RENAL CELL CARCINOMA − JAVELIN RENAL 100

A PHASE 1B, OPEN-LABEL, DOSE-FINDING STUDY TO EVALUATE SAFETY, PHARMACOKINETICS AND PHARMACODYNAMICS OF AVELUMAB (MSB0010718C) IN COMBINATION WITH AXITINIB (AG-013736) IN PATIENTS WITH PREVIOUSLY UNTREATED ADVANCED RENAL CELL CANCER

STATISTICAL ANALYSIS PLAN – B9991002

Compounds:	MSB0010718C, AG-013736
Compound Name:	Avelumab, Axitinib
Version:	3.0
Date:	20-Dec-2017

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

CCI

APPENDICES

1. VERSION HISTORY

CCI

CCI

Version	Version Date	Summary of Changes
CCI		

		CCI
CCI		

		CCI [redacted]

		CCI
CCI		

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B9991002. 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.

Statistical analyses will be performed using cleaned eCRF data as well as non-CRF data (ie, pharmacokinetic (PK) data, and CCI tumor derived biomarker data). The primary analysis will include DLT data within the first 4 weeks (2 cycles) of treatment with avelumab in combination with axitinib and all data up to a cut-off date corresponding to 18 months (78 weeks) after the last patient receives the first dose of study treatment. The final analysis of the data will be performed after last patient last visit (LPLV).

Additional analyses of the data may be performed for publication or regulatory reporting purposes.

Throughout this document, 'start date' refers to date of randomization for randomized phase and first dose of study treatment for non-randomized phase.

2.1. Study Objectives

Primary Objectives

- To assess the safety and tolerability of avelumab in combination with axitinib in patients with previously untreated advanced RCC in order to estimate the Maximum Tolerated Dose (MTD) and select the Recommended Phase 2 Dose (RP2D).

Secondary Objectives

- To evaluate the overall safety profile of avelumab in combination with axitinib in patients with previously untreated advanced RCC.
- To assess the preliminary anti-tumor activity of avelumab in combination with axitinib in patients with previously untreated advanced RCC
- To evaluate the OS of avelumab in combination with axitinib in patients with previously untreated advanced RCC.
- To characterize the PK of avelumab and axitinib when administered in combination, and to assess the effect of avelumab on the PK of axitinib.
- To evaluate candidate predictive biomarkers in pre-treatment tumor tissue that may aid in the identification of a patient subpopulation most likely to benefit from treatment with avelumab in combination with axitinib.
- To assess the immunogenicity of avelumab when combined with axitinib.

CCI [redacted]

- CCI [redacted]
- CCI [redacted]

2.2. Study Design

This is a Phase 1b, open-label, multi-center, multiple-dose, safety, PK and pharmacodynamic study of avelumab in combination with axitinib in adult patients with previously untreated aRCC. This clinical study will be composed of a Dose-finding Phase and a Dose-expansion Phase.

The Dose-finding Phase will estimate the MTD and RP2D in patients with aRCC with clear cell histology who did not receive prior systemic therapy for advanced disease, using the modified toxicity probability interval (mTPI) method. Dose finding will follow an mTPI design, with up to 3 potential dose levels (DL) to be tested:

- (DL1): avelumab 10 mg/kg Q2W + axitinib 5 mg BID
- (DL-1A): avelumab 5 mg/kg Q2W + axitinib 5 mg BID
- (DL-1B): avelumab 10 mg/kg Q2W + axitinib 3 mg BID

DL-1A and DL-1B will be explored concurrently in a randomized fashion only if the MTD is exceeded in DL1.

To understand the extent of any effects of avelumab on axitinib PK, a 7-day lead-in period with single-agent axitinib will be included prior to Cycle 1 in all patients in the Dose-finding Phase of the study. Since avelumab has a half-life of 3-5 days, it would not be feasible to run a lead-in with avelumab alone to study the PK of avelumab alone. Therefore, the effect of axitinib on avelumab will be evaluated by comparing avelumab trough concentrations at steady state in the presence of axitinib with those reported for avelumab alone in prior studies.

The Dose-finding Phase will lead to the identification of an Expansion Test Dose for avelumab in combination with axitinib in patients with aRCC who did not receive prior systemic therapy for their advanced disease. The Expansion Test Dose will either be the MTD (ie, the highest dose of avelumab and axitinib associated with the occurrence of DLTs in <33% of patients) or the RP2D, ie, the highest tested dose that is declared safe and tolerable by the investigators and sponsor. Once the Expansion Test Dose is identified, the Dose-expansion Phase will be opened, and avelumab in combination with axitinib will be evaluated in up to approximately 40 patients with previously untreated aRCC.

Based on the emerging PK data and after the completion of the dose-finding phase of the study based on mTPI method, it may be necessary to enroll up to 8 additional patients to further assess the effect of avelumab on the PK of axitinib. These additional patients will undergo the same evaluations and procedures as those described in the Schedule of Activities table for the Dose-finding Phase of the study and will be treated concurrently with the initiation of the Dose-expansion Phase of the study. With the exception of the mTPI-related assessments leading to determination of MTD, any other assessments or procedures described for the Dose-finding Phase will also apply to these additional patients.

The number of patients to be enrolled in the Dose-finding Phase will depend on the observed safety profile, and the number of tested dose levels. Up to approximately 55 patients (including Dose-finding and Dose-expansion phases) are projected to be enrolled in the study.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoints

- DLTs within the first 4 weeks (2 cycles) of treatment with avelumab in combination with axitinib.

  Severity of adverse events will be graded according to CTCAE version 4.03. For the purpose of dose finding, any of the following adverse events occurring during the primary DLT observation period (4 weeks, from the time of first administration of avelumab [Cycle 1 Day 1] until the planned administration of the third dose of avelumab [Cycle 3 Day 1]) that are attributable to one, the other, or both agents in combination will be classified as DLTs:

  Hematologic:

  - Grade 4 anemia
  - Grade 4 neutropenia lasting >7 days;
  - Febrile neutropenia, defined as absolute neutrophil count (ANC) <1000/mm^3 with a single temperature of >38.3 degrees C (>101 degrees F) or a sustained temperature of ≥38 degrees C (100.4 degrees F) for more than 1 hour;
  - Grade ≥3 neutropenic infection;
  - Grade ≥3 thrombocytopenia with bleeding;
  - Grade 4 thrombocytopenia.

  Non-Hematologic (including Non-Laboratory):

  - Any Grade ≥3 toxicity, except for any of the following:
    - Transient (≤6 hours) Grade 3 flu-like symptoms or fever, which is controlled with medical management;

  - Transient (≤24 hours) Grade 3 fatigue, local reactions, or headache that resolves to Grade ≤1;
  - Grade 3-4 nausea and vomiting controlled by optimal medical therapy within 72 hours;
  - Grade 3 hypertension controlled by medical therapy;
  - Grade 3 diarrhea or Grade 3 skin toxicity that resolves to Grade ≤1 in less than 7 days after medical management (eg, immunosuppressant treatment) has been initiated;
  - Any Grade ≥3 amylase or lipase abnormality that is not associated with symptoms or clinical manifestations of pancreatitis;
  - Tumor flare phenomenon defined as local pain, irritation, or rash localized at sites of known or suspected tumor.

- Grade 3-4 liver function test elevation (alanine aminotransferase [ALT], aspartate aminotransferase [AST]) concurrent with Grade 2 elevation of total bilirubin.
- Non-hematologic Grade ≥3 laboratory abnormality if medical intervention is required to treat the patient, or if the abnormality leads to hospitalization.
  - Single laboratory values out of normal range that are unlikely related to trial treatment according to the investigator, do not have any clinical correlate, and resolve to Grade ≤1 within 7 days with adequate medical management are not to be considered DLTs.
- Inability to complete at least 75% of the first 2 cycles doses of axitinib (starting from Cycle 1 Day 1 after completion of the Lead-in PK period) or 2 infusions of avelumab within the DLT observation period due to investigational product-related toxicity.

In absence of associated clinical abnormalities, abnormal laboratory tests should be repeated to confirm relevance. While the rules for adjudicating DLTs in the context of dose finding/dose expansion phases are specified above, an AE not listed above, or an AE meeting the DLT criteria above but occurring outside of the DLT observation period may be defined as a DLT after consultation between sponsor and investigator, based on the emerging safety profile.

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 version 4.03 and coded using the Medical Dictionary for Regulatory Activities (MedDRA)

- Vital signs (blood pressure, pulse rate).

3.2.2. Efficacy endpoints

- Objective Response (OR) and Disease Control (DC) as assessed by Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1.

  OR is defined as complete response (CR) or partial response (PR) according to RECIST v1.1 from the 'start date' until the date of the first documentation of progressive disease (PD). Both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met.

  DC is defined as CR, PR, non-CR/non-PD or SD. Both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met. Criteria for SD and non-CR/non-PD must have been met at least 6 weeks after the 'start date'.

- Time-to-event endpoints: Duration of Response (DR), Progression-Free Survival (PFS), and Time to Tumor Response (TTR);

  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 or death due to any cause.

  TTR is defined, for patients with an OR, as the time from the 'start date' to the first documentation of objective response (CR or PR) which is subsequently confirmed.

  PFS is defined as the time from the 'start date' to the date of the first documentation of PD or death due to any cause, whichever occurs first.

- Overall Survival (OS);

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

3.2.3. Pharmacokinetic endpoints

- PK parameters including C_{max}, T_{max}, AUC_{τ}, and $t_{1/2}$ as appropriate for axitinib; C_{trough} and C_{max} for avelumab;

Table 2. PK Parameters to be Determined for Avelumab[a] and Axitinib

Parameter	Definition	Method of Determination
AUC_{last}	Area under the plasma concentration-time profile from time zero to the time of the last quantifiable concentration (C_{last})	Linear/Log trapezoidal method
AUC_{0-12}	Area under the plasma concentration-time profile after single dose from time zero to 12 hours	Linear/Log trapezoidal method
$AUC_{ss,\tau}$	Area under the plasma concentration-time profile after single dose from time zero to the next dose (at steady state)	Linear/Log trapezoidal method
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
$t_{½}$ [b]	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/F[b]	Apparent clearance	Dose / AUCτ for steady state
V_z/F[b]	Apparent volume of distribution	Dose / (AUCτ ·kel) for steady state
AUC_{last}(dn)	Dose normalized AUC_{last}	AUC_{last} / Dose
C_{max}(dn)	Dose normalized C_{max}	C_{max} / Dose

[a] Only C_{max} and C_{trough}
[b] If data permit

3.2.4. Immunogenicity endpoints

- Anti-Drug Antibody (ADA; Neutralizing antibodies (nAb)) of avelumab when combined with axitinib.

3.2.5. Biomarker endpoints

- Tumor tissue biomarker status (ie, positive or negative based on, for example, PD L1 expression and/or quantitation of tumor infiltrating CD8+ T lymphocytes as assessed by immunohistochemistry [IHC]).

Table 3. Biomarker Definition and Determination

Parameter	Definition	Method of Determination
PD-L1 expression	The number of PD-L1 positive cells and/or qualitative assessment of PD-L1 staining on tumor and inflammatory cells in regions of interest that are defined by tumor cell morphology and the presence or absence of CD8+ and CD68+ cells	Pathologist, assisted by image analysis
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
Tumor associated CD68+ macrophages	The number of CD68+ cells per unit area and the percent of counted cells that are scored as CD68+	Pathologist, assisted by image analysis

CCI

- CCI

- CCI

3.4. Baseline Variables

3.4.1. Study drug, study treatment and baseline definitions

In this study '**study drug**' refers to avelumab or axitinib and '**study treatment**' (or '**treatment group**') refers to one of the following:

- DL1 with lead-in: axitinib 5 mg oral BID followed by avelumab 10 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1A with lead-in: axitinib 5 mg oral BID followed by avelumab 5 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1B with lead-in: axitinib 3 mg oral BID followed by avelumab 10 mg/kg IV Q2W in combination with axitinib 3 mg oral BID, or

- DL1: avelumab 10 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1A: avelumab 5 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1B: avelumab 10 mg/kg IV Q2W in combination with axitinib 3 mg oral BID.

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 the study drugs in the combination.

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

Definition of baseline:

Definition of baseline for efficacy analyses in randomized phase

The last measurement prior to randomization will serve as the baseline measurement for efficacy analyses. If such a value is missing, the last measurement prior to the first dose of study treatment will be used as the baseline measurement except for analyses of tumor 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 efficacy analyses in non-randomized phase and for safety analyses

The last available assessment prior to the start of study treatment is defined as 'baseline' value or 'baseline' assessment for safety and efficacy (for non-randomized phase) 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 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.

3.4.2. Baseline characteristics

Baseline characteristics (including demographics, physical measurements, disease history and prior anti-cancer therapies) 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 for the first time, or if the worsening of an event is 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 Appendices 1 and 2, respectively.

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.

4.1. Full Analysis Set

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

For the non-randomized phase: The FAS will include all patients who receive at least one dose of study drug. Patients will be classified according to the study treatment actually received. If a patient receives more than one study treatment the patient will be classified according to the first study treatment received.

4.2. Safety Analysis Set

For the randomized phase: 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.

For the non-randomized phase: 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 actually received. If a patient receives more than one study treatment, the patient will be classified according to the first study treatment received. In the non-randomized phase of the study, the FAS and the safety analysis set are identical.

4.3. Other Analysis Set

4.3.1. DLT-evaluable set

The DLT-evaluable set includes all enrolled patients during the dose-finding phase who:

- receive at least 1 dose of avelumab and axitinib, and
- either experience DLT during the first 2 cycles of combination axitinib + avelumab treatment, or complete the primary DLT observation period for the first 2 cycles of combination treatment (4 weeks).

Patients who withdraw from study treatment before receiving at least 75% of the planned first 2 cycles of axitinib starting from Cycle 1 Day 1 after completion of the Lead-in PK period or 2 infusions of avelumab within the primary DLT observation period due to reasons other than investigational product-related adverse events are not evaluable for DLT.

This analysis set will be used only in the Dose-finding phase of the study for determination of MTD/RP2D.

4.3.2. PK analysis set

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 axitinib.

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 axitinib.

4.3.3. Biomarker analysis set

The biomarker analysis set for biomarkers that are measured only at screening is a subset of the safety analysis set and will include patients who have at least one screening biomarker assessment.

The biomarker analysis set for biomarkers that are measured sequentially is a subset of the safety analysis set and will include patients who have at least one screening and one on-treatment biomarker assessment for the same marker.

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.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and sample size determination

Dose-finding phase

There is no formal hypothesis testing planned for this study. The objective of the dose-finding phase is to determine MTD or RP2D.

Due to the dynamic nature of the dose allocation procedure and unknown safety profile of the combination, the sample size using mTPI approach (see Section 5.1.2 for details) cannot be determined in advance. It is expected that approximately 15 DLT-evaluable patients will be required for the dose-finding phase based on the mTPI method.

Based on the emerging PK data and after the completion of the dose-finding phase of study based on the mTPI method, it may be necessary to enroll up to 8 additional patients to further assess the effect of avelumab on the PK of axitinib. These additional patients will be treated concurrently with the initiation of the dose-expansion phase.

Dose-expansion phase

Up to approximately 40 patients will be included in the dose-expansion phase of the study for each MTD dose level cohort expanded. A sample size of 40 patients will provide at least 90% probability to observe at least 1 AE if the true incidence of the AE in the population is ≥6%.

It is expected that at most approximately 55 patients will be enrolled to achieve all study objectives.

5.1.2. Decision rules

Dose-finding phase

Many alternative designs have been proposed to the standard 3+3 design for Phase 1 dose escalation trials that improve accuracy, efficiency and statistical validity.

The modified toxicity probability interval (mTPI) design is an interval-type of design that uses a Bayesian statistics framework and a beta/binomial hierarchical model to compute the posterior probability of 3 dosing intervals that reflect the relative difference in probability between the toxicity rate of each dose level to the target rate (pT = 0.33). If the toxicity rate of the currently used dose level is far smaller than pT, the mTPI will recommend escalating the dose level; if it is close to pT, the mTPI will recommend continuing at the current dose; if it is far greater than pT, the mTPI will recommend de-escalating the dose level. These rules are conceptually similar to those used by the 3+3 design, except the decisions of an mTPI design are based on posterior probabilities calculated under a coherent probability model.

Being a model-based design, mTPI automatically and appropriately tailors dose-escalation and de-escalation decisions for different trials with different toxicity parameters. More importantly, all the dose-escalation decisions for a given trial can be pre-calculated under the mTPI design. Thus, compared to other advanced model-based designs published in the literature, the mTPI design is logistically less complicated and easier to implement.

Decision rules are based on calculating unit probability mass (UPM) of 3 dosing intervals corresponding to under, proper, and over dosing in terms of toxicity. Specifically, the underdosing interval is defined as (0; pT-e_1), the over-dosing interval (pT+e_2), and the proper-dosing interval (pT- e_1, pT+ e_2), where e_1 and e_2 are small fractions. Based on the safety profile of avelumab as a single-agent in Study EMR100070-001, e_1 is selected as 0.05, and e_2 is selected as 0.03. Therefore, the target dosing interval for the DLT rate is (0.25, 0.33).

The 3 dosing intervals are associated with 3 different dose-escalation decisions. The under-dosing interval corresponds to a dose escalation (E), over-dosing corresponds to a dose de-escalation (D), and proper-dosing corresponds to remaining at the current dose (R). Given a dosing interval and a probability distribution, the unit probability mass (UPM) of that dosing interval is defined as the probability of a patient belonging to that dosing interval divided by the length of the dosing interval. The mTPI design calculates the UPMs for the 3 dosing intervals, and the one with the largest UPM informs the corresponding dose-finding decision, which is the dose level to be used for future patients. For example, if the under-dosing interval has the largest UPM, the decision will be to escalate, and the next cohort of patients will be treated at the next higher dose level. Ji and collaborators [1] have demonstrated that the decision based on UPM is optimal in that it minimizes a posterior expected loss (ie, minimizes the chance of making a wrong dosing decision).

The dose finding phase of the trial uses a modified version of the mTPI design that maximizes the number of evaluable patients treated at each dose level at 6. Dose finding is completed when 6 DLT-evaluable patients have been treated at the highest dose with a DLT rate < 0.33.

Dose-expansion phase

There are no decision rules associated with this phase.

5.2. General Methods

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

- DL1 with lead-in: axitinib 5 mg oral BID followed by avelumab 10 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1A with lead-in: axitinib 5 mg oral BID followed by avelumab 5 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1B with lead-in: axitinib 3 mg oral BID followed by avelumab 10 mg/kg IV Q2W in combination with axitinib 3 mg oral BID, or
- DL1: avelumab 10 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1A: avelumab 5 mg/kg IV Q2W in combination with axitinib 5 mg oral BID, or
- DL-1B: avelumab 10 mg/kg IV Q2W in combination with axitinib 3 mg oral BID.

Baseline characteristics, disposition and efficacy data will be summarized based on the FAS by treatment group.

DLTs will be summarized based on the DLT-evaluable set by treatment group including data from the dose-finding phase only.

Other safety data, exposure data, concomitant medications and non-drug treatments will be summarized based on the safety analysis set by treatment group.

PK data will be summarized based on the PK analysis sets by treatment group.

Biomarker data will be summarized based on the biomarker analysis sets by treatment group.

Immunogenicity data will be summarized based on the immunogenicity analysis set by treatment group.

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/treated 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.

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

Start date of new anti-cancer therapy (drug, radiation, surgery) is used for censoring in efficacy analyses (see Section 6.2.2).

The start date of new anti-cancer therapy is the earliest date after the first dose of study treatment for the non-randomized phases or after the date of randomization for the randomized phase 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 'Concomitant Radiation Therapy', and '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', 'Concomitant Radiation 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 not 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 31 days prior to and including the 'start date'.
- 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 CR, PR, SD, non-CR/non-PD, or PD can be determined (see Section 6.2.2.1). 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'.

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

Whether actual or nominal PK sampling time will be used for the derivation of PK parameters will be determined by the results of interim PK analyses. 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 efficacy analyses and in the derivation of the end of on-treatment period. PD date below refers to PD date by investigator assessment.

- 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.

At the time of the SAP amendment 01 (SAP V2), the dose-finding phase had been completed and the only dose levels tested in the study will be DL1 and DL1 with lead-in. Therefore, no randomization will be implemented in the study.

Data will be summarized separately for DL1 and for DL1 with lead-in as well as for both treatment groups combined.

6.1. Primary Endpoints

6.1.1. DLT in Dose-finding Phase

6.1.1.1. Primary analysis

The following analyses will be based on the DLT-evaluable set for patients in the dose-finding phase only. DLTs will be listed and summarized by treatment group.

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 group. Assessment of response by the Investigator will be made using RECIST v1.1.

6.2.2.1. Objective response and tumor shrinkage from baseline

Best overall response (BOR) will be assessed based on reported overall lesion responses at different evaluation time points from the 'start date' until the first documentation of PD, according to the following rules. Only tumor assessments performed on or before the start date of any further anti-cancer therapies will be considered in the assessment of BOR. Clinical deterioration will not be considered as documentation of disease progression.

BOR Based on Confirmed Responses:

- CR = at least two determinations of CR at least 4 weeks apart and before first documentation of PD
- PR = at least two determinations of PR or better (PR followed by PR or PR followed by CR) at least 4 weeks apart and 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 the 'start date' and before first documentation of PD (and not qualifying for CR or PR).
- Non-CR/non-PD (applicable only to patients with non-measurable disease at baseline) = at least one non-CR/non-PD assessment (or better) ≥ 6 weeks after the 'start date' and before first documentation of PD (and not qualifying for CR or PR).
- PD = first documentation of PD ≤ 12 weeks after the 'start date' (and not qualifying for CR, PR, SD or non-CR/non-PD).
- NE: all other cases.

An objective status of PR or SD cannot follow one of CR. SD can follow PR 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. If this occurs, the sequence PR-SD-PR is considered a confirmed PR. A sequence of PR – SD – SD – PD would be a best response of SD if the window for SD definition has been met.

Objective Response (OR) is defined as confirmed BOR of CR or PR according to RECIST v1.1.

Patients who do not have a post-baseline radiographic tumor assessment due to early progression, who receive anti-cancer therapies other than the study treatments prior to reaching a CR or PR, or who die, progress, or drop out for any reason prior to reaching a CR or PR will be counted as non-responders in the assessment of OR. Each patient will have an objective response status (0: no OR; 1: OR). OR rate (ORR) is the proportion of patients with OR in the analysis set.

ORR by treatment group will also be calculated along with the 2-sided 95% CI using the Clopper-Pearson 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 BOR of CR, PR, SD, non-CR/non-PD (applicable only to patients with non-measurable disease at baseline), PD, and NE will be tabulated. Patients with BOR 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
- New anti-cancer therapy started before first post-baseline assessment
- SD of insufficient duration (<6 weeks after the ‘start date’)
- PD too late (>12 weeks after the ‘start date’)

Special and rare cases where BOR is NE due to SD of insufficient duration and late PD will be classified as ‘SD of insufficient duration’.

Tumor shrinkage will be summarized as the percent change from baseline in target lesions (sum of longest diameter for non-nodal lesion and short axis for nodal lesion) per time point. It will be derived as:

- ((Sum of target lesions at week XX – sum of target lesions at baseline)/sum of target lesions at baseline) × 100

The maximum reduction in target lesions from baseline will be derived across all the post-baseline assessments until documented disease progression, excluding assessments after the start of subsequent anti-cancer therapy, as:

- Minimum of ((sum of target lesions at week XX – sum of target lesions at baseline)/sum of target lesions at baseline) × 100

A waterfall plot of maximum percent reduction in the sum of longest diameter for non-nodal lesions and short axis for nodal lesions from baseline will be created by treatment group. These plots will display the best percentage change from baseline in the sum of the diameter of all target lesions for each patient with measurable disease at baseline and at least one post-baseline assessment.

6.2.2.2. Disease control

Disease Control (DC) is defined as BOR of CR, PR, non-CR/non-PD or SD. DC rate (DCR) is the proportion of patients with DC.

DCR will be summarized by frequency counts and percentages.

6.2.2.3. Duration of response

Duration of Response (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 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 assessment. The censoring rules for DR are as described for PFS in Table 4.

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

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median DR time with 2-sided 95% CIs. In particular, the DR rate at 3, 6 and 12 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley 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 (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 OR is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided.

6.2.2.4. Time to response

Time to response (TTR) is defined, for patients with OR, as the time from the 'start date' to the first documentation of objective response (CR or PR) which is subsequently confirmed.

TTR (in months) = [first date of OR – 'start date' +1]/30.4375

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

6.2.2.5. Progression-free survival

Progression-Free Survival (PFS) is defined as the time from the 'start date' to the date of the first documentation of PD or death due to any cause, whichever occurs first.

PFS data will be censored on the date of the last adequate tumor assessment for patients who do not have an event (PD or death), for patients who start a new anti-cancer therapy prior to an event (see Section 5.2.6) or for patients with an event after 2 or more missing tumor assessments. Patients who do not have an adequate baseline tumor assessment or who do not have an adequate post-baseline tumor assessment will be censored on the 'start date' unless death occurred on or before the time of the second planned tumor assessment (ie, ≤ 12 weeks after the 'start date') in which case the death will be considered an event.

In this study antitumor activity will be assessed through radiological tumor assessments conducted at screening and every 6 weeks until PD regardless of initiation of subsequent anti-cancer therapy. After 1 year from the 'start date', tumor assessments will be conducted less frequently, ie, at 12-week intervals.

The censoring and event date options to be considered for the PFS and DR analysis are presented in Table 4.

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

Table 4. Outcome and event dates for PFS and DR analyses

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment	'Start date'	Censored [a]
PD or death - After at most one missing or inadequate post-baseline tumor assessment, OR - ≤ 12 weeks after the 'start date'	Date of PD or death	Event
PD or death - after 2 or more missing or inadequate post-baseline tumor assessments	Date of last adequate tumor assessment [b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
No PD and no death	Date of last adequate tumor assessment [b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
Treatment discontinuation due to 'Disease progression' without documented progression	Not applicable	Information is ignored. Outcome is derived based on documented progression only.
New anti-cancer therapy given	Date of last adequate tumor assessment [b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored

[a] However if the patient dies ≤ 12 weeks after the 'start date' the death is an event with date on death date

[b] If there are no adequate post-baseline assessments prior to PD or death, then the time without adequate assessment should be measured from the 'start date'; if the criteria were met the censoring will be on the 'start date'.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median PFS time with 2-sided 95% CIs. In particular, the PFS rate at 3, 6, 9, 12, 15, 18 and 24 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley 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 (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 group. Reasons for censoring will be summarized according to the categories in Table 5 following the hierarchy shown.

Table 5. PFS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment	No adequate baseline assessment
2	Start of new anti-cancer therapy before event.	Start of new anti-cancer therapy
3	Event after 2 or more missing or inadequate post-baseline tumor assessments/ 'start date'	Event after 2 or more missing assessments [a]
4	No event and [withdrawal of consent date ≥ 'start date' OR End of study (EOS) = Subject refused further follow-up]	Withdrawal of consent
5	No event and lost to follow-up in any disposition page	Lost to follow-up
6	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 tumor assessment	No adequate post-baseline tumor assessment
7	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

[a] 2 or more missing or inadequate post-baseline tumor assessments.

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

Time of Follow-Up for PFS

A plot will be generated to compare planned and actual relative day of tumor assessments by treatment group. A Kaplan-Meier plot for PFS follow-up duration will also be generated to assess the follow-up time in the treatment groups reversing the PFS censoring and event indicators. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median time of follow-up for PFS with 2-sided 95% CIs. In particular, the rate at 3, 6, 9, 12, 15, 18 and 24 months will be estimated with corresponding 2-sided 95% CIs.

6.2.2.6. Overall survival

Overall survival (OS) is defined as the time from the 'start date' 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– 'start date' +1]/30.4375

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median OS time with 2-sided 95% CIs. In particular, the OS rate at 3, 6, 9, 12, 15, 18 and 24 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley 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 (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 group. Reasons for censoring will be summarized according to the categories in Table 6 following the hierarchy shown.

Table 6. OS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and [withdrawal of consent date ≥ 'start date' OR End of study (EOS) = Subject 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 > 16 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 groups reversing the OS censoring and event indicators. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median time of follow-up for OS with 2-sided 95% CIs. In particular, the rate at 3, 6, 9, 12, 15, 18 and 24 months will be estimated with corresponding 2-sided 95% CIs.

6.2.3. Pharmacokinetic endpoints

6.2.3.1. Pharmacokinetic Analysis of Avelumab and Axitinib

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

C_{trough} and C_{max} for avelumab and axitinib will be summarized descriptively (n, mean, SD, CV, median, minimum, maximum, geometric mean, its associated CV, and 95% CI) by treatment group, cycle, and day. Other standard parameters for axitinib will be calculated including, but not limited to T_{max}, AUC_{last}, T_{last}, $t_{1/2}$, CL, and $V_{z,}$ as data permit. Multiple Dose (MD) - $T_{ss,max}$, $AUC_{ss,\tau}$, $t_{1/2}$, $C_{ss,av}$, CL, and $V_{ss,}$ as data permit.

Dose normalized parameters (eg, CDN- C_{max}, CDN- C_{trough}) will be reported as appropriate. The trough concentrations for avelumab and axitinib will be plotted for each dose using a box whisker plot by cycle and day in order to assess the attainment of steady state.

Pharmacokinetic parameters for avelumab and axitinib will be taken from observed values or derived using non-compartmental methods from plasma concentration-time data as described in Section 3.2.3.

Presentation of pharmacokinetic data will include:

- Descriptive statistics (n, mean, SD, %CV, median, minimum, maximum) of plasma concentrations for axitinib will be presented in tabular form by treatment group, dose level, cycle, day and nominal time. Additionally similar descriptive statistics will also be generated for dose-normalized axitinib pharmacokinetic parameters

- Linear-linear and log-linear plots of mean and median plasma concentrations by nominal time for axitinib will be presented for PK sampling days by cycle and study day. Similar plots will be presented for axitinib for each individual patient concentrations. Only patients who have matched pairs of PK collections available on both planned treatments (when administered alone and in combination with avelumab) will be included in the axitinib plasma concentration descriptive summary and median concentration profiles. Patients who have undergone intrapatient dose reduction or escalation will be excluded from the median plasma concentration-time plots.

- Pharmacokinetic parameters for axitinib will be listed and summarized by cycle and study day using descriptive statistics (n, mean, SD, %CV, median, minimum, maximum, geometric mean and its associated %CV, and 95% CI). For Tmax, the range (min, max) will also be provided. Only patients who have a matched pair of estimated axitinib PK parameters available on both planned treatments (when administered alone and in combination) will be included in the axitinib PK parameter summary tables. PK parameters with zero values will be excluded from the calculation of geometric means and its associated %CV. If an intrapatient dose escalation or reduction occurs, dose-dependent PK parameters (AUC and Cmax) for that patient may be dose-normalized when it is known that the drug exhibits linear PK within the dose range and other PK parameters will be reported as estimated; or may only be included in descriptive statistics and summary plots up to the time of the dose change. In addition, dose-normalized axitinib Cmax and AUC parameters will be summarized (as described above) using data pooled across cohorts in which different axitinib doses were administered.

- Box plots for AUCτ and Cmax for axitinib (during treatments when given alone and in combination) will be generated. Individual data points, the geometric mean and the median of the parameter in each treatment will be overlaid on the box plots. If a cohort has limited evaluable PK data (n<4), matchstick plots showing changes in AUC and Cmax for each drug (during treatments when given alone and in combination) in individual patients will then be generated. The geometric mean of the parameter in each treatment will be overlaid in the plots. In addition, box plots for dose-normalized axitinib Cmax and AUCτ parameters will be created using data pooled across cohorts in which different axitinib doses were administered.

- Ctrough and Cmax for avelumab will be plotted for each treatment group using a box-whisker plot by cycle and day within cycle in order to assess the attainment of steady-state. Individual data points, the geometric mean and the median of the parameter in each treatment will be overlaid on the box plots.

6.2.3.2. Effect of avelumab on axitinib pharmacokinetics

The effect of repeated avelumab dosing on steady-state axitinib PK will be evaluated using Cmax and AUC_{τ} of axitinib on Lead-in Day 7 and Cycle 4 Day 1 as the primary PK parameters. The ratio of geometric means of Cmax and AUCτ, (axitinib in presence of avelumab/axitinib alone) will be computed to assess the magnitude of the effect.

The associated 90% CIs will also be computed for the geometric mean ratios.

6.2.4. 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 and axitinib exposure and biomarkers or significant safety/efficacy endpoints. The results of these analyses, if performed, may be reported separately.

6.2.5. Biomarker endpoints

Biomarkers will be summarized at all appropriate timepoints. For soluble proteins this is screening, Lead-in PK D1, C1D1, C2D1, C3D1, C4D1, C6D1, EOT and follow-up at day 30. For Anti-avelumab Antibodies, this is C1D1, C2D1, C3D1, C4D1, C6D1, C8D1, then Q12W, and 30 days after end of avelumab treatment. For biomarkers from Archival FFPE Tumor Tissue Block, this is screening only. For biomarkers from DeNovo Tumor Biopsy, this is screening and end of treatment. Appropriate summary statistics at each timepoint and ratio to screening (for continuous markers) will be provided separately for each dose escalation cohort, the dose expansion cohort, and dose expansion cohort plus the same dose escalation cohort. For continuous biomarkers, ratio to screening within each cohort will be tested using the Wilcoxon Signed-Rank Test when $n \geq 5$ for a specific timepoint. P-values will be compared to unadjusted and adjusted alpha; alpha will be adjusted using the Bonferroni adjustment (eg, 0.05/# tests). The adjustment will be determined before data base lock.

To assess whether biomarkers are related to outcome, screening (and ratio to screening for follow-up timepoints, if available) will be summarized versus OR status (0: no OR; 1: OR). For continuous variables, the value of the biomarker will also be displayed in a box plot vs OR status (0: no OR; 1: OR), for each visit.

6.2.6. Immunogenicity Endpoints

All analyses described below are performed for all treatment groups combined.

Blood samples for avelumab immunogenicity testing will be collected pre-dose on Cycle 1, 2, 3, 4, 6 and 8. Subsequently, testing should be performed approximately every 12 weeks. All samples should be drawn within 2 hours before start of avelumab infusion. Additional samples for anti-avelumab antibodies (and simultaneous PK draws for measurement of avelumab) will be collected at the Day 30 Follow-up visit after the end of avelumab treatment.

Samples positive for ADA will be analyzed for titer and may be analyzed for nAb. As of the finalization of this SAP, the nAb assay is not yet available, therefore the analyses of nAb data described in the following sections will only be conducted contingent upon assay and data availability at the time of reporting.

Patients will be characterized into different ADA categories based on the criteria defined in Table 7.

Table 7 Patients Characterized Based on Anti-Drug Antibody Results (ADA Status)

Category	Definition	Subjects 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 past-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 8. For nAb, treatment-boosted is not applicable since no titer result is available.

Table 8 Patients Characterized Based on Neutralizing Antibody Results (nAb Status)

Category	Definition	Subjects 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 past-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.

6.2.6.1. 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 ongoing treatment with avelumab, or

- the last ADA assessment is positive AND patient discontinued treatment with avelumab AND the last planned ADA assessment (day 30 follow-up visit) 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) and the CIs for the survival function estimates will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (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.

6.2.6.2. 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.

6.2.6.3. Analysis of PK, safety and efficacy 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 as outlined in the "Schedule for Pharmacokinetic Sample Collection" table of the Schedule of Assessments section of the protocol.

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 1 of Cycle 2 and the first positive ADA result is observed on Day 1 of Cycle 3, then the relative PK day for this C_{trough} is -14. 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, 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.

Efficacy and immunogenicity status

For the ADA ever-positive patients, a listing will be prepared with patient ID, start and stop of avelumab treatment, date of first positive ADA result, time to ADA response, duration of ADA response, date of last ADA positive result, BOR, DR, PFS time or censoring time and reason for censoring, and OS time or censoring time and reason for censoring. If applicable, date of first positive nAb result, time to nAb response, duration of nAb response, date of last nAb positive result will also be presented.

For the ADA ever-positive patients, the percent change from baseline in target lesions as well as the first occurrence of a new lesion and patient off avelumab treatment will be displayed against time point (weeks) in a line plot. Additional symbols will indicate the first and last ADA positive result and, if applicable, the first and last nAb positive result.

CCI [redacted]

CCI [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI [redacted]

6.4. Subset Analyses

Subset analyses will be performed for OR, DR and PFS per Investigator assessment, and OS based on the FAS for the subgroups defined below.

- MSKCC prognostic criteria at baseline (detailed in Section 6.5.1.3)
  - Favorable (Reference)
  - Intermediate or Poor
- Heng prognostic criteria at baseline (detailed in Section 6.5.1.3)
  - Favorable (Reference)
  - Intermediate
  - Poor

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 group.

6.5.1.1. Demographic characteristics

Demographic characteristics and physical measurements will be summarized by treatment group 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/Latino (Yes/No)
  - Age (years): summary statistics
  - Age categories :
    - < 65 years, ≥ 65 years
    - <65, 65-<75, 75-<85, ≥ 85
  - Pooled Geographical Region:

  - North America
  - 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 treated population)
- Geographic Region (as applicable):
  - North America
  - Latin America
  - Western Europe
  - Eastern Europe
  - Australasia
  - Middle East
  - 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 group, 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 Abuse' and RECIST eCRF pages will be summarized overall and by treatment group. Summary statistics will be presented for the following.

From the 'Primary Diagnosis' eCRF page:

- Body side
- Site of primary tumor
- Primary diagnosis (summarize all categories collected in the 'Primary Diagnosis' eCRF page)
- Time since initial diagnosis to 'start date' (months), defined as ('start date' – date of initial diagnosis)/30.4375
- Time since diagnosis of local/regional recurrence of disease (months), defined as ('start date'– date of diagnosis of local/regional recurrence of disease)/30.4375
- Time since diagnosis of recurrence/metastatic disease (months), defined as ('start date' – date of diagnosis of recurrence/metastatic disease)/30.4375

Memorial Sloan-Kettering Cancer Center (MSKCC) prognostic criteria (Motzer et al., 1999) are based on the detection of five risk factors:

- Karnofsky PS < 80% (since Karnofsky PS is not collected in the eCRF, the derivation of risk factors will use ECOG PS ≥ 2 (Oken et al., 1982))
- high LDH (> 1.5 times the ULN)
- serum hemoglobin < LLN
- high corrected serum calcium (> 10 mg/dL)
- prior nephrectomy

According to MSKCC prognostic criteria, patients will be assigned to one of the following risk groups:

- Favorable (0 risk factors)
- Intermediate (1 or 2 risk factors)

- Poor (3 or more risk factors)

Heng prognostic criteria (Heng et al., 2009) are based on the detection of six risk factors:

- Karnofsky PS < 80% (since Karnofsky PS is not collected in the eCRF, the derivation of risk factors will use ECOG PS ≥ 2 (Oken et al., 1982))
- high corrected serum calcium (> 10 mg/dL)
- serum hemoglobin < LLN
- time from diagnosis to start of systemic treatment < 12 months
- absolute neutrophils count > ULN
- platelets count > ULN

According to Heng prognostic criteria, patients will be assigned to one of the following risk groups:

- Favorable (0 risk factors)
- Intermediate (1 or 2 risk factors)
- Poor (3 or more risk factors)

From the 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 vs current vs former smoker
  - Smoking exposure (pack-years): 0, <20, 20-<40, ≥40 and summary statistics
  - Years since quitting: never smoker, current smoker, <5, 5-<10, ≥10 and summary statistics

  Specifications for computation:

  - Cigarette equivalents are calculated as follows: one cigar is regarded equivalent to 5 cigarettes and 1 pipe is regarded equivalent to 3 cigarettes

- Duration of nicotine consumption [years]:

  (end of nicotine consumption – start of nicotine consumption + 1) / 365.25

- Pack-years:
  - calculate cigarette equivalents per day using the conversion factors given above
  - convert to packs per day where 20 cigarettes are regarded as 1 pack
  - pack-years = packs per day × duration of nicotine consumption [years]

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.

6.5.1.4. Prior anti-cancer therapies

The prior anti-cancer therapies are collected under the 'Prior Cancer Therapy', 'Prior Radiation Therapy' and 'Prior Surgery' eCRF pages.

The number and percentage of patients with at least one prior anti-cancer surgery will be tabulated.

Prior anti-cancer therapies will be included in the listings that follow with a flag to identify prior therapies. These will include the patient identification number, and all the relevant collected data-fields on the corresponding eCRF pages.

- Listing of anti-cancer drug therapies
- Listing of anti-cancer radiotherapy
- Listing of anti-cancer surgeries

6.5.2. Study conduct and patient disposition

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

6.5.2.1. Patient disposition

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

- Total number of patients screened overall
- Number of patients who discontinued from the study prior to treatment with study drug overall and by the main reason for discontinuation
- Number and percentage of patients in each of the analysis sets defined in Section 4
- Number and percentage of patients with study drug ongoing (separately for each study drug)

- Number and percentage of patients who discontinued study drug overall and by the main reason for discontinuation of study drug (separately for each study drug)
- Number and percentage of patients who entered follow-up
- Number and percentage of patients who 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

In addition the following will be summarized:

- Number and percentage of treated patients overall, by region (Europe, EEA (required by EudraCT), North America, Latin America, Middle East, Asia, Australasia Africa), by country within region
- Number and percentage of treated patients by center

A cross tabulation of patients who have discontinued/are ongoing treatment with avelumab vs patients who have discontinued/are ongoing treatment with axitinib will also be provided.

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 group.

Cycle definitions for study drugs that are administered in combination apply to all the study drugs in the combination. Ie, cycle is patient-dependent, rather than study-drug-dependent when study drugs are administered in combination.

For Cycle X, 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, any study drug with 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, all study drugs had dose=0 at that visit). Use start date in the exposure page if available; if start date is not available then use date of collection of vital signs on Cycle X day 1 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 + 14 days – 1 day

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

When summarizing exposure for each study drug, only cycles from first dose of study treatment until the last cycle with non-zero dose of at least one of the study drugs should be included.

Exposure will be summarized (per cycle and/or overall) as dose received (cumulative dose, actual dose intensity) and as dose received relative to intended dose (relative dose intensity [RDI]).

The derivations below are provided for the following:

- Avelumab administered as a 1-hour IV infusion at a dose of 10 mg/kg or 5 mg/kg once every 2 weeks in 2-week cycles.
- Axitinib administered BID PO at a dose of 5 mg or 3 mg.

Analysis of exposure will be based on the calculated actual dose levels

- Avelumab - total dose / weight
- Axitinib – total dose.

Exposure and compliance will be summarized by cycle (avelumab only) and overall (avelumab and axitinib) for each study drug in each treatment group.

6.5.3.1. Exposure to avelumab

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.

Intended duration of treatment with avelumab (weeks) =

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

where end date = start date of last cycle with non-zero dose of avelumab + 14 – 1

Duration of exposure to avelumab (weeks) =

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

Cumulative dose in a cycle or overall is the sum of the actual dose levels of avelumab received in a cycle or overall, respectively.

Actual Dose Intensity (DI)

- By cycle actual DI (mg/kg/2-week cycle) = [cumulative dose in the cycle (mg/kg)]/[cycle duration (weeks)/2]
- Overall actual DI (mg/kg/2-week cycle) = [overall cumulative dose (mg/kg)] / [intended duration of treatment with avelumab (weeks)/2].

Relative Dose Intensity (RDI)

- Intended DI (mg/kg/2-week cycle) = [intended cumulative dose per cycle] / [intended number of 2 weeks in a cycle] = [d (mg/kg)] / [1(2-week cycle)] = d (mg/kg/2-week cycle)
- By cycle RDI (%) = 100 × [by cycle actual DI] / [intended DI]
  = 100 × [by cycle actual DI] / [d (mg/kg/2-week cycle)]
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [d (mg/kg/2-week cycle)]

where d can be 5 or 10.

6.5.3.2. Exposure to axitinib

The dose level is calculated as actual dose administered (mg/day).

The Intended duration of treatment with axitinib (weeks) =

(last dose date of axitinib – first dose date of axitinib + 1)/7

Duration of exposure to axitinib (weeks) =

(last dose date of axitinib – first dose date of axitinib + 1)/7

Cumulative dose is the sum of the actual doses of axitinib received in the study.

Actual Dose Intensity (DI)

- Overall actual DI (mg/week) = [overall cumulative dose (mg)] / [intended treatment duration (weeks)]

Relative Dose Intensity (RDI)

- RDI (%) = 100 × [overall cumulative dose] / [intended cumulative dose per week × number of weeks from first dose of study drug to last dose of axitinib]
- = 100 × [overall cumulative dose] / [7×2×d×duration of exposure to axitinib (weeks)], where d = 3 or 5.

6.5.3.3. Dose Reductions

Applicable to avelumab. Dose reduction is defined as actual non-zero dose < 90% of the planned dose.

Applicable to axitinib. Dose reduction is defined as a change to a non-zero dose level lower than that planned in the protocol.

The number and percentage of patients with at least one dose reduction as well as a breakdown of dose reductions (1, 2 , 3 , ≥4) will be summarized.

6.5.3.4. Dose escalations

Applicable to axitinib only.

Dose escalation is defined as a change to a dose level higher than that planned in the protocol.

The number and percentage of patients with at least one dose escalation as well as a breakdown of dose escalations (1, 2, ≥3) will be summarized.

6.5.3.5. Dose interruptions

Applicable to axitinib only.

An interruption is defined a 0 mg dose administered on one or more days. What follows defines how dose interruptions will be counted in the case of multiple dose interruptions.

- If an interruption occurs consecutively for at least two days due to the same reason, then it will be counted only once (example: If the actual dose on days 1-3 is 10 mg and actual dose on days 4-5 is 0 mg and dose interruption on days 4-5 is due to AE, then the total number of dose interruptions is 1).
- If an interruption occurs consecutively for at least two days due to different reasons, then it will be counted for each reason (example: If the actual dose on days 1-3 is 10 mg and actual dose on days 4-5 is 0 mg and dose interruption on day 4 is due to AE and dose interruption on day 5 is due to dosing error, then the total number of dose interruptions is 2).
- If an interruption occurs for more than one day due to the same reason, but the days are not consecutive, ie there is at least one dosing day in between, then each dose interruption will be counted as a different occurrence (example: If the actual dose on days 1, 3 and 5, is 10 mg and actual dose on days 2 and 4 is 0 mg, and dose interruptions on day 2 and 4 are both due to dosing error, the total number of dose interruptions is 2).

A dose interruption is not considered a dose reduction.

The number and percentage of patients with dose interruptions and the corresponding reasons will be summarized.

6.5.3.6. Dose Delays

Applicable to avelumab only.

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.

In Cycle 1,

Dose Delay (days) = day of first dose of avelumab – 1.

In Cycle >1

Dose Delay (days) = Date of first dose of avelumab in Cycle x – Date of first dose of avelumab in Cycle (x-1) – 14.

Dose Delays will be grouped into the following:

- No delay
- 1-3 days delay
- 4-6 days delay
- 7 or more days delay

For example, for avelumab, administered on a 2-week schedule, if one patient receives avelumab on Day 1, then the next avelumab administration date will be on Day 15; however, if the patient receives avelumab at Day 16, 17 or 18, this is considered as 1-3 days delay.

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.

6.5.3.7. Infusion rate reductions

Applicable to avelumab only.

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.

6.5.3.8. Infusion interruptions

Applicable to avelumab only.

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.

6.5.4. Concomitant medications and non-drug treatments

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

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

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

6.6.1. Adverse events

Treatment-emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period for the first time, or if the worsening of an event is during the on-treatment period as defined in Section 3.5.1.

All analyses described below 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 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 Safety Review Plan and finalized for analysis of the current studies 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 Safety Review Plan and finalized for analysis of the current studies 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 group, primary SOC and PT in decreasing frequency for all patients.

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 group:
  - TEAEs
  - TEAEs, Grade ≥ 3
  - Related TEAEs
  - Related TEAEs, Grade ≥ 3
  - TEAEs leading to dose reduction of avelumab
  - TEAEs leading to dose reduction of axitinib

  - TEAEs leading to interruption of avelumab
  - TEAS leading to interruption of axitinib
  - TEAEs leading to discontinuation of avelumab
  - TEAEs leading to discontinuation of axitinib
  - TEAEs leading to discontinuation of any study drug
  - TEAEs leading to discontinuation of all study drugs
  - Related TEAEs leading to discontinuation of avelumab
  - Related TEAEs leading to discontinuation of axitinib
  - Related TEAEs leading to discontinuation of any study drug
  - Related TEAEs leading to discontinuation of all study drugs
  - 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 related to avelumab by SOC and PT and worst grade
- TEAEs related to axitinib by SOC and PT and worst grade
- TEAEs related to any study drug by SOC and PT and worst grade
- TEAEs leading to death by SOC and PT
- Related TEAEs leading to death by SOC and PT
- TEAEs Excluding SAEs, with frequency ≥ 5% in any treatment group by SOC and PT

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 group:

- TEAEs leading to dose reduction of avelumab by SOC and PT
- TEAEs leading to dose reduction of axitinib 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 6.5.3.6). 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 6.5.3.6).

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 group:

- TEAEs leading to interruption of avelumab by SOC and PT
- TEAEs leading to interruption of axitinib 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 group:

- TEAEs leading to both interruption and dose reduction of avelumab by SOC and PT
- TEAEs leading to both interruption and dose reduction of axitinib by SOC and PT

This summary will take into account PTs with both actions as defined in Section 6.6.1, even though 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 group:

- 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 axitinib by SOC and PT
- Related TEAEs leading to discontinuation of axitinib by SOC and PT
- TEAEs leading to discontinuation of any study drug by SOC and PT
- Related TEAEs leading to discontinuation of any study drug by SOC and PT
- TEAEs leading to discontinuation of all study drugs by SOC and PT
- Related TEAEs leading to discontinuation of all study drugs by SOC and PT

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

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 group.

- 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 group:

- SAEs by SOC and PT

- Related SAEs by SOC and PT

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 group:

- 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 axitinib, by Cluster and PT
- irAEs leading to discontinuation of any study drug, by Cluster and PT
- irAEs leading to discontinuation of all study drugs, by Cluster and PT
- Serious irAEs, by Cluster and PT

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 group:

- IRRs leading to death, by PT
- IRRs, by PT
- IRRs, Grade ≥ 3, by PT
- IRRs leading to discontinuation of avelumab, by PT
- Serious IRRs, by PT
- Time related to first onset of an IRR (infusion 1, infusion 2, infusion 3, infusion 4 or later)

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.

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 $\geq$ 800/mm3
- Neutrophil count decreased
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count $\geq$ 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 group:

- 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 groups, 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 a post-baseline TBILI ≥ 2×ULN, ALT≥ 3×ULN or AST ≥ 3×ULN 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 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).

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), TSH, Free T4 , ACTH, ANA (performed at baseline and if clinically indicated), ANCA (performed at baseline and if clinically indicated), RF (performed at baseline and if clinically indicated), HBV and HCV (performed at baseline and if clinically indicated).

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), prothrombin time (PT), international normalized ratio (INR), and PTT.
- Urinalysis: all urinalysis parameters, including protein, glucose, blood and albumin
- Other parameters: hormone, and immunology parameters
- 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. 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). The average of the replicate measurements will 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. If QTcB and QTcF methods do not adequately correct for HR and there are a sufficient number of patients (eg >30) with baseline ECGs, an alternate correction to achieve the goal of getting uncorrelated QTc and RR is based on a linear regression methods which yields, theoretically, uncorrelated QTc and RR.

Linear regression method:

- Fit a model QT = a + b × RR to baseline data
- Use the estimated slope, $\hat{b}$, to correct QT
- Corrected QT for heart rate will be computed as follows:

$$QTcP = QT + \hat{b} \times (1 - RR)$$

Data will be summarized using QTcF and QTcB. However, if these are not appropriate for the data set due to an observed large correlation between corrected QT and HR using the baseline assessments, the results will also be summarized using QTcP.

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 group, 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, if applicable, QTcP) 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. Physical examination

Number and percentage of patients with abnormal findings in physical examination will be summarized by body system.

6.6.9. ECOG performance status

The ECOG shift from baseline to highest score during the on-treatment period will be summarized by treatment group. 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

Not applicable.

7.1. Introduction

Not applicable.

7.2. Interim Analyses and Summaries

Not applicable.

8. REFERENCES

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

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

3. Ji Y et al. A modified toxicity probability interval method for dose-finding trials. Clinical Trials 2010; 7:653 663

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

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

6. Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon-Alfa as a comparative treatment for clinical interferon-alfa as a comparative treatment for clinical cell carcinoma. Journal of Clinical Oncology. 20: 289-296, 2002.

7. Heng DYC, Xie W, Regan MM, Warren MA, Golshayan AR, Sahi C, Eigl BJ, Ruether JD, Cheng T, North S, Venner P, Knox JJ, Chi KN, Kollmannsberger C, McDermott DF, Oh WK, Atkins MB, Bukowski RM, Rini BI, Choueiri TK. Prognostic Factors for Overall Survival in Patients With Metastatic Renal Cell Carcinoma Treated With Vascular Endothelial Growth Factor–Targeted Agents: Results From a Large, Multicenter Study. Journal of Clinical Oncology. 27: 5794-5799, 2009.

8. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, Carbone PP. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol. 1982 Dec;5(6):649-55.

9. APPENDICES

Appendix 1. Immune-Related Adverse Events

The MedDRA PTs and clusters for irAEs are defined in the Safety Review Plan (SRP) for avelumab.

Immune-related AEs (irAEs) will be programmatically identified as outlined in Table 8. This case definition is hierarchical, ie, each step is only checked for patients and events that have already met the prior step.

Table 10. 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) below are met then the AE is in for the next step A) conmed ATC code is in (H02A, H02B, H02C, D07, A01AC, S01BA, S01BB, L04AA, L04AB, L04AC, L04AD, L04AX) 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"

5	A) No clear etiology (other than immune mediated etiology) 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]	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) 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 avelumab 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 avelumab infusion (during or after the infusion) or the day after the avelumab 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 9 and will be identified for IV drugs only.

Table 11. 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