning overall survival analyses will have the following operating characteristics. These calculations are unadjusted for the pre-testing of PFS.

Table 8. Simulated cumulative probabilities to stop for efficacy on overall survival at PFS interim analysis, final PFS analysis, third interim OS analysis, fourth interim OS analysis, or final OS analysis

Scenario	Look	Number of deaths	Calendar Time (months)	P(Reject H_{0j})
H_{02} is true (HR=1)	Interim PFS	139	38	<0.0001
	Final PFS	198	51	0.0005
	3rd Interim OS	270	69	0.0067
	4th Interim OS	345	93	0.0160
	Final OS	392	112	0.0247
H_{a2} is true (HR=0.75)	Interim PFS	122	38	0.0104
	Final PFS	176	51	0.1060
	3rd Interim OS	270	78	0.4613
	4th Interim OS	345	106	0.6973
	Final OS	392	127	0.8015
H_{a2} is true (HR=0.80)	Interim PFS	126	38	0.0037
	Final PFS	181	51	0.0470
	3rd Interim OS	270	76	0.2625
	4th Interim OS	345	103	0.4738
	Final OS	392	124	0.5825

Interim and final PFS analyses calendar time expected at 38 and 51 months, respectively, under H_{a1} (HR for PFS=0.68).

Simulations performed in EAST 6.4.1 with number of simulations = 10,000 and randomization seed =7282016.

5.2. General Methods

As described in Section 3.4, in this study '**treatment arm**' refers to one of the following:

- Arm A = Avelumab + SOC CRT;
- Arm B = Placebo + SOC CRT.

Endpoints will be summarized based on the analysis sets described in Table 5 by treatment arm, unless otherwise specified.

5.2.1. Data handling after the cut-off date

Data after the cut-off date may not undergo the cleaning process and will not be displayed in any listings or used for summary statistics, statistical analyses or imputations.

5.2.2. Pooling of centers

In order to provide overall estimates of treatment effects, data will be pooled across centers. The 'center' factor will not be considered in statistical models or for subgroup analyses due to the high number of participating centers in contrast to the anticipated small number of patients randomized at each center.

5.2.3. Presentation of continuous and qualitative variables

Continuous variables will be summarized using descriptive statistics ie, number of non-missing values and number of missing values [ie, n (missing)], mean, median, standard deviation (SD), minimum, maximum and first and third quartile (Q1 and Q3).

Qualitative variables will be summarized by frequency counts and percentages. Unless otherwise specified, the calculation of proportions will include the missing category. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

In case the analysis refers only to certain visits, percentages will be based on the number of patients still present in the study at that visit, unless otherwise specified.

Confidence intervals (CIs) may also be reported as noted in subsequent sections.

5.2.4. Definition of study day

Start day of study treatment is the day of the first dose of study treatment.

The study day for assessments occurring on or after the start of study treatment (eg, adverse event onset, tumor measurement) will be calculated as:

Study day = Date of the assessment/event - start of study treatment + 1.

The study day for assessments occurring prior to the first dose of study treatment (eg, baseline characteristics, medical history) will be negative and calculated as:

Study day = Date of the assessment/event - start of study treatment.

The study day will be displayed in all relevant data listings.

5.2.5. Definition of start of new anti-cancer drug therapy

Start date of new anti-cancer drug therapy is used to determine the end of the on-treatment period (see Section 5.2.7).

The start date of new anti-cancer drug therapy is the earliest start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages that is after the first dose of study treatment. When start date of anti-cancer drug therapy is missing or partially missing, the imputation rules described in Section 5.3.3.4 should be applied using only data from the 'Follow-up Cancer Therapy' eCRF pages.

5.2.6. Definition of start of new anti-cancer therapy

The start date of new anti-cancer therapy is the earliest date after randomization amongst the following:

- Start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages
- Start date of radiation therapy recorded in 'Follow-up Radiation Therapy' eCRF pages with 'Treatment Intent' = 'Curative in intent'
- Surgery date recorded in 'Concomitant Surgery', and 'Follow-up Surgery' eCRF pages when 'Surgery Outcome' = 'Resected' or 'Partially Resected'.

When start date of anti-cancer therapy is missing or partially missing, the imputation rules described in Section 5.3.3.4 should be applied using 'Follow-up Cancer Therapy', 'Follow-up Radiation Therapy', 'Concomitant Surgery', and 'Follow-up Surgery' eCRF pages.

5.2.7. Definition of on-treatment period

Safety endpoints will be summarized based on the on-treatment period unless otherwise specified.

On-treatment period is defined as the time from the first dose of study treatment through minimum (30 days + last dose of study treatment, start day of new anti-cancer drug therapy – 1 day).

Safety data collected outside the on-treatment period as described above will be listed and flagged in listings but will not be summarized.

5.2.8. Standard derivations and reporting conventions

The following conversion factors will be used to convert days into weeks, months or years: 1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

Demographics and physical measurements:

- Age [years]:
  - (date of given informed consent - date of birth + 1) / 365.25;
  - In case of missing day, day only: Age [years]: (year/month of given informed consent – year/month of birth);
  - In case only year of birth is given: Age [years]: (year of given informed consent - year of birth).

The integer part of the calculated age will be used for reporting purposes.

- BMI (kg/m^2) = weight (kg)/[height (m)]^2

- BSA (m^2) = ([height (cm) × weight (kg)] / 3600)$^{0.5}$

For reporting conventions, mean and median should generally be displayed one more decimal place than the raw data and standard deviation should be displayed to two more decimal places than the raw data. Percentages will be reported to one decimal place. The rounding will be performed to closest integer / first decimal using the common mid-point between the two consecutive values. Eg, 5.1 to 5.4 will be rounded to an integer of 5, and 5.5 to 5.9 will be rounded to an integer of 6.

5.2.9. Unscheduled visits

Generally, data collected at unscheduled visits will be included and analyzed for both safety and efficacy analyses in the same fashion as the data collected at scheduled visits except where otherwise noted in the sections that follow. Descriptive statistics (mean, SD, median, minimum, maximum, quartiles) by nominal visit or time point for safety endpoints such as laboratory measurements, ECGs and vital signs will include only data from scheduled visits.

5.2.10. Adequate baseline tumor assessment

Adequate baseline is defined using the following criteria:

- All baseline assessments must be within 28 days prior to and including the date of randomization;
- All documented lesions must have non-missing assessments (ie, non-missing measurements for target lesions and non-missing lesions assessment status at baseline for non-target lesions).

5.2.11. Adequate post-baseline tumor assessment

An adequate post-baseline assessment is defined as an assessment where a response of uCR, uPR, SD, non-uCR/non-PD, or PD can be determined (see Section 6.2.2.8). Time points where the response is not evaluable (NE) or no assessment was performed will not be used for determining the censoring date.

5.3. Methods to Manage Missing Data

5.3.1. Missing data

Unless otherwise specified, all data will be evaluated as observed, and no imputation method for missing values will be used.

In all patient data listings imputed values will be presented. In all listings imputed information will be flagged.

Missing statistics, eg when they cannot be calculated, should be presented as ‘ND’ or ‘NA’. For example, if N=1, the measure of variability (SD) cannot be computed and should be presented as ‘ND’ or ‘NA’.

5.3.1.1. Pharmacokinetic concentrations

Concentrations Below the Limit of Quantification

For all calculations, figures and estimation of individual pharmacokinetic parameters, all concentrations assayed as below the level of quantification (BLQ) will be set to zero. In log-linear plots these values will not be represented. The BLQ values will be excluded from calculations of geometric means and their CIs. A statement similar to 'All values reported as BLQ have been replaced with zero' should be included as a footnote to the appropriate tables and figures.

Deviations, Missing Concentrations and Anomalous Values

In summary tables and plots of median profiles, concentrations will be set to missing if one of the following cases is true:

1. A concentration has been reported as ND (ie, not done) or NS (ie, no sample);
2. A deviation in sampling time is of sufficient concern or a concentration has been flagged as anomalous by the clinical pharmacologist.

Summary statistics will not be presented at a particular time point if more than 50% of the data are missing. For analysis of pharmacokinetic concentrations, no values will be imputed for missing data.

5.3.1.2. Pharmacokinetic parameters

Actual or nominal PK sampling time will be used for the derivation of PK parameters. If a PK parameter cannot be derived from a patient's concentration data, the parameter will be coded as NC (ie, not calculated). NC values will not be generated beyond the day that a patient discontinues.

In summary tables, statistics will be calculated by setting NC values to missing. Statistics will not be presented for a particular treatment if more than 50% of the data are NC. For statistical analyses (ie, analysis of variance), PK parameters coded as NC will also be set to missing.

If an individual patient has a known biased estimate of a PK parameter (due for example to a deviation from the assigned dose level), this will be footnoted in summary tables and will not be included in the calculation of summary statistics or statistical analyses.

5.3.2. Handling of incomplete dates

5.3.2.1. Disease history

Incomplete dates for disease history (eg, initial diagnosis date, date of documented, locally advanced, inoperable or metastatic disease diagnosis, date of response or progression in prior treatment) will be imputed as follows:

- If the day is missing, it will be imputed to the 15th day of the month.

- If both day and month are missing and the year is prior to the year of the first study treatment, the month and day will be imputed as July 1st.
- If both day and month are missing and the year is same as the year of the first study treatment, the month and day will be imputed as January 1st.
- If the date is completely missing, no imputation will be performed.

5.3.2.2. Adverse events

Incomplete AE-related dates will be imputed as follows:

- If the AE onset date is missing completely, then the onset date will be replaced by the start of study treatment.
- If only the day part of the AE onset date is missing, but the month and year are equal to the start of study treatment, then the AE onset date will be replaced by the start of study treatment. For example, if the AE onset date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed AE onset date will be 15/JAN/2015.
- If both the day and month of the AE onset date are missing but the onset year is equal to the start of study treatment, then the onset date will be replaced by the start of study treatment. For example, if AE onset date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed AE onset date will be 19/NOV/2014.
- In all other cases the missing onset day or missing onset month will be replaced by 1.
- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of patient's death. In the latter case the date of death will be used to impute the incomplete stop date.
- In all other cases the incomplete stop date will not be imputed. If stop date of AE is after the date of cut-off outcome of AE is ongoing at cut-off.

5.3.2.3. Prior and concomitant medications

Incomplete prior/concomitant medication dates will be imputed as follows:

- If the medication date is missing completely, then the medication date will be replaced by the start of study treatment.
- If the day of medication date is missing, but the month and year are equal to the start of study treatment, then the medication date will be replaced by the start of study treatment. For example, if the medication start date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed medication start date will be 15/JAN/2015.
- If both the day and month of medication start date are missing but the start year is equal to the start of study treatment, then the medication date will be replaced by the start of study treatment. For example, if the medication start date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed medication start date will be 19/NOV/2014.

- In all other cases the missing medication day or missing medication month will be replaced by 1.
- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of patient's death. In the latter case the date of death will be used to impute the incomplete stop date.
- In all other cases the incomplete medication stop date will not be imputed.

5.3.2.4. Exposure

No imputation will be done for first dose date. Date of last dose of study drug, if unknown or partially unknown, will be imputed as follows:

- If the last date of study drug is completely missing and there is no End of Treatment eCRF page and no death date, the patient should be considered to be ongoing and use the cut-off date for the analysis as the last dosing date;
- If the last date of study drug is completely or partially missing and there is EITHER an End of Treatment eCRF page OR a death date available (within the cut-off date), then imputed last dose date is:

  = 31DECYYYY, if only Year is available and Year < Year of min (EOT date, death date)

  = Last day of the month, if both Year and Month are available and Year = Year of min (EOT date, death date) and Month < the month of min (EOT date, death date)

  = min (EOT date, death date), for all other cases.

5.3.3. Imputation rules for date of last contact and efficacy assessments

5.3.3.1. Date of last contact

The date of last contact will be derived for patients not known to have died at the analysis cut-off using the latest complete date among the following:

- All patient assessment dates (blood draws (laboratory, PK), vital signs, performance status, ECG, tumor assessments);
- Start and end dates of anti-cancer therapies administered after study treatment discontinuation;
- AE start and end dates;
- Last date of contact collected on the ‘Survival Follow-up’ eCRF (do not use date of survival follow-up assessment unless status is ‘alive’);
- Study drug start and end dates;
- Randomization date;
- Withdrawal of consent date;

- Date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up).

Only dates associated with actual examinations of the patient will be used in the derivation. Dates associated with a technical operation unrelated to patient status such as the date a blood sample was processed will not be used. Assessment dates after the cut-off date will not be applied to derive the last contact date.

5.3.3.2. Death date

Missing or partial death dates will be imputed based on the last contact date:

- If the date is missing it will be imputed as the day after the date of last contact
- If the day or both day and month is missing, death will be imputed to the maximum of the full (non-imputed) day after the date of last contact and the following:
  - Missing day: 1st day of the month and year of death
  - Missing day and month: January 1st of the year of death

5.3.3.3. Tumor assessments

All investigation dates (eg, X-ray, CT scan) must be completed with day, month and year.

If there are multiple scan dates associated with an evaluation, ie, radiological assessments occur over a series of days rather than the same day, the choice of date of assessment could impact the date of progression and/or date of response. If there are multiple scan dates associated with an evaluation, the earliest of the scan dates associated with the evaluation will be used as the date of assessment.

If one or more investigation dates for an evaluation are incomplete but other investigation dates are available, the incomplete date(s) are not considered for calculation of the assessment date and assessment date is calculated as the earliest of all investigation dates (eg, X-ray, CT-scan).

If all measurement dates for an evaluation have no day recorded, the 1st of the month is used.

If the month is not completed, for any of the investigations for an evaluation, the respective assessment will be considered to be at the date which is exactly between the previous and the following assessment. If both a previous and following assessments are not available, this assessment will not be used for any calculations.

5.3.3.4. Date of start of new anti-cancer therapy

Incomplete dates for start date of new anti-cancer therapy (drug therapy, radiation, surgery) will be imputed as follows and will be used for determining censoring dates for sensitivity efficacy analyses and in the derivation of the end of on-treatment period. PD date below refers to PD date by modified RECIST V1.1.

- The end date of new anti-cancer therapy will be included in the imputations for start date of new anti-cancer therapy. If the end date of new anti-cancer therapy is
  - completely missing then it will be ignored in the imputations below
  - partially missing with only year (YYYY) available then the imputations below will consider 31DECYYYY as the end date of the new anti-cancer therapy
  - partially missing with only month and year available then the imputations below will consider the last day of the month for MMMYYYY as the end date of the new anti-cancer therapy
- For patients who have not discontinued study treatment at the analysis cut-off date, last dose of study treatment is set to the analysis cut-off date in the imputations below.
- If the start date of new anti-cancer therapy is completely or partially missing then the imputed start date of new anti-cancer therapy is derived as follows:
  - <u>Start date of new anti-cancer therapy is completely missing</u>

    Imputed start date = min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]
  - <u>Only year (YYYY) for start of anti-cancer therapy is available</u>

    IF YYYY < Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy] THEN imputed start date = 31DECYYYY;

    ELSE IF YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

    THEN imputed start date = min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

    ELSE IF YYYY > Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

    THEN imputed start date = 01JANYYYY
  - <u>Both Year (YYYY) and Month (MMM) for start of anti-cancer therapy are available</u>

    IF

    YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

    MMM < Month of min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

    THEN

    imputed start date = DAY (Last day of MMM) MMM YYYY;

    ELSE IF

    YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

MMM = Month of min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

THEN

imputed start date = min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]);

ELSE IF

YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

MMM > Month of min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

THEN

imputed start date = 01 MMM YYYY;

ELSE IF

YYYY < Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN

imputed start date = DAY (Last day of MMM) MMM YYYY;

ELSE IF

YYYY > Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN

imputed start date = 01 MMM YYYY.

6. ANALYSES AND SUMMARIES

Refer to Section 4 for definitions of analysis sets and Section 5.2 for general methodology.

6.1. Primary Endpoints

6.1.1. Progression-free survival per modified RECIST v1.1

6.1.1.1. Primary analysis

The following analyses will be based on the FAS using the strata assigned at randomization. PD below refers to PD by investigator assessment per modified RECIST v1.1. Tumor-related endpoints in this study will be based on RECIST v1.1 modified with respect to the definition of what constitutes PD as follows (Appendix 3 of the B9991016 protocol). Any of the following will constitute PD per modified RECIST v1.1:

- Locoregional disease progression per RECIST v1.1 which is subsequently confirmed by pathology. Pathologic confirmation of progression will verify that radiographic changes

represent true tumor progression and not radiation effects or non-malignant contrast enhancement;

- Locoregional clinically detectable progression that is confirmed by pathology;
- Surgical removal (salvage) of primary tumor with tumor present on final pathology;
- Salvage neck dissection >20 weeks (140 days) after completion of CRT with tumor present on final pathology;
- Metastatic (distant metastases) disease progression per RECIST v1.1. Confirmation of pathology is recommended unless medically contraindicated or lesion location too high risk for biopsy.

Local failure is defined as residual or recurrent viable tumor on pathologic evaluation from the site of original tumor location. Presence of tumor is required for declaration of a progression event which requires identification of viable tumor cells on final pathology. Tumor reappearing within the initial and immediately adjoining anatomical region of the primary will be considered local recurrence.

Regional failure is defined as residual or recurrent viable tumor on pathologic evaluation from the regional lymph node basins (eg, neck nodes). Presence of tumor is required for declaration of a progression event which requires identification of viable tumor cells on final pathology.

Distant metastatic disease is defined as new tumor identified at a site distant from the head and neck anatomic region or draining lymph nodes. Biopsy of any presumed distant metastatic disease is strongly recommended. A solitary, speculated lung mass/nodule is a second primary neoplasm and is not a disease progression event unless proven otherwise by biopsy in a patient with a smoking history. Multiple lung nodules/masses are considered distant metastases from the index cancer and constitutes a disease progression event unless proven otherwise by biopsy.

Irradiation of the primary tumor and radiographically enlarged lymph nodes is considered part of the treatment under investigation in this clinical trial and therefore these target lesions remain evaluable lesions for response assessment.

It is expected that the status of the primary tumor is assessed thoroughly at the beginning of the surgical procedure before undertaking nodal dissection. Presence of persistent disease at the primary site, confirmed by frozen section will be considered disease progression.

Neck dissection parameters are described in Section 3.4 of the B9991016 protocol and will consist of a selective neck dissection unless a cytologic sampling of the nodes that appear enlarged is negative. Positive neck specimens removed within 140 days after completion of CRT will be considered part of the initial treatment plan and not considered as failures of initial management; positive specimens upon neck dissection beyond 140 days will be considered regional failures.

Progression-Free Survival (PFS) is defined as the time from the date of randomization to the date of the first documentation of PD or death due to any cause, whichever occurs first. For example, the date of first documentation of PD when there is locoregional PD per RECIST v1.1 which is subsequently confirmed by pathology, will be the date of the locoregional PD per RECIST v1.1 rather than the date of pathology confirmation.

PFS data will be censored on the date of the last adequate tumor assessment or clinical evaluation for patients who do not have an event (PD or death); the analysis will consider any PD or death as an event regardless of the number of prior missing tumor assessments or clinical evaluations or timing of the event with respect to initiation of subsequent anti-cancer therapy.

The censoring and event date options to be considered for the PFS and DR analysis are presented in Table 9.

PFS (months) = [date of event or censoring– date of randomization +1]/30.4375

Table 9. Outcome and Event Dates for PFS Analyses

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment and no PD and no death	Date of randomization	Censored
No adequate post-baseline tumor assessment or clinical evaluation and no death	Date of randomization	Censored
PD or death	Date of PD or death	Event
No PD and no death	Date of last adequate tumor assessment or clinical evaluation documenting no PD	Censored

The primary efficacy analysis will compare the PFS time between Arm A and Arm B, and will be performed using a 1-sided stratified log-rank test as described in Section 5.1.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=Arm B, 1= Arm A) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

In order to account for the group sequential design in this study, the repeated CI (RCI) method (Jennison and Turnbull, 2000[11]), will be used to construct the 2-sided RCIs for the hazard ratio at the interim and the final analyses of PFS.

In addition, the unadjusted 95% CIs for the hazard ratio will also be reported at the interim and the final analyses for PFS.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median PFS time with 2-sided 95% CIs. In particular, the PFS rates at 12, 24, 36, 48 and 60 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[4] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[12] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with each event type (PD or death) and censoring reasons will be presented by treatment arm.

Reasons for censoring will be summarized according to the categories in Table 10 following the hierarchy shown.

Table 10. PFS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment AND no PD AND no death	No adequate baseline assessment
2	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Patient refused further follow-up]	Withdrawal of consent
3	No event and lost to follow-up in any disposition page	Lost to follow-up
4	No event and [EOS present OR disposition page for any epoch after screening says patient will not continue into any subsequent phase of the study] and no adequate post-baseline assessment	No adequate post-baseline assessment
5	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

The PFS time or censoring time and the reasons for censoring will also be presented in a patient listing.

Frequency (number and percentage) of patients with each subtype of PD will be presented by treatment arm. The subtype of PD is collected in the eCRF and includes the following:

- Locoregional PD per RECIST v1.1 with pathologic confirmation
- Locoregional clinically detectable progression with pathologic confirmation
- Surgical removal (salvage) of primary tumor with tumor present on final pathology
- Salvage neck dissection >20 weeks after completion of SOC CRT with tumor present on final pathology

- Metastatic (distant metastases) disease progression per RECIST v1.1.

Time of Follow-Up for PFS

A plot will be generated to compare planned and actual relative day of tumor assessments by treatment arm. A Kaplan-Meier plot for PFS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the PFS censoring and event indicators. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median time of follow-up for PFS with 2-sided 95% CIs. In particular, the rates at 12, 24, 36, 48 and 60 months will be estimated with corresponding 2-sided 95% CIs.

6.2. Secondary Endpoint(s)

6.2.1. Safety endpoints

Refer to Section 6.6.

6.2.2. Efficacy endpoints

The following analyses will be based on the FAS by treatment arm unless otherwise specified.

Response endpoints (Best Overall Response [BOR], unconfirmed Objective Response [uOR], DR, TTR) will be analyzed separately based on investigator assessment per modified RECIST v1.1 and based on investigator assessment per programmatically derived standard RECIST v1.1 (Appendix 3).

6.2.2.1. Overall survival

The following analyses will be based on the FAS using the strata assigned at randomization.

Overall survival (OS) is defined as the time from the date of randomization to the date of death due to any cause. Patients last known to be alive will be censored at date of last contact.

OS (months) = [date of death or censoring– date of randomization +1]/30.4375

The primary analysis of OS will compare the OS time between Arm A and Arm B, and will be performed using a 1-sided stratified log-rank test as described in Section 5.1.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=Arm B, 1= Arm A) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

In order to account for the group sequential design in this study, the repeated CI (RCI) method (Jennison and Turnbull, 2000[11]), will be used to construct the 2-sided RCIs for the hazard ratio at the interim and the final analyses of OS.

In addition, the unadjusted 95% CIs for the hazard ratio will also be reported at the interim and the final analyses for OS.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median OS time with 2-sided 95% CIs. In particular, the OS rates at 12, 24, 36, 48, 60, 72, 84, and 96 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[4] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[12] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with an event (death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 11 following the hierarchy shown.

Table 11. OS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Patient refused further follow-up]	Withdrawal of consent
2	No event and [lost to follow-up in any disposition page OR data cut-off date – last contact date > 18 weeks]	Lost to follow-up
3	No event and none of the conditions in the prior hierarchy are met	Alive

The OS time or censoring time and the reasons for censoring will also be presented in a patient listing.

Time of Follow-Up for OS

A Kaplan-Meier plot for OS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the OS censoring and event indicators. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median time of follow-up for OS with 2-sided 95% CIs. In particular, the rates at at 12, 24, 36, 48, 60, 72, 84, and 96 months will be estimated with corresponding 2-sided 95% CIs.

6.2.2.2. Sensitivity analyses for progression-free survival

The following sensitivity analyses will be performed to explore the robustness of the primary analysis results. These analyses are regarded as purely exploratory. The sensitivity analyses

will repeat the primary analysis (p-value, HR and 95% CIs) described in Section 6.1.1.1 with the modifications below:

- M1 - PFS per modified RECIST v1.1 following the censoring rules outlined in Table 12
- M2 - PFS per modified RECIST v1.1 on the PP analysis set using the censoring rules outlined in Table 9
- M3- PFS per modified RECIST v1.1 using an unstratified analysis using the censoring rules outlined in Table 9
- M4 - PFS per modified RECIST v1.1 using strata derived according to eCRF data instead of that entered in IRT using the censoring rules outlined in Table 9
- S0 - PFS with PD programmatically derived based on investigator assessment per standard RECIST v1.1 with the censoring rules defined in Table 13
- S1, S2, S3, S4 replicating the sensitivity analyses M1, M2, M3 and M4, respectively, for PFS with PD programmatically derived based on investigator assessment per standard RECIST v1.1. The censoring rules associated with S1 are outlined in Table 14
- S0_B - PFS with PD based on Blinded Independent Central Review (BICR) assessment per standard RECIST v1.1 with the censoring rules defined in Table 13
- M1_B - PFS with PD based on BICR assessment per standard RECIST v1.1 with the censoring rules defined in Table 12.

Table 12. Outcome and Event Date for PFS Sensitivity Analysis (M1)

Scenario	Date of Progression/Censoring	Outcome
No adequate baseline assessment	Date of randomization [a]	Censored [a]
PD or death - following at most one missing or indeterminate post-baseline tumor assessment or clinical evaluation, OR - ≤ 36 weeks after the date of randomization	Date of PD or death	Event
PD or death following two or more missing or indeterminate post-baseline tumor assessments or clinical evaluations [b]	Date of last adequate tumor assessment [b] or clinical evaluation documenting no PD prior to new anti-cancer therapy or missed assessments	Censored
No PD and no death		
New anti-cancer therapy given		
[a] If the patient dies ≤36 weeks after date of randomization, the death is an event with date on death date. [b] If there are no adequate post-baseline tumor assessments or clinical evaluations prior to the PD or death, then the time without adequate assessment should be measured from the date of randomization; if the criteria were met the censoring will be on the date of randomization		

Table 13. Outcome and Event Date for PFS Sensitivity Analysis (S0)

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment and no PD and no death	Date of randomization	Censored
No adequate post-baseline tumor assessment and no death	Date of randomization	Censored
PD or death	Date of PD or death	Event
No PD and no death	Date of last adequate tumor assessment documenting no PD	Censored

Table 14. Outcome and Event Date for PFS Sensitivity Analysis (S1)

Scenario	Date of Progression/Censoring	Outcome
No adequate baseline assessment	Date of randomization [a]	Censored [a]
PD or death - following at most one missing or indeterminate post-baseline tumor assessment, OR - ≤ 36 weeks after the date of randomization	Date of PD or death	Event
PD or death following two or more missing or indeterminate post-baseline tumor assessments [b]	Date of last adequate tumor assessment [b] documenting no PD prior to new anti-cancer therapy or missed assessments	Censored
No PD and no death		
New anti-cancer therapy given		

[a] If the patient dies ≤36 weeks after start date, the death is an event with date on death date.
[b] If there are no adequate post-baseline tumor assessments prior to the PD or death, then the time without adequate assessment should be measured from the date of randomization; if the criteria were met the censoring will be on the date of randomization

Methods for evaluating the validity of model assumptions

The proportional hazards assumption will be checked visually by plotting log(-log(PFS)) versus log(time) within each randomization stratum.

Schoenfeld residuals for the stratified Cox proportional regression model will be plotted to investigate graphically violations from the proportional hazards (PH) assumption; a non-zero slope is evidence of departure from PH. The PH assumption will be formally tested using Schoenfeld's residual test (Schoenfeld, 1980[18]; Therneau & Grambsch, 2000[20]). Large departures from PH will be evidenced by a p-value <0.05.

If these show large departures from proportional hazards, then PFS by modified RECIST v1.1 will also be analyzed based on restricted mean survival time (RMST) differences (Zhang, 2013[22]).

Restricted Mean Survival Time (RMST)

The hazard ratio estimate from the Cox proportional hazard model is routinely used to empirically quantify the between-arm difference under the assumption that the ratio of the two hazard functions is constant over time. When this assumption is plausible, such a ratio estimate may capture the relative difference between two survival curves. However, the clinical meaning of such a ratio estimate is difficult, if not impossible, to interpret when the underlying PH assumption is violated (ie, the hazard ratio is not constant over time).

The RMST is a robust and clinically interpretable summary measure of the survival time distribution. Unlike median survival time, it is estimable even under heavy censoring. There is a considerable body of methodological research (eg, Royston and Parmar, 2011[16]; Uno,Wei, et al., 2014[21]; Zhang, 2013[22]) about the use of RMST to estimate treatment effects as an alternative to the hazard ratio approach.

The RMST methodology is applicable independently of the PH assumption and can be used, at a minimum, as a sensitivity analysis to explore the robustness of the primary analysis results. However, when large departures from the PH assumption are observed, the log-rank test is underpowered to detect differences between the survival distributions for the treatment arms, and a test of the difference between the RMST for the experimental arm and the control arm may be more appropriate to determine superiority of the experimental arm compared to the control arm with respect to the time-to-event endpoint.

In particular, as it pertains to the **cut-off point (τ)** to evaluate the RMST, it is noted that the cut-off point should not exceed the minimum of the largest observed time for both treatment arms so that the RMST of all treatment arms being evaluated can be adequately estimated and comparison between treatments is feasible; τ should be clinically meaningful and closer to the end of the study follow-up so that the majority of survival outcomes will be covered by the time interval. The RMST up to time τ can then be interpreted as the expected survival time restricted to the common follow-up time τ among all patients. The selection of τ should ensure that the RMST evaluation will not go beyond the maximum time point where the evaluation can be performed while also taking into account a large period of time that is expected to provide a meaningful assessment of treatment effect. To avoid arbitrary selection of the common cut-off τ for both treatment arms, three sets of analyses will be performed:

- τ_1 = minimum of (largest observed PFS time for the experimental arm, largest observed PFS time for the control arm).
- τ_2 = minimum of (largest PFS event time for the experimental arm, largest PFS event time for the control arm).
- τ_3 = midpoint between τ_1 and τ_2.

The treatment effect between the experimental arm and the control arm will be assessed based on the difference in RMST. The associated 95% CI for the difference in means and 1-sided p-value will be generated.

Exploratory analyses to investigate the impact of potential prognostic or effect modifying (predictive) factors

See subgroups as defined in Section 6.4.

Multivariable Cox regression analysis will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The Cox's Proportional Hazard model is defined as:

$$h(t) = h(0;t)\, e^{Xb}$$

where h(0;t) defines the baseline hazard function and X defines the vector of explanatory variables and b the unknown vector of regression parameters.

In the stepwise selection procedure, variables are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the subsequent backward elimination. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald test). This analysis will be performed using the stepwise selection method in SAS (Proc PHREG). Once this procedure stops, the factor 'treatment arm' will be added to the last selected model in order to evaluate the effect of treatment on PFS time when adjusted for the selected explanatory variables. The hazard ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% CIs. No interactions will be considered. Post-baseline factors will not be considered for the model.

PFS per modified RECIST v1.1 vs PFS per standard RECIST v1.1

Summaries of PFS assessment per modified RECIST v1.1 versus standard RECIST v1.1 will be provided including numbers of concordant and discordant assessments as well as the number of cases where PFS is determined to occur at different timepoints based on the censoring rules described in Table 13 (S0 based on programmatically derived investigator assessment or S0_B based on BICR assessment).

Table 15 outlines the possible outcomes per modified RECIST v1.1 and standard RECIST v1.1, a similar classification used for possible outcomes of assessment by investigator and BICR (Amit O, et al., 2011).[1]

Table 15. Possible Outcomes per Modified RECIST v1.1 vs Standard RECIST v1.1

		Standard RECIST v1.1	
		Event	**No Event**
Modified RECIST v1.1	**Event**	a = a1 + a2 + a3	b
	No Event	c	d

a1: number of agreements on timing and occurrence of event;
a2: number of times agreement on event but modified RECIST v1.1 declares event later than RECIST v1.1;
a3: number of times agreement on event but modified RECIST v1.1 declares event earlier than RECIST v1.1;
N= a+b+c+d.
The timing agreement of event is defined with a window of 28 days

The following measure of discordance will be calculated for each treatment arm:

- Total Event Discrepancy Rate: (b+c) / N
- Early Discrepancy Rate (EDR): (a3+b) / (a+b)
- Late Discrepancy Rate (LDR): (a2+c) / (a2+a3+b+c)
- Overall Discrepancy Rate: (a2+a3+b+c) / N

The EDR represents the positive predictive value of modified RECIST v1.1 assessment and quantifies the frequency with which modified RECIST v1.1 will declare PFS event earlier than standard RECIST v1.1 within each arm as a proportion of the total number of modified RECIST v1.1 assessed PFS events.

The LDR quantifies the frequency with which modified RECIST v1.1 criteria assigns PFS event later than standard RECIST v1.1 criteria as a proportion of the total number of discrepancies within the treatment arm.

Biopsy

The following will be listed and summarized by treatment arm.

Disease site, lesion number, clinical evaluation date and outcome, imaging evaluation date and outcome, whether biopsy was done, reason, date of biopsy sample and biopsy outcome, will be listed for each patient. Lesions evaluated after PD per modified RECIST v1.1 will be flagged in the listing.

In what follows, lesions evaluated from randomization until PD per modified RECIST v1.1 will be considered.

The number of times a lesion underwent clinical evaluation and/or underwent imaging evaluation will be summarized overall and by disease site.

For each clinical evaluation outcome and imaging evaluation outcome and for the combination of outcomes), the following will be summarized. The denominator to calculate percentages is the total number of lesions evaluations; all lesion evaluations are counted if the same lesion is evaluated at multiple visits.

- Number of lesion evaluations
- Number and percentage of biopsies performed, reason and outcome
- Number of and percentage of biopsies not performed and reason

The following cross tabulations will be provided. The denominator to calculate the percentages will be based on the total number of lesions evaluations; all lesion evaluations are counted if the same lesion is evaluated at multiple visits. Biopsy status and outcome will include all combination of categories for biopsy performed (yes, no), reason to perform or not perform, and outcome if performed (positive, negative, inconclusive)

- Clinical evaluation outcome vs. biopsy status and outcome
- Imaging evaluation outcome vs. biopsy status and outcome
- Combination of clinical and imaging outcome vs biopsy status and outcome
- Clinical evaluation outcome vs. imaging evaluation outcome.

6.2.2.3. Sensitivity analyses for overall survival

The following sensitivity analyses will be performed to explore the robustness of the primary analysis results for OS. These analyses are regarded as purely exploratory. The sensitivity analyses will repeat the primary analysis (p-value, HR and 95% CIs) described in Section 6.2.2.1 with the modifications below:

- PP analysis set;
- unstratified;
- using strata derived according to eCRF data instead of that entered in IRT.

Methods for evaluating the validity of model assumptions

The same methodology described in Section 6.2.2.2 for PFS will be used for OS.

Exploratory analyses to investigate the impact of potential prognostic or effect modifying (predictive) factors

The same methodology described in Section 6.2.2.2 for PFS will be used for OS.

6.2.2.4. Locoregional failure

Locoregional failure (LRF) is defined as locoregional PD per RECIST v1.1 with pathologic confirmation or locoregional clinically detectable progression with pathologic confirmation or salvage of primary tumor with tumor present on final pathology or salvage neck dissection >20 weeks after completion of CRT with tumor present on final pathology.

Time to LRF is defined as the time from the date of randomization to the date of the first documentation of locoregional PD per RECIST v1.1 with pathologic confirmation, locoregional clinically detectable progression with pathologic confirmation, salvage of primary tumor with tumor present on final pathology, salvage neck dissection >20 weeks after completion of CRT with tumor present on final pathology or death due to any cause, whichever occurs first.

Time to LRF data will be censored on the date of the last adequate tumor assessment or clinical evaluation for patients who do not have an event (LRF or death); the analysis will consider any LRF or death as an event regardless of the number of prior missing tumor assessments or clinical evaluations or timing of the event with respect to initiation of anti-cancer therapy.

The censoring and LRF event date options to be considered are presented in Table 16.

LRF (months) = [date of event or censoring– date of randomization +1]/30.4375

Table 16. Outcome and Event Dates for Time to LRF Analysis

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment and no LRF and no death	Date of randomization	Censored
No adequate post-baseline tumor assessment or clinical evaluation and no death	Date of randomization	Censored
LRF or death	Date of LRF or death	Event
No LRF and no death	Date of last adequate tumor assessment or clinical evaluation documenting no LRF	Censored

The time to LRF in Arm A and Arm B will be compared based on the FAS using the strata assigned at randomization and using a 1-sided stratified log-rank test at 1-sided nominal alpha level of 0.025.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie, for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=Arm B, 1= Arm A) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median LRF time with 2-sided 95% CIs. In particular, the LRF rates at 12, 24, 36, 48 and 60 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to

Brookmeyer and Crowley (1982)[4] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[12] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with each event type (LRF or death) and censoring reasons will be presented by treatment arm.

Reasons for censoring will be summarized according to the categories in Table 17 following the hierarchy shown.

Table 17. Time to LRF Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment AND no LRF AND no death	No adequate baseline assessment
2	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Patient refused further follow-up]	Withdrawal of consent
3	No event and lost to follow-up in any disposition page	Lost to follow-up
4	No event and [EOS present OR disposition page for any epoch after screening says patient will not continue into any subsequent phase of the study] and no adequate post-baseline assessment	No adequate post-baseline assessment
5	No event (no LRF AND no Death) AND Distant Metastatic (DM) failure	Distant Metastatic failure only
6	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

6.2.2.5. Distant metastatic failure

Distant metastatic failure or distant metastasis (DM) is defined as new tumor identified at a site distant from the head and neck anatomic region or draining lymph nodes.

Time to DM is defined as the time from the date of randomization to the date of the first documentation of metastatic disease progression per RECIST v1.1 or death due to any cause, whichever occurs first. DM data will be censored on the date of the last adequate tumor assessment or clinical evaluation for patients who do not have an event (DM or death); the analysis will consider any DM or death as an event regardless of the number of prior missing tumor assessments or clinical evaluations or timing of the event with respect to initiation of anti-cancer therapy.

The censoring and DM event date options to be considered are presented in Table 18.

DM (months) = [date of event or censoring– date of randomization +1]/30.4375

Table 18. Outcome and Event Dates for Time to DM analysis

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment and no DM and no death	Date of randomization	Censored
No adequate post-baseline tumor assessment or clinical evaluation and no death	Date of randomization	Censored
DM or death	Date of DM or death	Event
No DM and no death	Date of last adequate tumor assessment or clinical evaluation documenting no DM	Censored

The time to DM in Arm A and Arm B will be compared based on the FAS using the strata assigned at randomization and using a 1-sided stratified log-rank test at 1-sided nominal alpha level of 0.025.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie, for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=Arm B, 1= Arm A) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median DM time with 2-sided 95% CIs. In particular, the DM rates at 12, 24, 36, 48 and 60 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[4] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[12] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with each event type (DM or death) and censoring reasons will be presented by treatment arm.

Reasons for censoring will be summarized according to the categories in Table 19 following the hierarchy shown.

Table 19. Time to DM Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment AND no DM AND no death	No adequate baseline assessment
2	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Patient refused further follow-up]	Withdrawal of consent
3	No event and lost to follow-up in any disposition page	Lost to follow-up
4	No event and [EOS present OR disposition page for any epoch after screening says patient will not continue into any subsequent phase of the study] and no adequate post-baseline assessment	No adequate post-baseline assessment
5	No event (no DM AND no Death) AND Locoregional (LRF) failure	Locoregional failure only
6	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

6.2.2.6. Neck dissection

A patient who had neck dissection can be identified by the "Surgery Details –Salvage Surgery Neck" eCRF. If the box "Yes" was checked for the question "Was neck dissection performed as expected per protocol?", the patient had a neck dissection.

The rate of neck dissection on each treatment arm will be estimated by dividing the number of patients with neck dissection recorded from randomization until PD per modified RECIST v1.1 or death due to any cause by the number of patients randomized to the respective treatment arm. The corresponding exact 2-sided 95% CIs will be provided by treatment arm.

The rate of pathologically-positive neck dissection and rate of pathologically-negative neck dissection will be summarized using the same methodology as for the rate of neck dissection.

The final pathology of neck dissection (positive or negative) can be identified from the "Outcome" field in the "Pathology Details" eCRF with "Lesion number" and "Date of sample" matching "Lesion number" and "Date of neck dissection" in the "Surgery Details – Salvage Surgery Neck" eCRF.

- Positive pathology includes the following outcomes from the "Pathology Details" eCRF: (1) live tumor cells present, or (2)10% or greater vital tumor tissues.
- Negative pathology includes the remaining outcomes: (3) no live tumor cells present, (4) complete tumor regression, no evidence of vital tumor tissues, (5) less than 10% vital tumor tissue, or (6) not consistent with disease under study.

The outcome of the pre-surgery biopsy sample (positive or negative) can be identified from the "Outcome of biopsy" field in the "Biopsy Sample Details" eCRF with "Lesion number" and "Date of biopsy sample" matching "Lesion number" and "Date of pre-surgery biopsy sample" in the "Surgery Details –Salvage Surgery Neck" eCRF.

The positive and negative predictive values (PPV and NPV respectively) of the biopsy to the final pathology will be calculated based on the subset of patients who had neck dissection and a pre-surgery biopsy sample (matching lesion number and date).

- PPV = TP / (TP + FP) where TP is the number of lesions with true positive pre-surgery biopsy (pre-surgery biopsy positive and positive pathology) and FP is the number of lesions with false positive pre-surgery biopsy (pre-surgery biopsy positive and negative pathology).

- NPV = TN / (TN + FN) where TN is the number of lesions with true negative pre-surgery biopsy (pre-surgery biopsy negative and negative pathology) and FN is the number of lesions with false negative pre-surgery biopsy (pre-surgery biopsy negative and positive pathology).

6.2.2.7. Pathologic complete response

Pathologic complete response (pCR) is defined as the absence of histologically identifiable residual cancer in any resected specimen and will be summarized separately for randomized patients who underwent salvage surgery at the primary site and for randomized patients who had neck dissection. Further, a cross tabulation of pCR at the primary site (responder, not a responder, no salvage surgery) and in the neck (responder, not a responder, no salvage surgery) will be provided.

Patients who underwent salvage surgery are identified in the “Surgery Details – Salvage Surgery Primary Site” or “Surgery Details – Salvage Surgery Neck” with answer “Yes” to “Was salvage surgery at primary site performed?” or “Was neck dissection performed as expected per protocol?”, respectively.

Pathologic complete responders are patients who underwent, prior to PD per modified RECIST v1.1, salvage surgery/ neck dissectionwith a negative pathology, ie, with outcome in “Pathology Details eCRF” as “no live tumor cells present”, “complete tumor regression”, “no evidence of vital tumor tissues”, “less than 10% vital tumor tissue”, or “not consistent with disease under study”. Patients should be matched based on the lesion number and date of the sample if the “Pathology Details eCRF” and date of the surgery at the primary site/date of neck dissection.

If more than one lesion was resected prior to PD per modified RECIST v1.1, a pathologic complete responder is a patient with 1) no resected lesion with a positive pathology AND 2) with at least one resected lesion with negative pathology.

The pCR rate at primary site/neck in each treatment arm will be estimated by dividing the number of patients with pCR recorded at any visit from randomization until PD per modified RECIST v1.1 or death due to any cause by the number of randomized patients who had salvage surgery at the primary site/neck dissection, respectively, in each treatment arm. The corresponding exact 2-sided 95% CIs will be provided by treatment arm.

6.2.2.8. Objective response

Best overall response (BOR) will be assessed based on reported overall lesion responses at different evaluation time points from the date of randomization until the first documentation of PD, according to the following rules.

BOR Based on Unconfirmed Responses

- uCR = one CR documented before first documentation of PD
- uPR = one PR documented before first documentation of PD (and not qualifying for a CR)
- SD (applicable only to patients with measurable disease at baseline) = at least one SD assessment (or better) ≥ 6 weeks after date of randomization and before first documentation of PD (and not qualifying for uCR or uPR).
- Non-uCR/Non-PD (applicable only to patients with non-measurable disease at baseline) = at least one non-uCR/non-PD assessment (or better) ≥ 6 weeks after date of randomization and before first documentation of PD (and not qualifying for uCR or uPR).
- PD = PD after date of randomization (and not qualifying for uCR, uPR or SD).
- NE: all other cases.

An objective status of uPR or SD cannot follow one of uCR. SD can follow uPR only in the rare case that tumor increases by less than 20% from the nadir, but enough that a previously documented 30% decrease from baseline no longer holds.

Unconfirmed Objective Response (uOR) is defined as BOR of uCR or uPR. A patient will be considered to have achieved a uOR if the patient has a uCR or uPR which does not need to be confirmed at a subsequent assessment.

Patients who do not have a post-baseline radiographic tumor assessment or clinical evaluation due to early progression, who die, progress, or drop out for any reason prior to reaching a uCR or uPR will be counted as non-responders in the assessment of uOR. Each patient will have an objective response status (0: no uOR; 1: uOR). uOR rate (uORR) is the proportion of patients with uOR in the analysis set.

uORR by treatment arm will be calculated along with the 2-sided 95% CI using the Clopper-Pearson[7] method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option).

In addition, the frequency (number and percentage) of patients with uBOR of uCR, uPR, SD, non-uCR/non-PD (applicable only to patients with non-measurable disease at baseline), PD and NE will be tabulated. Patients with uBOR of NE will be summarized by reason for having NE status. The following reasons will be used:

- No baseline assessment
- No post-baseline assessments due to death

- No post-baseline assessments due to other reasons
- All post-baseline assessments have overall response NE
- SD of insufficient duration (<6 weeks after date of randomization)

The association of treatment effect and uOR will be tested by the General Association Statistic of the Cochran-Mantel-Haenszel test (CMH) with the randomization strata taken into account. The null hypothesis of no association in any of the randomization strata is tested against the alternative, which specifies that there is an association between treatment effect and tumor response at least in one randomization stratum. The CMH test will be performed at 1-sided nominal alpha level of 0.025.

The stratified odds ratio in terms of uOR will also be estimated along with its 95% CI to compare study treatments. The odds ratio is defined as the odds of uOR with experimental treatment divided by the odds of uOR with control treatment. The Breslow-Day test will be used to check the homogeneity of the odds ratio across the randomization strata. It tests the null hypothesis that odds ratios in all strata are equal against the alternative hypothesis that at least in one stratum the odds ratio is different.

In case the null hypothesis of homogeneity of odds ratios across strata is not rejected at the 2-sided alpha level of 0.05, the common odds ratio will be determined using the Mantel-Haenszel estimate (by the FREQ procedure using CMH option in SAS); if the null hypothesis of homogeneity of odds ratio across all strata is rejected, the odds ratio per stratum will be calculated with the corresponding exact CI.

6.2.2.9. Duration of response

Duration of Response (DR) is defined, for patients with uOR, as the time from first documentation of objective response (uCR or uPR) to the date of first documentation of PD or death due to any cause. If a patient has not had an event (PD or death), DR is censored at the date of last adequate tumor or clinical assessment. The censoring rules for DR are described in Table 20.

DR (months) = [date of event or censoring– first date of uOR +1]/30.4375

Table 20. Outcome and Event Dates for DR Analyses

Scenario	Date of event/censoring	Outcome
PD or death	Date of PD or death	Event
No PD and no death	Date of last adequate tumor assessment or clinical evaluation documenting no PD	Censored

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median DR time with 2-sided 95% CIs. In particular, the DR rates at 12, 24, 36, 48 and 60 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to

Brookmeyer and Crowley (1982)[4] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[12] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

DR will be displayed graphically and analyzed using Kaplan-Meier methodology. If the number of patients with uOR is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided.

Frequency (number and percentage) of patients with each event type (PD or death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 21 following the hierarchy shown.

Table 21. DR Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Patient refused further follow-up]	Withdrawal of consent
2	No event and lost to follow-up in any disposition page	Lost to follow-up
3	No event and [EOS present OR disposition page for any epoch after screening says patient will not continue into any subsequent phase of the study] and no adequate post-baseline assessment	No adequate post-baseline assessment
4	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

The DR time or censoring time and the reasons for censoring will also be presented in a patient listing.

6.2.2.10. Time to response

Time to response (TTR) is defined, for patients with uOR, as the time from the date of randomization to the first documentation of objective response (uCR or uPR).

TTR (in months) = [first date of uOR – date of randomization +1]/30.4375

TTR will be summarized using simple descriptive statistics (mean, SD, median, min, max. Q1, Q3).

6.2.3. PRO endpoints

The NCCN FHNSI-22 was developed using methods consistent with the Food and Drug Administration (FDA) PRO guidance (FDA, 2009[9]) and surveyed input from patients with advanced cancers and physician experts (Cella D, et al., 2011[6]). The FHNSI-22 questionnaire is specifically designed to be a stand-alone instrument to measure disease symptoms, treatment side effects and overall quality of life in patients with head and neck cancer.

Responses on the FHNSI-22 questionnaire are used to calculate a total score and scores for four subscales: TSE, DRS-P, DRS-E, and FWB. The questionnaire contains 22 items with 5-point Likert scales ranging from 'not at all' to 'very much'. Higher scores mean better symptomatology, quality of life or functioning. The expected questionnaire completion time is about 5 minutes.

The EuroQol EQ-5D-5L is a patient-completed questionnaire designed to assess health status in terms of a single index value or utility score (Herdman M, et al., 2011[10]). There are 2 components to the EuroQol EQ-5D-5L: a descriptive system in which individuals rate their level of problems (none, slight, moderate, severe, extreme/unable) in 5 areas (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and a Visual Analogue Scale (VAS) in which patients rate their overall health status from 0 (worst imaginable) to 100 (best imaginable). Published weights are available that allow for the creation of a single summary score for the 5 Likert scale items. The expected questionnaire completion time is about 3 minutes or less.

All PRO summaries/analyses will be based on the FAS, with two analysis periods:

- Analysis including PRO data through EOT (including EOT)
- Additional analysis including all PRO data (including data collected after EOT)

6.2.3.1. Instrument completion rate

For each treatment arm and at each time point, the number and percentage of patients who complete the FHNSI-22 and EQ-5D-5L will be summarized, as will the reasons for non-completion of these measures.

6.2.3.2. Descriptive summaries over time

Only scheduled assessments will be included in the summaries that follow.

For each treatment arm and at each assessment time point, summary statistics of the FHNSI-22 total and subscale scores, the EQ-5D-5L, and VAS at each time point will be reported as described in Section 5.2.3.

For continuous endpoints (the total and subscale scores of the FHNSI-22, EQ-5D-5L, and VAS), change from baseline will be reported with N (number of non-missing values), mean, standard error, and 95% confidence intervals. Only patients with both a baseline score and a post-baseline score will be included in the summary of change from baseline scores.

Line charts depicting the means and mean changes of subscales together with standard error bars over time will be also provided.

6.2.3.3. Time-to-event endpoints

There are three separate time-to-event endpoints based on time to deterioration (TTD) of

1) HRQol;

2) DRS-P;

3) 'swallowing'.

TTD is defined as the time from randomization to deterioration where deterioration is defined, respectively, as

1) a 3 point decrease on the FWB scale for health-related quality of life (HRQoL);

2) a 3 point decrease on the DRS-P for symptoms of disease;

3) a 1 point decrease, in two consecutive assessments, on the 'swallowing' item of the FHNSI-22.

The time to event for HRQoL and DRS-P will be calculated using the date of randomization and the date of the event, respectively. The time to event for "swallowing" will be calculated using the date of randomization and the date of the first of the two consecutive assessments used to identify the event. Note that a decline in FWB subscale of the FHNSI-22 represents a worsening in HRQol. A decline in DRS-P score represents an increase in physical symptoms of disease. A decline in the 'swallowing' item score represents a decrease in the ability to swallow naturally and easily.

Patients who do not have a TTD event will be censored on the date when they last completed a PRO assessment.

TTD will be compared between treatment arms using a log-rank test stratified by randomization stratification factors. The treatment effect will be estimated using a Cox Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie, for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=Arm B, 1= Arm A) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median TTD time with 2-sided 95% CIs. In particular, the TTD rates at 12, 24, 36, 48, and 60 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[4] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[12] (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

6.2.3.4. Binary and categorical endpoint

Responder is a patient with improvement in swallowing, defined as a 1 point increase on the 5-point Likert scale maintained for 2 consecutive assessments, in the single 'swallowing' item of the FHNSI-22, month 6 and 12 during the Maintenance Phase and also at EOT. In accordance with published recommendations (Osoba D, et al., 2005[15]; Brundage M, et al.,

2007[5]; Sloan JA, et al., 2007[19]), proportions will be calculated including all patients with baseline questionnaires and counting those patients with missing questionnaires at a subsequent time point as not having improved.

Responder on the swallowing item of the FHNSI-22 will be compared between treatment arms using the Cochran-Mantel-Haenszel test stratified by randomization stratification factors. In this analysis, patients with missing questionnaires will be assigned as non-responder.

A sensitivity analysis will also be conducted. Patients whose post baseline questionnaires were missing because of disease progression or death will be considered as non-responder, with other patients whose post baseline questionnaires were missing being excluded from the analysis.

6.2.3.5. Continuous endpoints

Longitudinal random intercept random slope mixed-effect model will be carried out for the total and subscale scores of the FHNSI-22, EQ-5D-5L, and VAS using PROC MIXED. Outcomes are PRO post-baseline scores and the predictors are the corresponding baseline PRO score, treatment, randomization stratification factor, time (treated as a continuous variable), and treatment-by-time interaction. Intercept and time are considered as random effects particular to each patient. All available data for each patient should be used in the analyses. All parameter estimates should be obtained using restricted maximum likelihood. The unstructured covariance structure should be used to define covariance between random effects (using option "Type=UN" as a part of the RANDOM statement in PROC MIXED). For the degrees-of-freedom calculations the Kenward and Roger algorithm should be used (using option "ddfm = kr" as a part of the MODEL statement in PROC MIXED).

6.2.4. Pharmacokinetic endpoints

The following pharmacokinetic analyses will be based on the PK analyses set by treatment arm.

6.2.4.1. Avelumab

C_{trough} and C_{max} for avelumab will be summarized descriptively (n, mean, SD, CV, median, minimum, maximum, geometric mean, its associated CV, and 95% CI) by treatment phase, cycle (if applicable), and day in Arm A. Other standard parameters will be calculated, including but not limited to single and multiple dose T_{max}, $T_{ss,max}$, as data permit.

C_{trough} and Cmax for avelumab will be plotted for each dose using a box whisker plot by treatment phase, cycle (if applicable), and day within in order to assess the attainment of steady state in Arm A.

Additionally, avelumab PK will be evaluated following cisplatin dosing by comparing the overall geometric mean ratios of C_{max} and C_{trough} in only Arm A on Day 8 of the CRT Phase to Day 1 of the Lead-in Phase.

Pharmacokinetic parameters for avelumab will be taken from observed values or derived from plasma concentration-time data as described in Section 3.2.4.

6.2.4.2. Cisplatin (Total and Free)

PK parameters for cisplatin (total and free) will be estimated using noncompartmental and/or compartment methods, if needed. Analyses will include C_{max}, time to T_{max}, AUC_{inf}, t½, CL, and Vz as data permit. Dose-normalized parameters [eg, C_{max} (dn), AUC_{inf} (dn)] will be reported as appropriate. Descriptive statistics for the PK parameters for cisplatin will be provided in tabular form. Cisplatin plasma concentrations will be summarized descriptively (n, mean, SD, CV, median, minimum, maximum, geometric mean, its associated CV, and 95% CI) by nominal time. Individual patient and median profiles will be presented on both linear-linear and log-linear scales.

Additionally, cisplatin PK (total and free) will be evaluated following avelumab or placebo dosing by comparing the overall geometric mean ratios of C_{max} and AUC_{inf} on Day 1 of the CRT Phase in Arm A to Day 1 of the CRT Phase in Arm B.

Pharmacokinetic parameters for cisplatin (total and free) will be taken from observed values or derived from plasma concentration-time data as described in Section 3.2.4.

6.2.5. Population pharmacokinetic endpoints

Pharmacokinetic and pharmacodynamic data from this study may be analyzed using modeling approaches and may also be pooled with data from other studies to investigate any association between avelumab exposure and biomarkers or significant safety/efficacy endpoints. The results of these analyses, if performed, may be reported separately.

In addition, the relationship between exposure and efficacy and safety endpoints may be explored, as necessary, based on emerging efficacy and safety data. The analysis details will be provided in the Population Modeling Analysis Plan. The results of these modeling analyses may be reported separately from the clinical study report. The analyses will include patients in both the FAS and the safety analysis set.

6.2.6. Biomarker endpoints

There are two sets of biomarkers, those collected at screening (pre-treatment tumor sample), and the optional tumor sample after one dose of avelumab. The optional tumor samples will be analyzed separately.

Summary statistics for pre-treatment biomarkers will be presented by treatment arm and overall. Summary statistics for on-treatment biomarkers will additionally include the ratio to baseline for continuous biomarkers, and a contingency table for non-continuous biomarkers, as appropriate.

Continuous biomarkers will be split at the median ($<$ and $\geq$ median), and will be included as a covariate in an exploratory Kaplan-Meier analysis for PFS, as described in Section 6.1.1.1; no other covariates will be included. Non-continuous biomarkers will either be split at an appropriate point in their distribution, and classified as above and below the split (0,1), or the actual categories will be included in the survival analysis.

6.2.7. Endpoints for immunogenicity data of avelumab

All analyses described below are performed for the treatment arm containing avelumab (Arm A) unless otherwise specified.

Blood samples for avelumab immunogenicity testing will be collected at pre-dose on Day 1 of the Lead-in Phase, and Days 8, and 25 of the CRT Phase. All samples should be drawn within 2 hours before the start of avelumab or placebo infusion.

Samples positive for ADA will be analyzed for titer and may be analyzed for nAb.

Patients will be characterized into different ADA categories based on the criteria defined in Table 22.

Table 22. Patients Characterized Based on Anti-Drug Antibody Results (ADA Status)

Category	Definition	Patients at Risk (Denominator for Incidence)
ADA never-positive	No positive ADA results at any time point; ADA-negative patients (titer < cutpoint)	Number of patients with at least one valid ADA result at any time point
ADA ever-positive	At least one positive ADA result at any time point; ADA-positive patients (titer ≥ cutpoint)	Number of patients with at least one valid ADA result at any time point
Baseline ADA positive	A positive ADA result at baseline	Number of patients with valid baseline ADA result
Treatment-boosted ADA	A positive ADA result at baseline and the titer ≥ 8×baseline titer at least once after treatment with avelumab	Number of patients with valid baseline ADA results and at least one valid post-baseline ADA result
Treatment-induced ADA	Patient is ADA-negative at baseline and has at least one positive post-baseline ADA result; or if patient does not have a baseline sample, the patient has at least one positive post-baseline ADA result	Number of patients with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)
Transient ADA response	If patients with treatment-induced ADA have (a single positive ADA result or duration between first and last positive result <16 weeks) and ADA result at the last assessment is not positive.	Number of patients with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)
Persistent ADA response	If patients with treatment-induced ADA have duration between first and last positive ADA result ≥16 weeks or a positive ADA result at the last assessment	Number of patients with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)

ADA: anti-drug antibody, NR = not reportable.

Patients will be characterized into different nAb categories based on the criteria in Table 23. For nAb, treatment-boosted is not applicable since no titer result is available.

Table 23. Patients Characterized Based on Neutralizing Antibody Results (nAb Status)

Category	Definition	Patients at Risk (Denominator for Incidence)
nAb never-positive	No positive nAb results at any time point	Number of patients with at least one valid ADA result at any time point
nAb ever-positive	At least one positive nAb result at any time point	Number of patients with at least one valid ADA result at any time point
Baseline nAb positive	A positive nAb result at baseline	Number of patients with valid baseline ADA result
Treatment-induced nAb	Patient is not nAb positive at baseline and has at least one positive post-baseline nAb result; or if patient does not have a baseline sample, the patient has at least one positive post-baseline ADA result	Number of patients with at least one valid post-baseline ADA result and without positive baseline nAb result (including missing, NR)
Transient nAb response	If patients with treatment-induced nAb have (a single positive nAb result or duration between first and last positive result <16 weeks) and nAb result at the last assessment is not positive.	Number of patients with at least one ADA valid post-baseline result and without positive baseline nAb result (including missing, NR)
Persistent nAb response	If patients with treatment-induced nAb have duration between first and last positive nAb result ≥16 weeks or a positive nAb result at the last assessment	Number of patients with at least one valid post-baseline ADA result and without positive baseline nAb result (including missing, NR)

ADA = antidrug antibody, nAb = neutralizing antibody, NR = no result.

The number and percentage of patients in each ADA and nAb category will be summarized.

6.2.7.1. Time to and Duration of ADA and nAb response

The ADA and nAb analyses described below will include patients with treatment-induced ADA or nAb, respectively.

Time (weeks) to ADA response is defined as:

(Date of first positive ADA result – date of first dose of avelumab + 1)/7.

Time to ADA response will be summarized using simple descriptive statistics (mean, SD, median, min, max. Q1, Q3).

Duration (weeks) of ADA response is defined as:

(Date of last positive ADA result – date of first positive ADA result + 1)/7.

Duration of ADA response will be censored if:

- the last ADA assessment is positive AND patient is continuing treatment with avelumab, or

- the last ADA assessment is positive AND patient discontinued treatment with avelumab AND the last planned ADA assessment (Day 25 of the CRT Phase) is after the cut-off date.

Time to nAb response and duration of nAb response are defined similarly based on first and last positive nAb result.

Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including the median ADA response time with 2-sided 95% CIs. ADA response rates at different timepoints will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[4] and the CIs for the survival function estimates will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[12] (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Duration of ADA response will be displayed graphically and analyzed using Kaplan-Meier methodology. If the number of patients with ADA response is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided

As data permit, the analyses described above will be repeated for patients with treatment-induced nAb.

6.2.7.2. ADA titer

For patients who are ADA ever positive, the maximum observed ADA titer for a patient will be summarized, overall and by ADA subcategories (baseline ADA positive, treatment-boosted ADA, treatment-induced ADA, transient ADA response, persistent ADA response) of patients having each discrete maximum titer value will be tabulated. The denominator to calculate the percentages will be the total number of patients in the associated ADA subcategory.

For patients with treatment-induced ADA, a cross tabulation of duration of ADA response and maximum ADA titer will be provided. The following categories for duration of ADA response will be used: ≤1, >1 to ≤3, >3 to ≤5, >5 to ≤7, >7 to ≤13, >13 to ≤16, >16 to ≤25, >25 weeks. In this categorization, the censoring in duration of ADA response is ignored.

6.2.7.3. Analysis of PK and safety by immunogenicity status

The following ADA and nAb status will be used for the analyses described below.

ADA

- ADA ever-positive versus ADA never-positive;
- ADA: treatment-induced ADA versus ADA never-positive or baseline ADA positive.

nAb

- nAb ever-positive versus nAb never-positive;
- nAb: treatment-induced nAb versus nAb never-positive or baseline nAb positive.

Data listings will include immunogenicity data together with relevant PK, safety and efficacy data.

PK parameters and immunogenicity status

The following analyses will include patients in both the immunogenicity analysis set and in the PK parameter analysis set. The PK endpoints pertinent to the immunogenicity analyses are C_{trough} and C_{max}.

Blood samples for avelumab PK will be collected pre-dose (within 2 hours before the start of avelumab/placebo infusion) and at the end of infusion (immediately before until 10 minutes after the end of avelumab/placebo infusion) on Days 1 of the Lead-in Phase and Days 8, and 25 of the CRT Phase.

C_{trough} and C_{max} will be summarized descriptively (n, mean, SD, CV, median, minimum, maximum, geometric mean, its associated CV, and 95% CI) by nominal time and ADA status. Linear-linear and log-linear plots of mean and median for C_{trough} and C_{max} over nominal time and by ADA status will be presented.

Among patients with treatment-induced ADA, analyses will be conducted to assess whether C_{trough} and C_{max} have any changes before and after the first positive ADA assessment. To be included in this analysis, patients must have the same PK parameter available both before and after the first positive ADA assessment. Relative PK day will be calculated as:

(PK assessment nominal day) – (first positive ADA assessment nominal day).

Nominal day is the protocol scheduled timing for an assessment. For example, if C_{trough} is collected on Day 8 of CRT Phase and the first positive ADA result is observed on Day 25 of CRT Phase, then the relative PK day for this C_{trough} is -17. Linear-linear and log-linear plots of mean and median for C_{trough} and C_{max} over relative PK day will be presented.

As data permit, the analyses described above will be repeated for nAb.

Safety and immunogenicity status

The following analyses will include patients in the immunogenicity analysis set.

The frequency (number and percentage) of patients with each of the following will be presented by ADA status.

- TEAEs, by SOC and PT;
- TEAEs leading to dose reduction of avelumab, by SOC and PT;

- TEAEs leading to discontinuation of avelumab, by SOC and PT;
- TEAEs leading to discontinuation of study treatment by SOC and PT;
- Grade ≥ 3 TEAEs, by SOC and PT;
- SAEs, by SOC and PT;
- IRRs associated with avelumab, by PT.

For patients who had at least one IRR and have treatment-induced ADA, time related to first onset of an IRR (infusion 1, infusion 2, infusion 3, infusion 4 or later) will be summarized taking into account whether the IRR occurred on or after the first ADA positive assessment or whether the IRR occurred before the first ADA positive assessment.

As data permit, the analyses described above will be repeated for nAb.

6.3. Other Endpoints

CCI [redacted]

CCI [redacted]

6.4. Subset Analyses

Subset analyses will be performed for PFS, OS, uOR and DR based on the FAS for the subgroups defined below.

The following subgroups will be defined and used for analyses:

- Randomization stratification factors
  - Tumor (T) stage (<T4 vs T4; T4=Reference);
  - Nodal (N) stage (N0/N1/N2a/N2b vs N2c/N3; N2c/N3=Reference);
  - HPV status (positive vs negative; negative=Reference) as measured by p16 expression by IHC.

- Age
  - Age < 65 years (Reference);
  - Age ≥ 65 years.
- Gender
  - Male (Reference);
  - Female.
- Race
  - White (Reference);
  - Asian;
  - Black or African American;
  - Other.
- Ethnicity
  - Hispanic or Latino;
  - Not Hispanic or Latino (Reference).
- Pooled Geographical Region
  - North America;
  - Western Europe (Reference);
  - Eastern Europe;
  - Asia;
  - Rest of the World (Australasia, Latin America, Africa and/or Middle East will be included as additional subgroups if including > 10% of the overall randomized population).
- ECOG Performance Status at Baseline
  - 0 (Reference);
  - 1.
- Subsite of Primary Diagnosis
  - Oral cavity (Reference);
  - Oropharynx;
  - Larynx;
  - Hypopharynx;
  - Other.

Subset analyses for PFS, OS and DR will use the primary censoring rules described in Sections 6.1.1.1 and 6.2.2.1. All the subgroup analyses are exploratory. Treatment arms will be compared for PFS and OS using a 2-sided unstratified log-rank test for each subgroup

level and the unstratified HR and its corresponding 95% CI will be computed per subgroup level.

All the subgroup analyses will be exploratory; no adjustment for multiplicity will be performed. In the case of a low number of patients within a category (<5% of the randomized population), the categories will be pooled.

To assess the heterogeneity of treatment effects for PFS and OS across the subgroup levels, two Cox regression model will be fitted with PFS or OS, respectively, as the dependent variable and subgroup, treatment, and with and without the treatment-by-subgroup interaction as explanatory variables.

- Model 1: factors + treatment + subgroup
- Model 2: factors + treatment + subgroup + treatment×subgroup-variable

A p-value for the interaction test (Wald's test) will be provided together with the HR and corresponding 95% CI for the interaction model parameter.

The HR for PFS and OS and corresponding 95% CIs for all subgroups will also be presented in a forest plot.

The ORR odds ratio for each subgroup and corresponding 95% CIs will also be presented in a forest plot for each treatment arm.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline summaries

The following analyses will be based on the FAS overall and separately by treatment arm.

6.5.1.1. Demographic characteristics

Demographic characteristics and physical measurements will be summarized by treatment arm using the following information from the 'Screening/Baseline Visit' eCRF pages.

- Demographic characteristics
  - Gender: Male, Female
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Unknown
  - Ethnic origin:
    - Hispanic or Latino
    - Not Hispanic or Latino
  - Age (years): summary statistics

  - Age categories:
    - < 65 years, ≥ 65 years;
    - < 65, 65-<75, 75-<85, ≥ 85 years.
  - Pooled Geographical Region (as applicable):
    - North America;
    - Western Europe;
    - Eastern Europe;
    - Asia;
    - Rest of the World (Australasia, Latin America, Africa and/or Middle East will be included as additional pooled geographical regions if including > 10% of the overall randomized population).
  - Geographic Region (as applicable):
    - North America;
    - Latin America;
    - Western Europe;
    - Eastern Europe;
    - Middle East;
    - Australasia;
    - Asia;
    - Africa.
  - Eastern Cooperative Oncology Group (ECOG) Performance Status: 0, 1, 2, 3, and 4
- Physical measurements
  - Height (cm);
  - Weight (kg);
  - Body Mass Index (BMI) (kg/m^2);
  - Body Surface Area (BSA) (m^2).

Center codes will be used for the determination of the patient's geographic region.

The listing of demographics and baseline characteristics will include the following information: patient identifier, treatment arm, age, sex, race, ethnicity, height (cm), weight (kg), BMI (kg/m^2), BSA (m^2) and ECOG performance status.

6.5.1.2. Medical history

Medical history will be coded using the most current available version of Medical Dictionary for Regulatory Activities (MedDRA) and will be summarized from the 'Medical History' eCRF page. Medical history will be summarized as the numbers and percentages of patients

by MedDRA preferred term (PT) as event category and MedDRA primary system organ class (SOC) as summary category. Each patient will be counted only once within each PT or SOC.

Medical history will be displayed in terms of frequency tables: ordered by primary SOC and PT in alphabetical order.

6.5.1.3. Disease characteristics

Information on disease characteristics collected on 'Primary Diagnosis', 'Substance Use' and modified RECIST eCRF pages will be summarized overall and by treatment arm. Summary statistics will be presented for the following.

From the 'Primary Diagnosis' eCRF page:

- Site of primary tumor
- Primary diagnosis (summarize all categories collected in the 'Primary Diagnosis' eCRF page)
- Time since initial diagnosis to date of randomization (months), defined as (date of randomization – date of initial diagnosis)/30.4375

From the modified RECIST eCRF page:

- Measurable disease (lesions) at baseline (Yes, No)
- Involved tumor sites at baseline

From the 'Substance Use' eCRF page:

- Smoking history (never smoker, former smoker, current smoker)

Listing of disease history will be provided with all relevant data (as collected on the 'Primary Diagnosis' and 'Substance Use' eCRF pages) and derived variables as above.

6.5.2. Study conduct and patient disposition

The following analyses will be performed based on the FAS overall and separately by treatment arm.

6.5.2.1. Patient disposition

The percentages below will be calculated based on the number of patients in the FAS in each treatment arm.

- Total number of patients screened overall;
- Number of patients who discontinued from the study prior to randomization overall and by the main reason for discontinuation;
- Number and percentage of randomized patients in each of the analysis sets defined in Section 4;

- Number and percentage of randomized patients with study drug ongoing during the Lead-in Phase;
- Number and percentage of randomized patients who completed or discontinued study drug overall and by the main reason for discontinuation of study drug during the Lead-in Phase;
- Number and percentage of randomized patients who entered CRT Phase;
- Number and percentage of randomized patients with study drug ongoing (separately for each study drug) during the CRT Phase;
- Number and percentage of randomized patients who completed or discontinued study drug overall and by the main reason for discontinuation of study drug (separately for each study drug) during the CRT Phase;
- Number and percentage of randomized patients with at least one study drug ongoing during the CRT Phase;
- Number and percentage of randomized patients who completed or discontinued all study drugs, discontinued at least one study drug, and completed all study drugs during the CRT Phase;
- Number and percentage of randomized patients who entered the Maintenance Phase;
- Number and percentage of randomized patients with study drug ongoing during the Maintenance Phase;
- Number and percentage of randomized patients who completed or discontinued study drug overall and by the main reason for discontinuation of study drug during the Maintenance Phase;
- Number and percentage of patients who entered follow-up;
- Number and percentage of patients who completed or discontinued follow-up overall and by the main reason for discontinuation;
- Number and percentage of patients who entered long-term follow-up;
- Number and percentage of patients who discontinued long-term follow-up overall and by the main reason for discontinuation.

The results of the randomization algorithm (according to IRT) will be summarized as follows:

- Number and percentage of randomized patients overall, by region (Europe, EEA (required by EudraCT), North America, Latin America, Asia, Australasia, Middle East, Africa), by country within region;
- Number and percentage of randomized patients by center;
- Number and percentage of randomized patients by randomization strata (IRT);
- Number and percentage of randomized patients by randomization strata (eCRF);
- Cross tabulation: stratum by IRT vs. stratum by eCRF;

- Cross tabulation: patients randomized (Arm A/Arm B/none) vs. patients treated (Arm A/Arm B/none).

6.5.2.2. Protocol deviations

All protocol violations that impact the safety of the patients and/or the conduct of the study and/or its evaluation will be reported. These include:

- Patients who are dosed on the study despite not satisfying the inclusion criteria;
- Patients who develop withdrawal criteria whilst on the study but are not withdrawn;
- Patients who receive the wrong treatment or an incorrect dose;
- Patients who receive an excluded concomitant medication;
- Deviations from GCP.

The identification of these and other CSR-reportable deviations will be based on the inclusion/exclusion criteria or other criteria presented in the protocol.

6.5.3. Study treatment compliance and exposure

The following analyses will be based on the safety analysis set by treatment arm.

The derivations below are provided for the following:

- Avelumab or matching placebo administered as a 1-hour IV infusion at a dose of 10 mg/kg
  - on Day 1 of the Lead-in Phase (1 week prior to the start of the CRT Phase), and
  - on Days 8, 25, and 39 during the CRT Phase, and
  - once every 2 weeks in 4-week cycles for 12 months (52 weeks) during the Maintenance Phase.
- Cisplatin 100 mg/m^2 administered in 500 mL normal saline over a 60 – 120 minute infusion with an additional 1 to 1.5 L of fluid give post-hydration, on Days 1, 22, 43 during the CRT Phase
- IMRT administered 70 Gy/35 fractions/7weeks, 1 fraction/day, 5 fractions/week)for 7 weeks during the CRT Phase.

Analysis of exposure for avelumab (or matching placebo) and cisplatin will be based on the calculated actual dose levels:

- Avelumab – total dose / weight
- Cisplatin - total dose/m^2

Analysis of exposure for IMRT will be based on the total dose.

6.5.3.1. Exposure to avelumab or matching placebo

Exposure to avelumab or matching placebo will be summarized by treatment phase (Lead-in, CRT, Maintenance). In what follows, references to avelumab are meant to apply also to matching placebo.

For Cycle X during the Maintenance Phase, actual cycle start date for each patient is

- the earliest start date of dosing in the Cycle X Day 1 visit eCRF exposure page, if the patient received study treatment on that visit (ie, avelumab dose>0 at that visit)
- the first day of assessments in the Cycle X Day 1 visit, if the patient did not receive study treatment on that visit (ie, avelumab dose=0 at that visit).

Actual cycle end date for each patient is,

- for all cycles X except the last cycle, actual cycle end date = actual cycle (X+1) start date – 1 day;
- for the last cycle, actual cycle end date = actual cycle start date + 28 days – 1 day

Cycle duration (weeks) = (actual cycle end date – actual cycle start date + 1)/7

The dose level for avelumab is calculated as actual dose administered/weight (mg/kg). The last available weight of the patient on or prior to the day of dosing will be used.

Lead-in Phase

One dose of avelumab is planned to be administered in the Lead-in Phase, the number and percentage of patients who receive the dose will be summarized.

CRT Phase

During the CRT Phase, three doses of avelumab are planned to be administered (on Days 8, 25, and 39 of CRT Phase) with a minimum of 14 days required to separate consecutive doses of avelumab or matching-placebo.

Intended duration of avelumab treatment in CRT Phase (days) = ((39+14-1)-8+1) = 45

Duration of exposure to avelumab in the CRT Phase (days) =

(last dose date of avelumab – first dose date of avelumab + 14)

Cumulative avelumab dose (mg/kg) is the sum of the actual doses of avelumab received in the CRT Phase.

Actual Dose Intensity (DI) for avelumab in the CRT Phase =

[cumulative dose in the CRT Phase (mg/kg)]/[intended duration of avelumab treatment in CRT Phase (days)/45] (mg/kg/45 days)

Relative Dose Intensity (RDI) for avelumab in the CRT Phase is calculated as follows:

- Intended DI (mg/kg/45 days) = 30 (mg/kg/45 days)
- RDI (%) = 100 × [actual DI] / [intended DI] = 100 × actual DI / 30.

The summary of treatment exposure and compliance for avelumab in the CRT Phase will include the following information:

- Duration of exposure to avelumab (days);
- Total number of infusions received;
- Cumulative dose (mg/kg);
- Actual DI (mg/kg/45 days);
- RDI (%).

Maintenance Phase

Each cycle for avelumab during the Maintenance Phase is defined by a 4-week period. The dose intensity (DI) and the relative dose intensity (RDI) for avelumab during the Maintenance Phase will be calculated for each patient across all cycles.

Intended duration of avelumab treatment in the Maintenance Phase (weeks) =

(end date–date of first dose of study drug +1)/7,

where end date = start date of last cycle with non-zero dose of study drug + 28 – 1

Duration of exposure to avelumab in the Maintenance Phase (weeks) =

(last dose date of avelumab – first dose date of avelumab + 14)/7

Cumulative dose (mg/kg) is the sum of the actual doses of avelumab received during the Maintenance Phase.

Actual DI for avelumab in the Maintenance Phase

- Overall actual DI (mg/kg/4-week cycle) = [overall cumulative dose (mg/kg)] / [intended duration of treatment with avelumab (weeks)/4].

RDI for avelumab in the Maintenance Phase

- Intended DI (mg/kg/4-week cycle) = [intended cumulative dose per cycle] / [intended number of 4-weeks in a cycle] = [20 (mg/kg)] / [1 (4-week cycle)] = 20 (mg/kg/4-week cycle)
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI] = 100 × [overall actual DI] / [20 (mg/kg/4-week cycle)]

The summary of treatment exposure and compliance for avelumab during the Maintenance Phase will include the following information:

- Duration of exposure to avelumab (weeks);
- Total number of infusions received;
- Cumulative dose (mg/kg);
- Actual DI (mg/kg/cycle);
- RDI (%).

Duration of exposure to avelumab across all treatment phases

Duration of exposure to avelumab (weeks) = (last dose date of avelumab – first dose date of avelumab + 14) /7

Total number of avelumab infusions received overall is the sum of the total number of infusions received in each treatment phase.

6.5.3.2. Exposure to Cisplatin

Cisplatin is to be administered only in the CRT Phase. Three doses of cisplatin (100 mg/m^2) are planned to be administered on Days 1, 22, and 43 of the CRT Phase.

Intended duration of cisplatin treatment in CRT Phase (weeks) = 63/7 = 9 weeks

Duration of exposure to cisplatin in the CRT Phase (weeks) =

(last dose date of cisplatin – first dose date of cisplatin + 21)/7

Cumulative cisplatin dose (mg/m^2) is the sum of the actual doses of cisplatin received in the CRT Phase.

Actual Dose Intensity (DI) for cisplatin in the CRT Phase =

[cumulative dose (mg/m^2)]/[intended duration of cisplatin treatment (weeks) /9] (mg/m^2/9-weeks)

Relative Dose Intensity (RDI) for cisplatin in the CRT Phase is calculated as follows:

- Intended DI (mg/m^2/9-weeks) = 300 (mg/m^2/9-weeks)
- RDI (%) = 100 × [actual DI] / [intended DI] = 100 × actual DI / 300.

The summary of treatment exposure and compliance for cisplatin will include the following information:

- Duration of exposure to cisplatin (weeks);

- Total number of doses [a] received;
- Cumulative dose (mg/m^2);
- Actual DI (mg/m^2/9-weeks);
- RDI (%).

[a] Per protocol, cisplatin dose can be split over a period of 2 to 4 consecutive days. Therefore, if cisplatin is given consecutively over 2-4 days, this will be viewed as a split dose. Split doses will be summed up and categorized as a single dose.

6.5.3.3. Exposure to IMRT

IMRT is to be administered only in the CRT Phase. The total dose and duration of IMRT administration will be calculated for each patient during the CRT Phase.

Duration of IMRT (days) = number of days IMRT is administered.

Total dose of IMRT (Gy) = duration of IMRT (days) × 2.

The number of days IMRT is administered and compliance with IMRT administration ("Appropriate", "Variation acceptable", "Deviation unacceptable") will be assessed by a central vendor for each patient during the CRT Phase.

The summary of treatment exposure and compliance for IMRT will include the following information:

- Duration of IMRT (days);
- Total dose of IMRT (Gy);
- Overall compliance (Appropriate, Variation acceptable, or Deviation unacceptable).

6.5.3.4. Dose reductions

Applicable to avelumab, matching placebo and cisplatin.

For avelumab or matching placebo, dose reduction is defined as actual non-zero dose < 90% of the planned dose. The number and percentage of patients with at least one dose reduction as well as a breakdown of the number of dose reductions (1, 2, 3, ≥4) will be summarized overall and by treatment phase (Lead-in, CRT, Maintenance).

For cisplatin, a dose reduction is defined as a reduction to dose level -1 (75 mg/m^2) or dose level -2 (50 mg/m^2). The number and percentage of patients with at least one one-dose level reduction, and with at least one two-dose level reduction as well as a breakdown of the number of dose reductions will be summarized.

6.5.3.5. Dose Delays

Applicable to avelumab, matching placebo, and cisplatin.

Dose Delay is the difference between the actual time between two consecutive non-zero doses and the planned time between the same two consecutive non-zero doses.

Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) – Planned days between two consecutive doses.

- Lead-in Phase: not applicable
- CRT Phase:
  - cisplatin: the number of planned days between two consecutive doses is 21 days;
  - avelumab or matching placebo: not applicable (note that the protocol mandates a minimum interval between doses but does not establish a fixed number of days between two consecutive doses);
- Maintenance Phase (avelumab or matching placebo only): the number of planned days between two consecutive doses is 14 days.

Dose delays will be grouped into the following categories:

- No delay;
- CRT Phase: 1-2 days delay; 3-6 days delay;
- Maintenance Phase: 1-3 days delay; 4-6 days delay;
- 7 or more days delay.

In the CRT Phase, no delay and 1-2 days delay will also be summarized together, and in the Maintenance Phase no delay and 1-3 days delay will also be summarized together.

The number and percentage of patients with delayed study drug administration and maximum length of delay, ie, the worst case of delay if patients have multiple dose delays will be summarized

- for avelumab or matching placebo: in the Maintenance Phase;
- for cisplatin: in the CRT Phase.

6.5.3.6. Infusion rate reductions

Applicable to avelumab, matching placebo and cisplatin.

The number and percentage of patients with at least one infusion rate reduction of ≥50% compared to the first infusion rate reported in the eCRF as well as the frequency of patients with 1, 2, 3, or ≥4 infusion rate reductions of ≥ 50% will be summarized

- for avelumab or matching placebo: overall and by treatment phase (Lead-in, CRT, Maintenance);

- for cisplatin: in the CRT Phase.

6.5.3.7. Infusion interruptions

Applicable to avelumab, matching placebo and cisplatin.

An infusion interruption is defined as an infusion that is stopped and re-started on the same day (ie, for a visit more than one infusion start time and infusion end time are recorded).

The number and percentage of patients with at least one infusion interruption as well as the frequency of patients with 1, 2, 3, or ≥4 infusion interruptions will be summarized

- for avelumab or matching placebo: overall and by treatment phase (Lead-in, CRT, Maintenance);
- for cisplatin: in the CRT Phase.

6.5.4. Concomitant medications and non-drug treatments

The following analyses will be based on the safety analysis set by treatment arm.

Concomitant medications are medications, other than study drugs, which started prior to first dose date of study treatment and continued during the on-treatment period as well as those started during the on-treatment period. **Prior medications** are medications, other than study drugs and pre-medications for study drug, which are started before the first dose of study treatment.

Prior and concomitant medications will be summarized from the 'General Concomitant Medications' eCRF page. Pre-medications for study drug will also be summarized separately from the 'Pre-Medication Treatment' eCRF page.

Summary of prior medications, summary of concomitant medications and summary of pre-medications will include the number and percentage of patients by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term. A patient will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. If any prior or concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes. The summary tables will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. In case any specific medication does not have ATC classification level 2 coded term, it will be summarized under 'Unavailable ATC classification' category.

A listing of prior medications and a listing of concomitant medications will be created with the relevant information collected on the 'General Concomitant Medications' eCRF page. A listing of pre-medications will be created with the relevant information collected on the 'Pre-Medication Treatment' eCRF page.

All concurrent procedures, which were undertaken any time during the on-treatment period, will be listed according to the eCRF page 'General Non-drug Treatments'.

A listing of concurrent procedures will be created with the relevant information collected on the 'General Non-drug Treatments' eCRF page.

6.5.5. Subsequent anti-cancer therapies

The following analyses will be based on the FAS by treatment arm.

Anti-cancer treatment will be provided in a data listing with data retrieved from 'Follow-up Cancer Therapy', 'Follow-up Radiation Therapy', 'Concomitant Surgery', and 'Follow-up Surgery' eCRF pages.

Number and percentage of patients with any anti-cancer therapy after discontinuation will be tabulated overall and by type of therapy based on the data collected from the 'Follow-up Cancer Therapy', 'Follow-up Radiation Therapy' and 'Follow-up Surgery' eCRF pages.

6.6. Safety Summaries and Analyses

The Safety Analysis Set will be the primary population for safety evaluations. Summaries of AEs and other safety parameters will be based on the safety analysis set by treatment arm.

6.6.1. Adverse events

Treatment-emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period as defined in Section 3.5.1.

All analyses described will be based on TEAEs (started during the on-treatment period) if not otherwise specified. The AE listings will include all AEs (whether treatment-emergent or not). AEs with onset outside the on-treatment period will be flagged in the listings.

- **Related Adverse Events:** adverse events with relationship to study treatment (as recorded on the AE eCRF page, Relationship with study treatment = Related) reported by the investigator and those of unknown relationship (ie, no answer to the question 'Relationship with study treatment'). Related AEs are those related to any study drug (ie, at least one of the study drugs).
- **Serious Adverse Events (SAE):** serious adverse events (as recorded on the AE eCRF page, Serious Adverse Event = Yes).
- **Adverse Events Leading to Dose Reduction:** adverse events leading to dose reduction of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Dose reduced).
- **Adverse Events Leading to Interruption of Study Treatment:** adverse events leading to interruption of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug interrupted). The eCRF does not allow for a clear separation between interruption of an infusion and delays of administration for a parenteral drug as both are recorded using the same term on the eCRF ("Drug interrupted"). IRRs will be excluded in the analysis of AEs leading to Drug Interruption in case they only led to an interruption of the infusion.

- **Adverse Events Leading to Permanent Treatment Discontinuation:** adverse events leading to permanent discontinuation of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug withdrawn).
- **Adverse Events Leading to Death:** adverse event leading to death (as recorded on the AE eCRF page, Outcome = Fatal, as well as AEs of Grade 5).
- **Immune-related Adverse Events (irAE):** irAEs (as identified according to the methodology outlined in Appendix 1 for a pre-specified search list of MedDRA PTs documented in the SRP and finalized for analysis of this study's data prior to DB lock)
- **Infusion-related Reactions (IRR)**: IRRs (as identified according to the methodology outlined in Appendix 2 for a pre-specified search list of MedDRA PTs documented in the SRP and finalized for analysis of this study's data prior to DB lock).

Unless otherwise specified, AEs will be summarized by number and percentage of patients with the AE in the category of interest as described above, by treatment arm, primary SOC and PT in decreasing frequency based on the frequencies observed for Arm A.

Each patient will be counted only once within each SOC or PT. If a patient experiences more than one AE within a SOC or PT for the same summary period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity.

6.6.1.1. All adverse events

Adverse events will be summarized by worst severity (according to NCI-CTCAE version 4.03) per patient, using the latest version of MedDRA preferred term (PT) as event category and MedDRA primary system organ class (SOC) body term as Body System category.

In case a patient has events with missing and non-missing grades, the maximum of the non-missing grades will be displayed. No imputation of missing grades will be performed.

The following tables will be created:

- The overall summary of AEs table will include the frequency (number and percentage) of patients with each of the following by treatment arm:
  - TEAEs;
  - TEAEs, Grade ≥ 3;
  - Related TEAEs;
  - Related TEAEs, Grade ≥ 3;
  - TEAEs leading to dose reduction of avelumab;
  - TEAEs leading to dose reduction of cisplatin;
  - TEAEs leading to dose reduction of IMRT;
  - TEAEs leading to interruption of avelumab;

  - TEAEs leading to interruption of cisplatin;
  - TEAEs leading to interruption of IMRT;
  - TEAEs leading to discontinuation of avelumab;
  - TEAEs leading to discontinuation of cisplatin;
  - TEAEs leading to discontinuation of IMRT;
  - TEAEs leading to discontinuation of any study drug (CRT Phase);
  - TEAEs leading to discontinuation of all study drugs (CRT Phase);
  - Related TEAEs leading to discontinuation of avelumab;
  - Related TEAEs leading to discontinuation of cisplatin;
  - Related TEAEs leading to discontinuation of IMRT;
  - Related TEAEs leading to discontinuation of any study drug (CRT Phase);
  - Related TEAEs leading to discontinuation of all study drugs (CRT Phase);
  - Serious TEAEs;
  - Related Serious TEAEs;
  - TEAEs leading to death;
  - Related TEAEs leading to death;
  - irAEs;
  - IRRs.
- TEAEs by SOC and PT and worst grade
- Related TEAEs by SOC and PT and worst grade
- TEAEs leading to death by SOC and PT
- Related TEAEs leading to death by SOC and PT

In addition, the following tables will be created by treatment phase:

- Overall summary of AEs table will include the frequency (number and percentage) of patients with each of the following by treatment arm and treatment phase [a] :
  - TEAEs;
  - TEAEs, Grade ≥ 3;
  - TEAEs leading to discontinuation of avelumab;
  - Serious TEAEs;
  - irAEs;
  - IRRs.
- TEAEs by SOC and PT and worst grade.

- TEAEs leading to death by SOC and PT.

[a] AEs will be included in the summary for each treatment phase if the AE onset occurs anytime from first dose of study treatment in the treatment phase through minimum(first dose of study treatment in next treatment phase, end of on-treatment period).

6.6.1.2. Adverse events leading to dose reduction

The frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to dose reduction of each study drug by treatment arm:

- TEAEs leading to dose reduction of avelumab by SOC and PT;
- TEAEs leading to dose reduction of cisplatin by SOC and PT;
- TEAEs leading to dose reduction of IMRT by SOC and PT.

The listing of all AEs leading to dose reduction will also be provided with the relevant information.

6.6.1.3. Adverse events leading to interruption of study treatment

The eCRF does not allow for a clear separation between interruption of an infusion and delays of administration for a parenteral drug as both are recorded using the same term on the eCRF ("Drug interrupted"). IRRs will be excluded in the analysis of AEs leading to Drug Interruption in case they only led to an interruption of the infusion (ie, did not lead to a dose reduction or a dose delay).

As such, AEs leading to interruption will be defined as AEs identified in the AE eCRF page with an action taken with study treatment of 'drug interrupted' excluding:

- IRRs that occurred on the day of infusion with ≥90% of the planned dose given (ie, IRRs that did not lead to a dose reduction) and subsequent administration of study drug had no delay (as defined in Section 6.5.3.5). These IRRs will be considered as IRRs leading to interruption of infusion.
- IRRs occurring on the day after infusion and subsequent dose administration had no delay (as defined in Section 6.5.3.5).

The frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to interruption of each study drug by treatment arm:

- TEAEs leading to interruption of avelumab by SOC and PT;
- TEAEs leading to interruption of cisplatin by SOC and PT;
- TEAEs leading to interruption of IMRT by SOC and PT.

The listing of all AEs leading to interruption of study treatment will also be provided with the relevant information.

In addition, the frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to interruption of each study drug by treatment arm:

- TEAEs leading to both interruption and dose reduction of avelumab by SOC and PT;
- TEAEs leading to both interruption and dose reduction of cisplatin by SOC and PT;
- TEAEs leading to both interruption and dose reduction of IMRT by SOC and PT.

This summary will take into account PTs with both actions as defined in Section 6.6.1, eventhough the actions may be captured for different PT records (ie, different onset for the PT with action "drug interrupted" and the PT with action "dose reduced".

6.6.1.4. Adverse events leading to discontinuation of study treatment

The frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to permanent discontinuation of each study drug and study treatment, by treatment arm:

- TEAEs leading to discontinuation of avelumab by SOC and PT;
- Related TEAEs leading to discontinuation of avelumab by SOC and PT;
- TEAEs leading to discontinuation of cisplatin by SOC and PT;
- Related TEAEs leading to discontinuation of cisplatin by SOC and PT;
- TEAEs leading to discontinuation of IMRT by SOC and PT;
- Related TEAEs leading to discontinuation of IMRT by SOC and PT;
- TEAEs leading to discontinuation of any study drug by SOC and PT (CRT Phase);
- Related TEAEs leading to discontinuation of any study drug by SOC and PT (CRT Phase);
- TEAEs leading to discontinuation of all study drugs by SOC and PT (CRT Phase);
- Related TEAEs leading to discontinuation of all study drugs by SOC and PT (CRT Phase).

The listing of all AEs leading to treatment discontinuation will also be provided with the relevant information.

In addition, the frequency (number and percentage) of patients with TEAEs leading to discontinuation of avelumab will be summarized by SOC and PT within each treatment arm and treatment phase.

6.6.2. Deaths

The frequency (number and percentage) of patients in the safety analysis set who died and who died within 30 days after last dose of study treatment as well as the reason for death, will be tabulated based on information from the 'Notice of Death' and 'Survival Follow-Up' eCRFs, by treatment arm, and by treatment arm and treatment phase.

- All deaths
- Deaths within 30 days after last dose of study treatment
- Reason for Death
  - Disease progression;

- Study treatment toxicity;
- AE not related to study treatment;
- Unknown;
- Other.

In addition, date and cause of death will be provided in individual patient data listing together with selected dosing information (study treatment received, date of first / last administration, dose) and will include the following information:

- AEs with fatal outcome (list preferred terms of AEs with outcome=Fatal, as well as AEs of Grade 5);
- Flag for death within 30 days of last dose of study treatment.

6.6.3. Serious adverse events

The frequency (number and percentage) of patients with each of the following will be presented for treatment-emergent SAEs by treatment arm:

- SAEs by SOC and PT;
- Related SAEs by SOC and PT.

In addition, the frequency (number and percentage) of patients with SAEs will be summarized by SOC and PT within each treatment arm and treatment phase.

The listings of all SAEs will also be provided with the relevant information with a flag for SAEs with onset outside of the on-treatment period.

6.6.4. Other significant adverse events

The frequency (number and percentage) of patients with each of the following will be presented for irAEs, by treatment arm:

- irAEs leading to death, by Cluster and PT;
- irAEs, by Cluster and PT;
- irAEs, Grade ≥ 3, by Cluster and PT;
- irAEs leading to discontinuation of avelumab, by Cluster and PT;
- irAEs leading to discontinuation of cisplatin, by Cluster and PT;
- irAEs leading to discontinuation of IMRT, by Cluster and PT;
- irAEs leading to discontinuation of any study drug, by Cluster and PT (CRT Phase);
- irAEs leading to discontinuation of all study drugs, by Cluster and PT (CRT Phase);
- Serious irAEs, by Cluster and PT.

In addition, the frequency (number and percentage) of patients with each of the following will be presented for irAEs by treatment arm and treatment phase:

- irAEs, by Cluster;
- irAEs, Grade ≥ 3, by Cluster;
- irAEs leading to death, by Cluster;
- irAEs leading to discontinuation of avelumab, by Cluster;
- Serious irAEs, by Cluster.

The listing of all irAEs will also be provided with the relevant information with a flag for irAEs with onset outside of the on-treatment period.

The frequency (number and percentage) of patients with each of the following will be presented for IRRs, by treatment arm:

- IRRs leading to death, by PT (separately for IRRs associated with avelumab, associated with cisplatin, associated with either avelumab or cisplatin);
- IRRs, by PT (separately for IRRs associated with avelumab, associated with cisplatin, associated with either avelumab or cisplatin);
- IRRs, Grade ≥ 3, by PT (separately for IRRs associated with avelumab, associated with cisplatin, associated with either avelumab or cisplatin);
- IRRs leading to discontinuation of avelumab, by PT;
- IRRs leading to discontinuation of cisplatin, by PT;
- IRRs leading to discontinuation of any study drug, by PT (CRT Phase);
- IRRs leading to discontinuation of all study drugs, by PT (CRT Phase);
- Serious IRRs, by PT (separately for IRRs associated with avelumab, associated with cisplatin, associated with either avelumab or cisplatin);
- Time related to first onset of an IRR associated with avelumab (infusion 1, infusion 2, infusion 3, infusion 4 or later).

In addition, the frequency (number and percentage) of patients with each of the following will be presented for IRRs by treatment arm and treatment phase

- IRRs associated with avelumab, by PT
- IRRs associated with avelumab, Grade ≥ 3, by PT
- IRRs associated with avelumab leading to death, by PT
- IRRs leading to discontinuation of avelumab, by PT;
- Serious IRRs associated with avelumab, by PT;
- Time related to first onset of an IRR associated with avelumab.

The listing of all IRRs will also be provided with the relevant information with a flag for IRRs with onset outside of the on-treatment period.

In addition, the following analyses will be presented for Tier-1 (irAEs and IRRs as defined in Appendix 1 and Appendix 2, respectively) and Tier-2 events separately. CIs for risk difference will be calculated based on the unconditional exact method by Santner and Snell (1980)[17]. No additional analyses will be presented for Tier-3 AEs.

- Frequency (number and percentage) of patients with each of the following by treatment arm and PT or Clustered Term:
  - Tier-2 AEs
  - Tier-2 AEs Grade ≥ 3
- Point estimate for risk difference and 95% CI for risk difference (experimental arm vs comparator arm) for each of the following by PT or Clustered Term:
  - irAEs
  - irAEs Grade ≥ 3
  - IRRs associated with avelumab
  - IRRs associated with avelumab, Grade ≥ 3
  - Tier-2 AEs
  - Tier-2 AEs Grade ≥ 3
- The CIs reported are not adjusted for multiplicity and should be used for screening purposes only. The 95% CIs are provided to help gauge the precision of the estimates for the risk difference and should be used for estimation purposes only.

6.6.5. Laboratory data

6.6.5.1. Hematology and chemistry parameters

Laboratory results will be classified according to the NCI-CTCAE criteria version 4.03. Non-numerical qualifiers (with the exception of fasting flags) will not be taken into consideration in the derivation of CTCAE criteria (eg, hypokalemia Grade 1 and Grade 2 are only distinguished by a non-numerical qualifier and therefore Grade 2 will not be derived). Additional laboratory results that are not part of NCI-CTCAE will be presented according to the categories: below normal limit, within normal limits and above normal limit (according to the laboratory normal ranges).

Quantitative data will be summarized using simple descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each nominal visit over time (unscheduled measurements would therefore not be included in these summaries as described in Section 5.2.9). End of Treatment visit laboratory results will be summarized separately. The changes computed will be the differences from baseline. Qualitative data based on reference ranges will be described according to the categories (ie, Low, Normal, High).

Abnormalities classified according to NCI-CTCAE toxicity grading version 4.03 will be described using the worst grade. For those parameters which are graded with two toxicities such as potassium (hypokalemia/hyperkalemia), the toxicities will be summarized separately. Low direction toxicity (eg, hypokalemia) grades at baseline and post baseline will be set to 0 when the variables are derived for summarizing high direction toxicity (eg, hyperkalemia), and vice versa.

For **WBC differential counts** (total neutrophil [including bands], lymphocyte, monocyte, eosinophil, and basophil counts), the absolute value will be used when reported. When only percentages are available (this is mainly important for neutrophils and lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

Derived differential absolute count = (WBC count) × (Differential %value / 100)

If the range for the differential absolute count is not available (only range for value in % is available) then Grade 1 will be attributed to as follows:

- Lymphocyte count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 800/mm3
- Neutrophil count decreased
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 1500/mm3

For **calcium**, CTCAE grading is based on Corrected Calcium and Ionized Calcium (CALCIO). Corrected Calcium is calculated from Albumin and Calcium as follows

Corrected calcium (mmol/L) = measured total Calcium (mmol/L) + 0.02 (40 - serum albumin [g/L])

Liver function tests: Alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total bilirubin (TBILI) are used to assess possible drug induced liver toxicity. The ratios of test result over upper limit of normal (ULN) will be calculated and classified for these three parameters during the on-treatment period.

Summary of liver function tests will include the following categories. The number and percentage of patients with each of the following during the on-treatment period will be summarized by treatment arm:

- ALT ≥ 3×ULN, ALT ≥ 5×ULN, ALT ≥ 10×ULN, ALT ≥ 20×ULN;
- AST ≥ 3×ULN, AST ≥ 5×ULN, AST ≥ 10×ULN, AST ≥ 20×ULN;

- (ALT or AST) ≥ 3×ULN, (ALT or AST) ≥ 5×ULN, (ALT or AST) ≥ 10×ULN, (ALT or AST) ≥ 20×ULN;
- TBILI ≥ 2×ULN;
- Concurrent ALT ≥ 3×ULN and TBILI ≥ 2×ULN;
- Concurrent AST ≥ 3×ULN and TBILI ≥ 2×ULN;
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN;
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and ALP > 2×ULN;
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and (ALP ≤ 2×ULN or missing).

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, ie, a patient with an elevation of AST ≥10×ULN will also appear in the categories ≥5×ULN and ≥3×ULN. Liver function elevation and possible Hy's Law cases will be summarized using frequency counts and percentages.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created, with different symbols for different treatment arms, by graphically displaying

- peak serum ALT(/ULN) vs peak total bilirubin (/ULN) including reference lines at ALT=3×ULN and total bilirubin=2×ULN;
- peak serum AST(/ULN) vs peak total bilirubin (/ULN) including reference lines at AST=3×ULN and total bilirubin=2×ULN.

In addition, a listing of all TBILI, ALT, AST and ALP values for patients with concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and (ALP ≤ 2×ULN or missing) will be provided.

Parameters with NCI-CTC grades available:

The laboratory toxicities will be tabulated using descriptive statistics (number of patients and percentages) during the on-treatment period. The denominator to calculate percentages for each laboratory parameter is the number of patients evaluable for CTCAE grading (ie those patients for whom a Grade 0, 1, 2, 3 or 4 can be derived).

- The summary of laboratory parameters by CTCAE grade table will include number and percentage of patients with Grade 1, 2, 3, 4, Grade 3/4 and any grade (Grades 1-4), laboratory abnormalities during the on-treatment period.
- The shift table will summarize baseline CTCAE grade versus the worst on-treatment CTCAE grade. The highest CTCAE grade during the on-treatment period is considered as the worst grade for the summary.
- The number and percentage of patients with newly occurring or worsening laboratory abnormalities during the on-treatment period will be summarized by worst grade on-treatment (Grade 1, 2, 3, 4, Grade 3/4 and any grade (Grades 1-4)).

The above analyses apply to hematology and chemistry evaluations which can be graded per CTCAE, ie:

- Hematology:

  Hemoglobin (HB), Leukocytes (white blood cell decreased), Lymphocytes (lymphocyte count increased/decreased), Neutrophils / Absolute Neutrophils Count (ANC) (neutrophil count decreased), Platelet Count (PLT) (platelet count decreased).

- Serum Chemistry:

  Albumin (hypoalbuminemia), Alkaline Phosphatase (alkaline phosphatase increased), Alanine Aminotransferase (ALT) (ALT increased), Amylase (serum amylase increased), Aspartate Aminotransferase (AST) (AST increased), Total Bilirubin (blood bilirubin increased, Cholesterol (cholesterol high), Creatinine (creatinine increased), Creatine Kinase (CPK increased), Potassium (hypokalemia/ hyperkalemia), Sodium (hyponatremia/ hypernatremia), Magnesium (hypomagnesemia/hypermagnesemia), Calcium (hypocalcemia/ hypercalcemia), Glucose (hypoglycemia/hyperglycemia), Gamma Glutamyl Transferase (GGT) (GGT increased), Lipase (lipase increased), Phosphates (hypophosphatemia), Triglycerides (hypertriglyceridemia)

In addition, hematology and chemistry laboratory parameters by CTCAE grade (any grade and Grade ≥ 3) will also be summarized by treatment phase. Laboratory tests will be included in the summary for each treatment phase if the laboratory test date/time occurs anytime from first dose date/time of study treatment in the treatment phase through minimum(first dose date/time of study treatment in next treatment phase, end of on-treatment period).

Parameters with NCI-CTC grades not available:

Hematology and chemistry evaluations which cannot be graded per CTCAE criteria will be summarized as frequency (number and percentage) of patients with:

- shifts from baseline normal to at least one result above normal during on-treatment period
- shifts from baseline normal to at least one result below normal during on-treatment period

In this study, these apply to the following parameters:

- Hematology: Absolute Monocytes, Absolute Eosinophils, Absolute Basophils
- Serum Chemistry: Chloride, Total Urea, Uric Acid, Total Protein, C-Reactive Protein, Lactate Dehydrogenase (LDH)

6.6.5.2. Other laboratory parameters

All other parameters collected on the eCRF will be listed in dedicated listings presenting all corresponding collected information on the eCRF.

- Coagulation: activated partial thromboplastin time (aPTT) and prothrombin time (INR).
- Urinalysis: protein, glucose, blood
- Thyroid function tests: TSH, free T4

- Other parameters: adrenocorticotropic hormone (ACTH)
- Pregnancy test

The listings of laboratory results will be provided for all laboratory parameters. The listings will be sorted by parameters and assessment dates or visits for each patient. Laboratory values that are outside the normal range will also be flagged in the data listings, along with corresponding normal ranges. A listing of CTCAE grading will also be generated for those laboratory tests.

In addition, listings of abnormal values will be provided for hematology, chemistry, urinalysis, coagulation parameters. If there is at least one abnormal assessment for any parameter, all the data for that laboratory parameter will be included into the listing.

For all tests not mentioned above but present in the clinical data, a listing of patients with at least one result for the relevant test will be provided.

6.6.6. Vital signs

Weight for the purposes of dose calculation will be recorded at screening and within 3 days pre-dose Day 1 of each cycle. Weight will not be collected at End of Treatment. Height will be measured at screening only.

Vital sign summaries will include all vital sign assessments from the on-treatment period. All vital sign assessments will be listed, and those collected outside the on-treatment period will be flagged in the listing.

All vital sign parameters will be summarized using descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each visit over time. End of Treatment visit will be summarized separately. The changes computed will be the differences from baseline.

6.6.7. Electrocardiogram

ECG summaries will include all ECG assessments from the on-treatment period. All ECG assessments will be listed, and those collected outside the on-treatment period will be flagged in the listing. QTcB and QTcF will be derived based on RR and QT (see below). When triplicates are collected (for approximately 30 subjects in each treatment arm), the average of the replicate measurements should be determined after the derivation of the individual parameter at each time point.

Selecting Primary QT Correction for Heart Rate

The analysis of QT data is complicated by the fact that the QT interval is highly correlated with heart rate. Because of this correlation, formulas are routinely used to obtain a corrected value, denoted QTc, which is independent of heart rate. This QTc interval is intended to represent the QT interval at a standardized heart rate. Several correction formulas have been proposed in the literature. For this analysis we will use some of those methods of correction,

as described below. The QT interval corrected for heart rate by the Bazett's formula, QTcB, is defined as

$$QTcB = \frac{QT}{\sqrt{RR}},$$

the QT interval corrected for heart rate by the Fridericia's formula, QTcF, is defined as

$$QTcF = \frac{QT}{\sqrt[3]{RR}},$$

where RR represents the RR interval of the ECG, in seconds, and can be estimated as 60/Heart Rate.

Although Bazett's correction is the historical standard, it does not perform well when heart rate fluctuates. Fridericia's formula may perform better under these conditions.

ECG Summaries

The following analyses will be performed for each applicable ECG parameters (RR, PR, QRS, QT, ventricular rate –denoted as HR in what follows-, and QTc) by treatment arm, during the on-treatment period. The denominator to calculate percentages for each category is the number of patients evaluable for the category.

- Pearson correlation between QT and HR, QTc (QTcB, QTcF) and HR using individual (non-averaged) baseline assessments
- For each of the ECG parameters (HR, and QT, QTc, QRS, PR intervals), descriptive statistics at baseline, at each post-baseline time point and changes from baseline at each post-baseline time point
- Frequency (number and percentage) of patients with notable ECG values according to the following categories:
  - QT/QTc increase from baseline >30 ms, >60 ms;
  - QT/QTc > 450 ms, > 480 ms, > 500 ms;
  - HR ≤ 50 bpm and decrease from baseline ≥ 20 bpm;
  - HR ≥ 120 bpm and increase from baseline ≥ 20 bpm;
  - PR ≥ 220 ms and increase from baseline ≥ 20 ms;
  - QRS ≥ 120 ms.

Patients with notable ECG interval values and qualitative ECG abnormalities will be listed for each patient and time point and the corresponding notable values and abnormality findings will be included in the listings.

Unscheduled ECG measurements will not be used in computing the descriptive statistics for change from baseline at each post-baseline time point. However, they will be used in the analysis of notable ECG changes and the shift table analysis of notable QT parameters.

6.6.8. ECOG performance status

The ECOG shift from baseline to highest score during the on-treatment period will be summarized by treatment arm. ECOG performance status with shift from ECOG=0 or 1 to ECOG 2 or higher will also be presented in a data listing.

7. INTERIM ANALYSES

7.1. Introduction

The goals of the interim analyses are to allow early stopping of the study for futility or to do an early determination of superiority of Arm A compared to Arm B based on PFS. The interim analysis of PFS and OS will be performed as described in Sections 5.1.1 and 5.1.2 using the methodology described in Section 6.1.1 for PFS and Section 6.2.2.1 for OS.

At the time of the interim analysis for PFS, the interim analysis for OS will be performed by an independent statistician. Unblinded results from the interim analysis for PFS and first interim analysis for OS will not be communicated to the Sponsor's clinical team or to any party involved in the study conduct (apart from the independent statistician and E-DMC members) until the E-DMC has determined that either (i) PFS analysis has crossed the pre-specified boundary for efficacy or (ii) the study needs to be terminated due to any cause, including futility or safety reasons. Further details will be described in the E-DMC charter.

At the time of final PFS analyses, both PFS and interim OS analysis will be performed by the Sponsor's clinical team. All patients will continue to be followed for OS until the final OS analysis (or earlier if OS reaches statistical significance at an interim analysis).

7.2. Interim Analyses and Summaries

At each analysis time point, the critical boundaries for the group sequential test will be derived from the predefined spending function(s) as described in Section 5.1. The calculations of boundaries will be performed using EAST.

7.2.1. Interim analysis for PFS

Let $u(t_1)$ and $u(t_F)$ denote the upper critical boundaries based on the test statistics Z_1 and Z_F for efficacy at the interim and the final analysis, respectively, and let $l(t_1)$ and $l(t_F)$ denote the lower critical boundary for futility at the interim and final analysis, respectively. For the final analysis, $l(t_F)=u(t_F)$.

The critical values $u(t_1)$ and $l(t_1)$ for the interim analysis of PFS are determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1) \quad \text{and} \quad P_a(Z_1 \leq l(t_1)) = \beta(t_1),$$

where P_0 and P_a denote the probabilities under the null hypothesis and the alternative hypothesis, respectively, and $\alpha(t_1)$ and $\beta(t_1)$ denote the α and β spent, respectively, based on

the predefined spending functions at information fraction t_1 (t_1 is calculated as the ratio of the number of PFS events observed at the time of the cut-off for the interim analysis and the total number of PFS events targeted for the final analysis).

The boundary for the final efficacy analysis will be calculated such that

$$\alpha(t_1) + P_0(Z_1 < u(t_1), Z_F \geq u_F) = 0.025$$

As described in Section 5.1.2, if the number of PFS events in the final analysis deviates from the target number of PFS events, the final analysis criteria will be determined as above taking into account the actual alpha spent at the interim analysis and the actual correlation between the two test statistics Z_1 and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at 0.025.

7.2.2. Interim analysis for OS

No futility analysis will be performed for OS. Let $u(t_i)$ and $u(t_F)$ denote the upper critical boundaries based on the test statistics Z_i and Z_F for efficacy at the i^{th} interim and the final analysis, respectively, where i=1, 2, 3, 4.

In what follows P_0 denotes the probability under the null hypothesis, and $\alpha(t_i)$ denotes the α spent based on the predefined α-spending function at information fraction t_i (t_i is calculated as the ratio of the number of OS events observed at the time of the cut-off for the i^{th} interim analysis and the total number of OS events targeted for the final analysis.

For each comparison, the critical value $u(t_1)$ for the 1^{st} interim analysis of OS is determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1).$$

Critical boundaries for the additional interim analyses and the final analysis of OS are calculated recursively as follows for each comparison

u(t2) is derived such that $\alpha(t_1) + P_0(Z_1 < u(t_1), Z_2 \geq u(t_2)) = \alpha(t_2)$,

u(t3) is derived such that $\alpha(t_2) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_3 \geq u(t_3)) = \alpha(t_3)$,

u(t4) is derived such that $\alpha(t_3) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_3 < u(t_3), Z_4 \geq u(t_4)) = \alpha(t_4)$.

The boundary for the final efficacy analysis is derived such that

$$\alpha(t_4) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_3 < u(t_3), Z_4 < u(t_4), Z_F \geq u_F) = 0.025.$$

As described in Section 5.1.2, if the number of OS events in the final analysis deviates from the target number of OS events, the final analysis criteria will be determined as above taking into account the actual alpha spent at the interim analyses and the actual correlation between the four test statistics Z_i (i=1, 2, 3, 4) and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at 0.025.

8. REFERENCES

1. Amit O, et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis and recommendation from a PhRMA working group. European Journal of Cancer 47:1772-1778, 2011.

2. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363 (1);24-35.

3. Ang KK, Zhang Q, Rosenthal, DI, et al. Randomized phase III trial of concurrent accelerated radiation plus cisplatin with or without cetuximab for stage III and IV head and neck carcinoma: RTOG 0522. J Clin Oncol 2014;32: 2940-2950.

4. Brookmeyer R, Crowley JJ. A confidence interval for the median survival time. Biometrics. 38: 29-41, 1982.

5. Brundage M, Osoba D, Bezjak A, et al. Lessons learned in the assessment of health-related quality of life: Selected examples from the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 2007, 25:5078-5081.

6. Cella D, Rosenbloom SK, Beaumont JL, et al. Development and validation of 11 symptom indexes to evaluate response to chemotherapy. JNCCN 2011;9: 268-278.

7. Clopper CJ, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika; 26, 404-413.

8. EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. Version 2.1, April 2015. Available at: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf

9. FDA. Guidance for Industry: Patient-reported outcomes measures; use in medical product development to support labeling claims. December 2009. Available at: https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf.

10. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20: 1727-1736.

11. Jennison C, Turnbull BW. In: Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, 2000.

12. Kalbfleisch JD, Prentice, RL. Statistical Analysis of Failure Time Data, 2nd Edition. Hoboken, Wiley Interscience. 2002.

13. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 53: 457-81, 1958.

14. NCCN/FACT Head & Neck Symptom Index-22 (FHNSI-22) Scoring Guidelines (Version 2). Available at: http://www.facit.org/facitorg/questionnaires.

15. Osoba D, Bezjak A, Brundage M, et al. Analysis and interpretation of health-related quality-oflife data from clinical trials: Basic approach of The National Cancer Institute of Canada Clinical Trials Group. Eur J Cancer 2005, 41:280-287.

16. Royston, P. and Parmar, M.K. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in medicine*, 2011; *30*(19), pp.2409-2421.

17. Santner TJ, Snell MK. (1980). Small-sample confidence intervals for p1-p2 and p1/p2 in 2 x 2 contingency tables. *Journal of the American Statistical Association*, 75:386–394

18. Schoenfeld D. Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 1980; 67:145-53.

19. Sloan JA, Dueck AC, Erickson PA, et al. Analysis and interpretation of results based on patient reported outcomes. Value Health 2007, 10:S106-S115 (suppl 2).

20. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health, 2000.

21. Uno H, et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. J Clin Oncol 2014; 32.

22. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 Bio-Algorithms and Med-Systems 2013; 9(4): 183–9.

9. APPENDICES

Appendix 1. Immune-Related Adverse Events

The MedDRA PTs and clusters for irAEs are defined in the SRP for avelumab.

Immune-related AEs (irAEs) will be programmatically identified as outlined in Table 24. This case definition is hierarchical, ie, each step is only checked for patients and events that have already met the prior step.

Table 24. Case Definition for irAEs

Step	Selection Criteria	Additional Notes
1	Event selected based on a list of pre-specified MedDRA PTs within clusters. These are included in the SRP as Tier1 events (Immune-mediated xxxx). If AE matches the list, then it is in for the next step.	
2	AE onset during 1st study drug administration or anytime thereafter through 90 days after last dose of study treatment.	This is regardless of start of new anti-cancer drug therapy and regardless of TEAE classifications
3	Answer in the AE eCRF page to 'Was another treatment given because of the occurrence of the event' is 'YES'	
4	AE treated with corticosteroids or other immunosuppressant therapy. For endocrinopathies only: AE required hormone replacement	Look in the conmed pages for AE identifiers that match the AEs from Step 3. For each of such AEs if A) OR B) OR C) below are met then the AE is in for the next step A) conmed ATC code is in (H02A, H02B, D07, A01AC, S01BA, S01BB, L04AA, L04AB, L04AC, L04AD, L04AX, A07EA) and AE PT is in any of the irAE clusters. B) conmed ATC code is in (H03A, H03B) and AE PT is in one of the irAE clusters associated with "Immune-mediated endocrinopathies" C) conmed ATC code is A10A and AE PT is in the irAE cluster associated with "Immune-mediated endocrinopathies: Type I Diabetes Mellitus"

5	A) No clear etiology (other than immune mediated etiology)	A) From the AE eCRF page. Is the AE clearly related to an etiology other than immune-mediated etiology? Yes / No If answer is Yes, check all that apply: • Underlying malignancy / progressive disease. • Other medical conditions. • Prior or concomitant medications / procedures. • Other. Specify.
	B) Histopathology / biopsy consistent with immune-mediated event Event is in if [Answer to 5B1 and 5B2 is YES (regardless of answer to 5A)] OR [Answer to 5B1 is YES AND answer to 5B2 is NO AND answer to 5A is NO] OR [Answer to 5B1 is NO AND answer to 5A is NO]	B) From the AE eCRF page. B1) Was there a pathology /histology evaluation performed to investigate the AE? Y/N B2) If answer to the above is Yes, does the pathology/histology evaluation confirms an immune mediated mechanism for the AE? Y/N B3) If pathology / histology evaluation performed to investigate the AE, provide summary of relevant findings of the pathology /histology report. (Free Text)

The data set associated with irAEs may be refined based on medical review. The final data set including any changes based on medical review (eg, addition of cases that are not selected by the programmatic algorithm) will be the basis of the irAE analyses.

Appendix 2. Infusion Related Reactions

For defining an AE as IRR the onset of the event in relation to the infusion of study drug and time to resolution of the event will be considered.

- All AEs identified by the MedDRA PT query describing signs and symptoms will be considered potential IRRs when onset is on the day of study drug infusion (during or after infusion) and the event resolved with end date within 2 days after onset.
- All AEs identified by the MedDRA PTs of Infusion related reaction, Drug hypersensitivity, Anaphylactic reaction, Hypersensitivity, Type 1 hypersensitivity, will be considered potential IRRs when onset is on the day of study drug infusion (during or after the infusion) or the day after the study drug infusion (irrespective of resolution date).

The list of MedDRA PTs for 'IRRs SIGNS and SYMPTOMS' and PTs 'IRRs CORE' are defined in the SRP for avelumab.

Infusion-related reactions (IRRs) will be programmatically identified as outlined in Table 25 or Table 26 and will be identified for IV drugs only.

Table 25. Case Definition for IRRs – IV Study Drugs Administered Alone Or In Combination With Non-IV Study Drugs

Condition	Selection criterion
If AE meets [1 AND 2] OR [3 AND (4A OR 4B)] then AE is classified as an IRR	
1	PT is included in the 'IRRs SIGNS and SYMPTOMS' list
2	• AE onset date = date of infusion of study drug AND • AE timing related to study drug ('DURING', 'AFTER') AND • AE outcome in ('RECOVERED/RESOLVED', 'RECOVERED/RESOLVED WITH SEQUELAE', 'RECOVERING/RESOLVING') AND • AE end date – AE onset date <=2
3	PT is included in the 'IRRs CORE' list
4A	• AE onset date = date of infusion of study drug AND • AE timing related to study drug in ('DURING', 'AFTER')
4B	AE onset on the day after infusion

Table 26. Case Definition for IRRs – IV Study Drugs Administered in Combination (eg, Doublets or Triplets)

Condition	Selection criterion
IRR can be associated with the first IV drug and/or subsequent IV drugs that are administered in combination. Without loss of generality assume triplet IV with D_1 administered first then D_2 then D_3. The IV study drug or drugs associated with the IRR need to be identified in the analysis data set to enable subsequent analysis. The following are not sequential and an AE can be classified as an IRR associated with multiple D_J from one or more of I, II, III, IV, V below: I - If the AE meets [1 AND 2A1] for a D_J then the AE is classified as an IRR associated with the D_J that meets the 2A1 criterion II - If the AE meets [1 AND 2A2] for a D_J then the AE is classified as an IRR associated with the D_J and associated with D_{J+1} that meets the 2A2 criterion III - If the AE meets [3 AND 4B] for any D_J then the AE is classified as an IRR associated with all D_J that meet the 4B criterion. IV- If the AE meets [3 AND 4A1] for a D_J then the AE is classified as an IRR associated with the D_J that meets the 4A1 criterion V- If the AE meets [3 AND 4A2] for a D_J then the AE is classified as an IRR associated with the D_J and associated with D_{J+1} that meets the 4A2 criterion	
1	PT is included in the 'IRRs SIGNS and SYMPTOMS' list
2A1	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • [AE timing related to study drug D_{J+1} ('BEFORE') OR AE onset date < date of infusion of study drug D_{J+1}] AND • AE outcome in ('RECOVERED/RESOLVED', 'RECOVERED/RESOLVED WITH SEQUELAE', 'RECOVERING/RESOLVING') AND • AE end date – AE onset date <=2
2A2	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • AE timing related to study drug D_{J+1} ('DURING', 'AFTER') AND • AE outcome in ('RECOVERED/RESOLVED', 'RECOVERED/RESOLVED WITH SEQUELAE', 'RECOVERING/RESOLVING') AND • AE end date – AE onset date <=2
3	PT is included in the 'IRRs CORE' list
4A1	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • [AE timing related to study drug D_{J+1} ('BEFORE') OR AE onset date < date of infusion of study drug D_{J+1}]
4A2	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • AE timing related to study drug D_{J+1} ('DURING', 'AFTER')
4B	AE onset on the day after infusion of study drug D_J

Appendix 3. Programmatic Derivation of Tumor Response Using RECIST v1.1

Methods of Tumor Assessment

The following are considered "interchangeable" methods: Conventional CT, Spiral CT, Enhanced CT, MRI and Enhanced MRI.

Period for Derivation of BOR and PFS

The period is defined from randomization until PD per RECIST v1.1 (Appendix 4 of B9991016 protocol) or death due to any cause.

Evaluation of Target Lesions at Each Visit

- The sum of lesion dimensions (SLD) is only considered if the methods of assessments are consistent with baseline. The "interchangeable" methods noted above are all considered consistent methods.
- In the SLD, the longest diameter will be used for non-nodal lesions and the short axis dimension will be used for each lymph node included in the sum.
- Complete Response (CR) is defined by the disappearance of all non-lymph node target lesions (where all target lesions are recorded with a length of 0 mm on the "Target Lesions" eCRF). Any pathological lymph nodes (recorded as target lesion) must have reduction in short axis to <10 mm. Note: the SLD may not be zero if lymph nodes are included as target lesions.
- Partial Response (PR) is defined by a 30% or more decrease in SLD of target lesions, taking as reference the baseline SLD.
- Progressive Disease (PD) is defined by a 20% or more increase in the SLD of target lesions relative to nadir (smallest SLD considering baseline and all assessments prior to the timepoint under evaluation), with a minimum absolute increase of 5 mm relative to nadir.
  - If only a subset of target lesions are assessed and the sum of the non-missing lesion diameters results in an increase above of at least 5 mm and at least 20% above the nadir, then the progression criteria will have been met.
- Stable Disease (SD) is assigned when neither sufficient shrinkage to qualify for CR or PR, nor sufficient increase to qualify for PD is observed, taking as reference the nadir.
- No Target Lesion at Baseline (NB) is assigned if "No Target Lesion" is checked, on the "Target Lesions" eCRF.
- Not All Evaluated (NAE) is assigned if:
  - Any individual target lesion is evaluated as "Indeterminate" on the "Target Lesions" eCRF
  - Inconsistent methods are used for any target lesions after randomization
  - One or more target lesions are not assessed

- One or more target lesions were excised or irradiated and have not reappeared or increased.

Evaluation of Non-Target Lesions at Each Visit

The lesions assessed are only considered for CR, Non-CR/Non-PD and PD if the methods of assessments are consistent with baseline. The "interchangeable" methods noted above are all considered consistent methods.

- CR is defined by the complete disappearance of all non-target lesions (where all non-target lesions are marked 'Absent Lesion or Normal Lymph Node' on the "Non-Target Lesion" eCRF).
- Non-CR/Non-PD is defined by persistence of one or more non-target lesions (ie, if all non-target lesions are marked 'Present/Not Increased' on the "Non-Target Lesion" eCRF).
- PD is assigned if any non-target lesion is marked "Increased" or "Present (First Appearance" on the "Non-Target Lesion" eCRF.
- No Non-Target Lesion at Baseline (NB) is assigned if "No Non-Target Lesions" is marked on the "Non-Target Lesions" eCRF.
- Not All Evaluated (NAE) is assigned if:
  - Any individual non-target lesion is evaluated as "Indeterminate" (marked as "Indeterminate" on the "Non-Target Lesions" eCRF)
  - Inconsistent methods are used for any non-target lesions after randomization
  - One or more non-target lesions are not assessed.

New Lesion: is defined by the appearance of 1 or more new lesions (where any lesion is marked 'New' on the "New Measurable Lesions" or "New Non-measurable Lesions" eCRFs). In addition, any lesion that is recorded for the first time after the start date without being marked as "New" on the "New Lesions" eCRF must be queried. In case the inconsistency is not resolved at time of data base snapshot/lock, the lesion will be considered as new and the patient in progression at that time point. Note: the requirement for consistent methods of assessment with baseline, obviously, does not apply for new lesions.

Objective Response Status At Each Assessment

- Overall response is determined from the derived target and non-target lesion data using conventions in Table 27 under the assumption that there are no new lesions identified at the visit. **If there are any new lesions at a time point, then the response is PD at that time point regardless of target or non-target lesion response**.
- Objective status after a change in modality that is not considered interchangeable is classified as NE.

- To group target and non-target lesion measurements to the corresponding actual post-baseline tumor assessment visit, a clustering algorithm is applied. Each cluster represents an actual visit. For each patient, the number of clusters is equal to the maximum number of assessments available among all target and non-target lesions. The SAS procedure, Proc Fastclus, is applied to a variable that represents the days from randomization to the date of the scan for lesion (date of scan – randomization +1). Then the assessments of lesions (target, non-target, new) that occurred close to each other in time will be assigned to the same cluster.

The rules for derived objective status for each visit are presented in Table 27.

Table 27. Derivation of Objective Status Based on Target and Non-Target Lesion Response Assuming No New Lesions

Target Lesion Response	Non-Target Lesion Response	Objective Status
CR	CR	CR
CR	Non-CR/Non-PD	PR
CR	PD	PD
CR	NAE	PR
PR	CR	PR
PR	Non-CR/Non-PD	PR
PR	PD	PD
PR	NAE	PR
SD	CR	SD
SD	Non-CR/Non-PD	SD
SD	PD	PD
SD	NAE	SD
PD	Any	PD
NAE	PD	PD
NAE	CR Non-CR/Non-PD NAE	NE
NB	CR	CR
NB	Non-CR/Non-PD	Non-CR/Non-PD
NB	PD	PD
NB	NAE	NE
CR	NB	CR
PR	NB	PR
SD	NB	SD
PD	NB	PD
NAE	NB	NE

CR=Complete Response, PR=Partial Response, SD=Stable Disease, PD=Progressive Disease,
NAE=Not All Evaluated; NE= Not Evaluable
Note: If non-target (or target) lesions are not collected at baseline, then the overall response is equivalent to the target (or non-target) lesions response, respectively.
If there are any new lesions at a time point, then the response is PD at that time point regardless of target or non-target lesion response.

Date of Response at Each Visit

If there are multiple assessments within a cluster, then the earliest scan date associated with the evaluation will be used as the date of the assessment. As such:

- The date of progressive disease is derived as the earliest date that PD has been documented, ie, the earliest date of: (target lesion assessments when the target lesion response is PD, non-target lesion assessments where the lesion status is 'increased' and the non-target lesion response is 'PD', new lesion)
- The date of CR, PR, non-CR/non-PR, SD, NE is derived as the date of the first radiographic evaluation included in the cluster.

Best Overall Response Evaluation for Each Patient

- Best overall response is derived from the sequence of objective responses reported during the "Period for Derivation of Best Overall Response".
- The best overall response is the best response: CR, PR, SD, Non-CR/Non-PD, PD or NE
  - For a patient to qualify for a best response of SD, the overall response evaluation must have met the stable disease criteria at least once ≥6 weeks (42 days) after randomization.
  - NE is assigned for all cases that do not qualify as CR, PR, SD, Non-CR/Non-PD or PD.
  - If a patient's first overall response other than NE is PD then the patient's best overall response is PD. If the first response of SD before the minimum period for SD defined above is followed by a PD then the best overall response is PD.