A PHASE 2, OPEN LABEL STUDY TO EVALUATE SAFETY AND CLINICAL ACTIVITY OF AVELUMAB (BAVENCIO®) IN COMBINATION WITH AXITINIB (INLYTA®) IN PATIENTS WITH ADVANCED OR METASTATIC PREVIOUSLY TREATED NON-SMALL CELL LUNG CANCER OR TREATMENT NAÏVE CISPLATIN-INELIGIBLE UROTHELIAL CANCER

JAVELIN MEDLEY VEGF

STATISTICAL ANALYSIS PLAN – B9991027

Compounds:	MSB0010718C AG-013736
Compound Name:	Avelumab Axitinib
Version:	3.0
Date:	15-Dec-2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

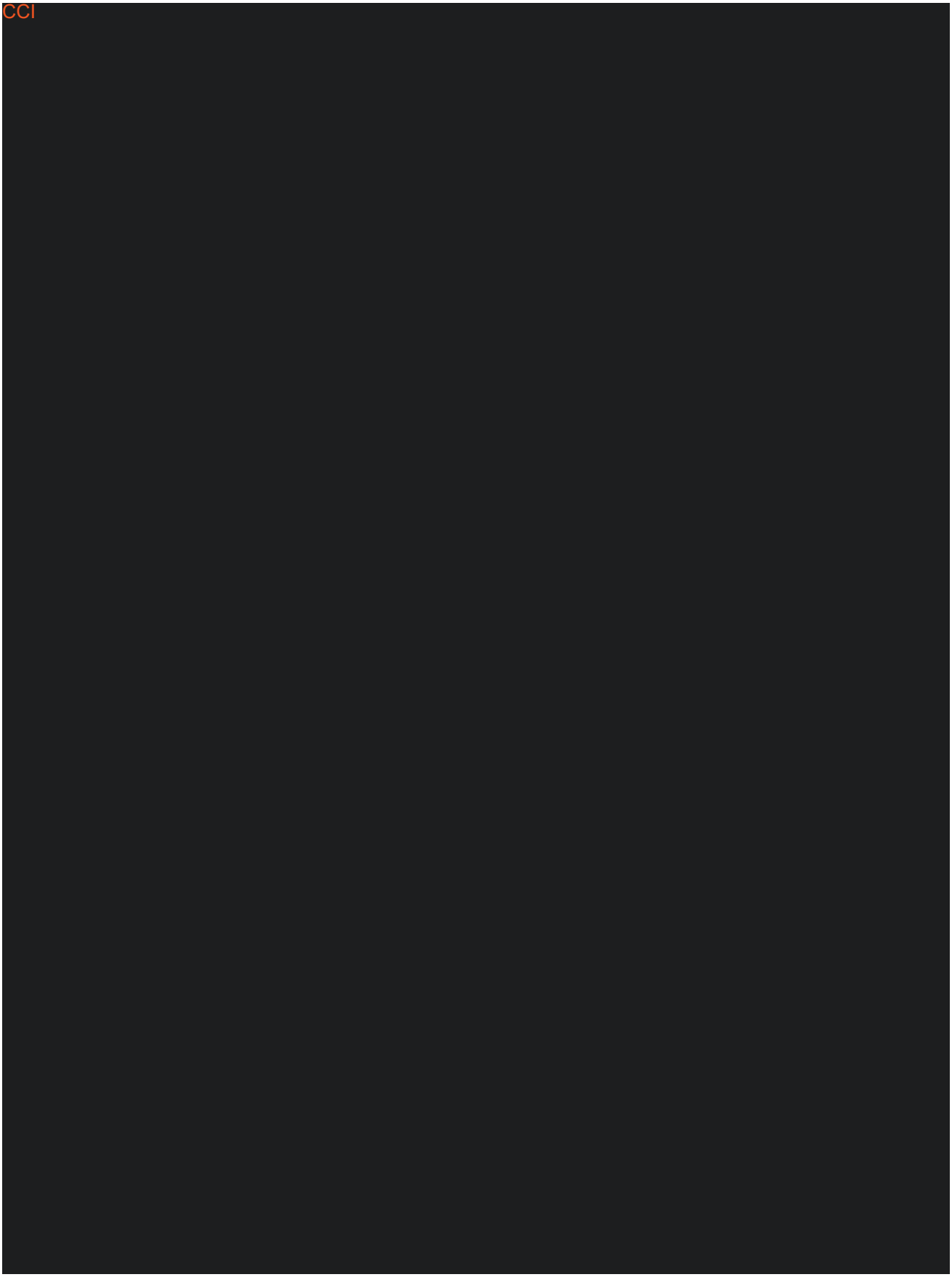
APPENDICES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B9991027 is based on the protocol amendment 1 dated 07MAR2018 and Protocol Administrative Change Letters dated 01JUN2018 and 14DEC2018.

CCI

CCI



CCI

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B9991027. 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, pharmacokinetics data, immunogenicity data, and biomarker data). The primary (also 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.

2.1. Study Objectives

Primary Objective

To evaluate the Objective Response Rate (ORR) based on Investigator assessment, per Response Evaluation Criteria in Solid Tumors (RECIST) version (v) 1.1 of avelumab in combination with axitinib in patients with advanced or metastatic non-small cell lung cancer (NSCLC) who have received at least one prior platinum-containing therapy and in treatment naïve patients with advanced or metastatic urothelial cancer (UC), who are ineligible for cisplatin-containing chemotherapy for their advanced disease.

Secondary Objectives

To evaluate the following in patients with advanced or metastatic NSCLC who have received at least one prior platinum-containing therapy and in treatment naïve patients with advanced or metastatic UC, who are ineligible for cisplatin-containing chemotherapy for their advanced disease:

- The safety and tolerability of avelumab in combination with axitinib;
- The anti-tumor activity of avelumab in combination with axitinib;
- The pharmacokinetics (PK) of avelumab and axitinib when administered in combination;
- Candidate predictive and/or early response biomarkers in pre-treatment tumor tissue and pre- and post-treatment blood samples that may aid in the identification of a patient subpopulation most likely to benefit from treatment with avelumab in combination with axitinib. Assessments may include but are not limited to tumor mutational burden, immune repertoire and PD-L1 expression within the tumor microenvironment;
- The immunogenicity of avelumab when combined with axitinib.

CCI

CCI

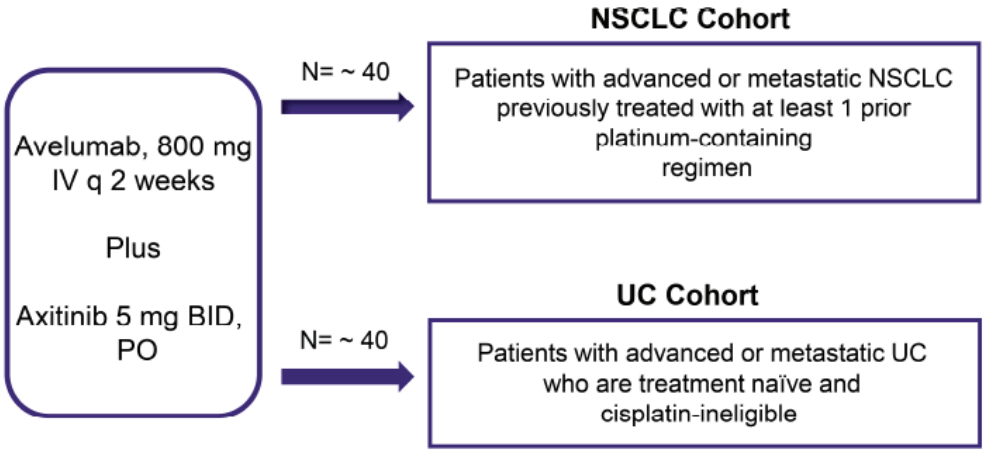
2.2. Study Design

This is an open-label, multi-center, Phase 2 study to evaluate the safety and efficacy of avelumab in combination with axitinib in patients with advanced or metastatic NSCLC who have received at least one prior platinum-containing therapy, and in patients with advanced or metastatic UC, who are treatment naïve and ineligible for cisplatin-containing chemotherapy for their advanced disease.

Patients will receive avelumab 800 mg intravenous (IV) every 2 weeks (Q2W) on Days 1 and 15 of each 28-day cycle in combination with axitinib 5 mg by mouth (PO) twice daily (BID) on a continuous dosing schedule.

Approximately 40 patients will be enrolled in each cohort, NSCLC and UC (Figure 1).

Figure 1. Study Design Schema



3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

- Confirmed Objective Response (OR) based on Investigator assessment per RECIST v1.1.

  OR is defined as complete response (CR) or partial response (PR) according to RECIST v1.1 from the date of first dose of study treatment 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.

3.2. Secondary Endpoints

3.2.1. Safety endpoints

- Adverse events (AEs) as characterized by type, severity as graded by National Cancer Institute [NCI] Common Terminology Criteria for Adverse Events [CTCAE] v4.03, timing, seriousness, and relationship to study treatments;

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

- Laboratory abnormalities as characterized by type, severity (as graded by NCI CTCAE v4.03) and timing.

3.2.2. Efficacy endpoints

- Time-to-event endpoints including time to tumor response (TTR), duration of response (DR), progression-free survival (PFS) based on Investigator assessment per RECIST v1.1, and overall survival (OS).

  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 date of first dose of study treatment to the first documentation of objective response (CR or PR) which is subsequently confirmed.

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

  OS is defined as the time from the date of first dose of study treatment to the date of death due to any cause.

3.2.3. Pharmacokinetic endpoints

- Pharmacokinetic parameters including trough and maximum concentrations (C_{trough}, C_{max}) of avelumab and axitinib.

  Pharmacokinetic parameters (C_{max} and C_{trough} after single dose and at steady state).

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

Parameter	Definition	Method of Determination
C_{max}	Maximum observed plasma concentration	Observed directly from data
C_{trough}	Pre-dose observed plasma concentration	Observed directly from data
C_{max}(dn)	Dose normalized C_{max} for axitinib	C_{max} / Dose
C_{trough}(dn)	Dose normalized C_{trough} for axitinib	C_{trough} / Dose

3.2.4. Immunogenicity endpoints

- Anti-drug antibody (ADA) titers and neutralizing antibodies (nAb) against avelumab.

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 mutational burden as well as characterization of the immune repertoire in peripheral blood and/or tumor).

Table 3. Biomarker Definition and Determination

Parameter	Definition	Method of Determination
PD-L1 expression	Assessment of the proportion (%) of tumor cells and immune cells displaying PD-L1 staining of any intensity within the tumor microenvironment	Pathologist, assisted by image analysis
Immune repertoire	Determination of the identity and abundance of individual T cell clones in peripheral blood and/or tumor tissue samples at baseline and on-treatment	TCR sequencing
Tumor mutational burden	Determination/estimation of the frequency of mutations (total and non-synonymous) present in baseline tumor derived nucleic acid samples	Whole exome or genome sequencing and/or RNAseq

CCI

[redacted]

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:

- NSCLC Cohort: avelumab 800 mg IV Q2W plus axitinib 5 mg PO BID;
- UC Cohort: avelumab 800 mg IV Q2W plus axitinib 5 mg PO BID.

Start and end dates of study treatment:

The date/time of first dose of study treatment in a combination group 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 in a combination group 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 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 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.

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.

3.4.2. Baseline characteristics

Baseline characteristics (including demographics, 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.

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

Adverse Events of Special Interest (AESIs)

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

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

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 treatment the patient will be classified according to the first study treatment received.

4.2. Safety Analysis Set

The safety analysis set will include all patients who receive at least one dose of study drug. Patients will be classified according to the study treatment actually received. If a patient receives more than one study treatment, the patient will be classified according to the first study treatment received. In this non-randomized study, the FAS and the safety analysis set are identical.

4.3. Other Analysis Set

4.3.1. PK analysis sets

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

4.3.2. Biomarker analysis set

Biomarker analysis set is a subset of the safety analysis set and will include patients who have at least one baseline biomarker assessment. Analysis sets will be defined separately for blood-based and tumor tissue-based biomarkers.

4.3.3. 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 assessment collected for avelumab.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and sample size determination

There is no formal hypothesis testing in this study. Approximately 40 patients will be enrolled from each of 2 tumor type cohorts: NSCLC, and UC. Thus, a total of approximately 80 patients will be enrolled. Within each treatment group, ORR will be estimated and the 2-sided exact 95% confidence intervals (CIs) will be calculated.

With 40 treated patients per tumor type cohort, ORR can be estimated with a maximum standard error of 0.079. Further, assuming beta (0.5, 0.5) prior:

- NSCLC Cohort: if 15 responders (out of 40 patients, ORR of 37.5%) are observed, the probability of a true ORR ≥ 30% (considered a clinically relevant effect above single-agent checkpoint inhibition) will be ≥80% (85.0%).

- UC Cohort: if 19 responders (out of 40 patients, ORR of 47.5%) are observed, the probability of a true response rate ≥ 40% (considered a clinically relevant effect above single-agent checkpoint inhibition) will be ≥80% (83.4%).

Table 4 provides the exact 95% CI for ORR based on different observed number of responders in a given tumor type cohort.

Table 4. Sample Size and Exact 95% CI for ORR

Number of Patients	Number of Responders	Observed ORR	Exact 95% CI for ORR
40	4	10.0%	(2.8% - 23.7%)
	6	15.0%	(5.7% - 29.8%)
	8	20.0%	(9.1% - 35.6%)
	10	25.0%	(12.7% - 41.2%)
	12	30.0%	(16.6% - 46.5%)
	14	35.0%	(20.6% - 51.7%)
	16	40.0%	(24.9% - 56.7%)
	18	45.0%	(29.3% - 61.5%)
	20	50.0%	(33.8% - 66.2%)
	24	60.0%	(43.3% - 75.1%)
	28	70.0%	(53.5% - 83.4%)

CI=confidence interval; ORR=objective response rate.

5.1.2. Decision rules

There are no formal decision rules in this study.

5.2. General Methods

As described in Section 3.4.1, in this study '**treatment group**' (ie, cohort) refers to one of the following:

- NSCLC Cohort: avelumab 800 mg IV Q2W plus axitinib 5 mg PO BID;
- UC Cohort: avelumab 800 mg IV Q2W plus axitinib 5 mg PO BID.

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

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 set by treatment group and for all treatment groups combined.

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

Immunogenicity data will be summarized based on the immunogenicity analysis set by treatment group and for all treatment groups combined.

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 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 6.1.1 and Section 6.2.2).

The start date of new anti-cancer therapy is the earliest date after the first dose of study treatment 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 Treatment', and 'Follow-up Radiation Treatment' eCRF pages with 'Treatment Intent' = 'Curative in intent'
- Surgery date recorded in 'On-Study Anti-Cancer Surgery', and 'Follow-up Anti-Cancer 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 Treatment', 'Follow-up Radiation Treatment', 'On-Study Anti-Cancer Surgery', and 'Follow-up Anti-Cancer 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.

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 (if such analyses are performed).

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 date of first dose of study treatment.
- 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 6.1.1.1). 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' or 'NA'.

5.3.1.1. Pharmacokinetic concentrations

Concentrations Below the Limit of Quantification

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

Deviations, Missing Concentrations and Anomalous Values

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

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

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

5.3.1.2. Pharmacokinetic parameters

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 is missing, 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
- 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

- Both Year (YYYY) and Month (MMM) for start of anti-cancer therapy are available

IF

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

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

THEN

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

ELSE IF

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

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

THEN

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

ELSE IF

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

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

THEN

imputed start date = 01 MMM YYYY;

ELSE IF

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

THEN

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

ELSE IF

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

THEN

imputed start date = 01 MMM YYYY.

6. ANALYSES AND SUMMARIES

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

6.1. Primary Endpoint

6.1.1. Objective response as assessed by the Investigator per RECIST v1.1

6.1.1.1. Primary analysis

The following analyses will be based on the FAS by treatment group. Assessment of response will be made using RECIST v1.1. Assessments below refer to investigator assessment.

Best overall response (BOR) will be assessed based on reported overall lesion responses at different evaluation time points from the date of first dose of study treatment 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 date of first dose of study treatment 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 date of first dose of study treatment and before first documentation of PD (and not qualifying for CR or PR).
- PD = first documentation of PD ≤ 12 weeks after the date of first dose of study treatment (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 [2] (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 a confirmed 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 confirmed 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 (i.e., death prior to 8 weeks after first dose of study treatment)
- 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 date of first dose of study treatment without further evaluable tumor assessments)
- PD too late (>12 weeks after the date of first dose of study treatment)

Special and rare cases where BOR is NE due to both SD of insufficient duration and late PD will be classified as ‘SD too early’ (ie, SD of insufficient duration).

6.2. Secondary Endpoints

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 will be made using RECIST v1.1. Tumor-related endpoints will be analyzed based on investigator assessment.

6.2.2.1. Tumor shrinkage from baseline

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 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 diameters of all target lesions for each patient with measurable disease at baseline and at least one post-baseline assessment.

6.2.2.2. 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 in Table 5.

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

Table 5. Outcome and Event Dates for DR Analyses

Scenario	Date of event/censoring	Outcome
PD or death - After at most one missing or inadequate post-baseline tumor assessment, OR - ≤ 16 weeks after the date of first dose of study treatment	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 [a] 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 [a] 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 [a] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored

[a] If there are no adequate post-baseline assessments prior to PD or death, then the time without adequate assessment should be measured from the first dose of study treatment; if the criteria were met the censoring will be on the first dose of study treatment.

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 rates at 2, 4, 6, 8, 10, 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 (1982)[1] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[3] (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.

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 6 following the hierarchy shown.

Table 6. DR Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	Start of new anti-cancer therapy	Start of new anti-cancer therapy
2	Event after 2 or more missing or inadequate post-baseline tumor assessments/date of randomization	Event after 2 or more missing assessments [a]
3	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Patient refused further follow-up]	Withdrawal of consent
4	No event and lost to follow-up in any disposition page	Lost to follow-up
5	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
6	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.

6.2.2.3. Time to response

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

TTR (in months) = [first date of OR – date of first dose of study treatment +1]/30.4375

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

6.2.2.4. Progression-free survival

Progression-free survival (PFS) is defined as the time from the date of first dose of study treatment 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 date of first dose of study treatment unless death occurred on or before the time of the second planned tumor assessment (ie ≤ 16 weeks after the date of first dose of study treatment) 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 8 weeks (± 7 days) for 1 year from start of study treatment, and then every 12 weeks (± 7 days) thereafter until PD regardless of discontinuation of study treatment or initiation of subsequent anti-cancer therapy.

The censoring and event date options to be considered for the PFS analysis are presented in Table 7.

PFS (months) = [date of event or censoring– date of first dose of study treatment +1]/30.4375

Table 7. Outcome and Event Dates for PFS Analyses

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment	Date of first dose of study treatment [a]	Censored [a]
PD or death - After at most one missing or inadequate post-baseline tumor assessment, OR - ≤ 16 weeks after the date of first dose of study treatment	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 ≤16 weeks after the date of first dose of study treatment 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 date of first dose of study treatment; if the criteria were met the censoring will be on the date of first dose of study treatment

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 rates at 2, 4, 6, 8, 10, 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 (1982)[1] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[3] (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 Table 8 following the hierarchy shown.

Table 8. 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	Start of new anti-cancer therapy
3	Event after 2 or more missing or inadequate post-baseline tumor assessments/date of first dose of study treatment	Event after missing assessments [a]
4	No event and [withdrawal of consent date ≥ date of first dose of study treatment OR End of study (EOS) = Patient 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, including the median time of follow-up for PFS with 2-sided 95% CIs.

6.2.2.5. Overall Survival

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

OS (months) = [date of death or censoring– date of first dose of study treatment +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 rates at 2, 4, 6, 8, 10, 12, 15, 18, 24 and 30 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982)[1] and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[3] (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 9 following the hierarchy shown.

Table 9. OS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and [withdrawal of consent date ≥ date of first dose of study treatment OR End of study (EOS) = Patient refused further follow-up]	Withdrawal of consent
2	No event and [lost to follow-up in any disposition page OR data cut-off date – last contact date > 14 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, including the median time of follow-up for OS with 2-sided 95% CIs.

6.2.3. Pharmacokinetic endpoints

The following pharmacokinetic analyses will be based on the PK analyses set by treatment group and for all patients.

Pharmacokinetic parameters (C_{trough} and C_{max}) for avelumab and axitinib will be calculated from observed concentrations. Dose normalized concentrations (eg, CDN- C_{max}, CDN-C_{trough}) for axitinib will be reported as appropriate.

Presentation of pharmacokinetic data will include:

- Pharmacokinetic plasma concentrations for avelumab and axitinib will be listed and summarized in tabular form by treatment group, cycle, study day and nominal time using descriptive statistics (n, mean, SD, %CV, median, minimum, maximum, geometric mean and its associated %CV, and 95% CI). For axitinib, if an intrapatient dose escalation or reduction occurs, dose-dependent PK parameters (C_{trough} and C_{max}) for that patient may be dose-normalized when the dose is confirmed within the linear 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, the dose-normalized axitinib C_{max} and C_{trough} parameter will be summarized (as described above).

- Box plots for C_{trough} and C_{max} for avelumab and axitinib 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 treatment group has limited evaluable PK data (n<4), matchstick plots showing changes in C_{trough} and C_{max} for each drug 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 C_{max} and C_{trough} parameters will be created using data pooled across treatment groups.
- C_{trough} and C_{max} 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.
- Observed concentrations of avelumab and axitinib from this study may be compared with the historical observed concentrations when each drug is administered alone.

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

Secondary biomarker endpoints include PD-L1 expression, tumor mutational burden, and immune repertoire in peripheral blood and/or tumor.

Biomarker data will be analyzed based on the biomarker analysis sets as defined in Section 4.3.2. Data will be presented by treatment group.

Biomarkers values at screening will be listed and summarized.

For these biomarker endpoints, patients may be classified as positive, negative, or some other category according to scoring algorithms and cut-offs established from internal or external sources. If no external standards exist, patients may be stratified using the median, quartiles and tertiles. Patients whose status cannot be determined are not considered to have screening biomarker assessment per the biomarker analysis set definition, and therefore will be excluded from the analyses.

For continuous measurement biomarker results, summary statistics (eg, the mean, standard deviation, median, percent of coefficient of variation, and minimum/maximum levels) will be determined at baseline and on-treatment/end of treatment time points, as appropriate.

Appropriate change from baseline measurements will be provided.

For discrete measurement biomarker results (eg, tumor marker status), frequencies and percentages of categorical biomarker measures will be determined at baseline and on-treatment/end of treatment time points. Shift tables may also be provided.

Clinical responses will be summarized by treatment group and by biomarker status following the methodology outlined in Section 6.1.1.1. The number of responders (patients with BOR of CR or PR) will be tabulated relative to biomarker classifications using a contingency table and a Fisher's exact test will be performed.

DR, PFS and OS (if meaningful) will be summarized by treatment group and by biomarker status following the methodology outlined in Sections 6.2.2.2, 6.2.2.4 and 6.2.2.5.

6.2.6. Endpoints for immunogenicity data of avelumab

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

Blood samples for avelumab immunogenicity testing will be collected pre-dose on Cycle 1 Day 1 and Day 15, Cycle 2 Day 1, and Day 15 of Cycles 3, 6, 9 and 12, as well as at the end of treatment and Day 30 follow-up visit. All samples should be drawn within 2 hours before start of avelumab infusion.

Samples positive for ADA will be analyzed for titer and may be analyzed for nAb. The analyses of nAb data described in the following sections will be conducted contingent upon data availability at the time of reporting.

Patients will be characterized into different ADA categories based on the criteria defined in Table 10.

Table 10 Patients Characterized Based on Anti-Drug Antibody Results (ADA Status)

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

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

Patients will be characterized into different nAb categories based on the criteria in Table 11. For nAb, treatment-boosted is not applicable since no titer result is available.

Table 11 Patients Characterized Based on Neutralizing Antibody Results (nAb Status)

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

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

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

6.2.6.1. 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)[1] and the CIs for the survival function estimates will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002)[3] (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 and safety by immunogenicity status

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

ADA

- ADA ever-positive versus ADA never-positive

- ADA: treatment-induced ADA versus ADA never-positive or baseline ADA positive

nAb

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

PK parameters and immunogenicity status

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

Blood samples for avelumab PK will be collected pre-dose and post-dose on Cycle 1 Day 1 and Day 15, Cycle 2 Day 1, and pre-dose on Day 15 of Cycles 3, 6, 9 and 12.

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 box plots of mean and median for C_{trough} and C_{max} over nominal time and by ADA status will be presented, as data permit.

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 -28.

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.

CCI

6.4. Subset Analyses

OR and DR (if meaningful) will also be summarized for patients in the following subsets:

Within each cohort:

- Baseline PD-L1 status: Positive, Negative.

NSCLC cohort:

- Histology: Non-Squamous, Squamous;
- Number of previous systemic therapies: 1, ≥2.

UC cohort:

- Baseline ECOG performance status (PS): ≤1, ≥2.

In the case of a low number of patients within a category (<10 patients), the category may not be summarized.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline summaries

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

6.5.1.1. Demographic characteristics

Demographic characteristics 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, Multiracial, Not Reported.
  - Ethnic origin:
    - Hispanic or Latino
    - Not Hispanic or Latino
  - Age (years): summary statistics
  - Age categories:
    - < 65 years, ≥ 65 years
    - < 65, 65-<75, 75-<85, ≥ 85 years
  - Pooled Geographical Region (as applicable):
    - North America
    - 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
    - Middle East

  - Australasia
  - Asia
  - Africa
- Eastern Cooperative Oncology Group (ECOG) Performance Status: 0, 1, 2, 3, and 4

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 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.

For the UC cohort, the reasons for ineligibility for cisplatin-containing front-line chemotherapy will be summarized using information collected on 'Oncology Stratification By Inclusion Criteria' and 'ECOG Performance Status' eCRF pages:

- Baseline ECOG PS 2;
- Renal dysfunction (defined as creatinine-clearance <60 ml/min) at baseline;
- Grade ≥2 peripheral neuropathy at baseline;
- Grade ≥2 hearing loss (hearing loss measured by audiometry of 25 decibels at two contiguous frequencies) at baseline.

6.5.1.3. Disease characteristics

Information on disease characteristics collected on 'Primary Diagnosis' and RECIST eCRF pages will be summarized by treatment group. Summary statistics will be presented for the following.

From the 'Primary Diagnosis' eCRF page:

- Site of primary tumor
- Primary diagnosis (summarize all categories collected in the 'Primary Diagnosis' eCRF page)

- Time since initial diagnosis to date of first dose of study treatment (months), defined as (date of first dose of study treatment – date of initial diagnosis)/30.4375

From the RECIST eCRF page:

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

Listing of disease history will be provided with all relevant data (as collected on the 'Primary Diagnosis' eCRF page) 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 Anti-Cancer Surgery' eCRF pages.

The number and percentage of patients in each of the following anti-cancer therapy categories will be tabulated:

- Patients with at least one type of prior anti-cancer therapy
- Patients with at least one prior anti-cancer drug therapy
- Patients with at least one prior anti-cancer radiotherapy
- Patients with at least one prior anti-cancer surgery

Prior anti-cancer drug therapy will be summarized as follows based on the number and percentage of patients with the following:

- At least one prior anti-cancer drug therapy
- Number of prior anti-cancer drug therapy regimens: missing, 1, 2, 3, ≥4
- Intent of Drug Therapy: Neo-Adjuvant, Adjuvant, Advanced – Metastatic, Locoregional Disease-Recurrence
- Best response: CR, PR, SD, PD, Unknown, Not applicable. Best response is derived from the last treatment regimen.

The prior anti-cancer drug therapies will also be summarized based on the number and percentage of patients by the drug class 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. The summary 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.

Prior anti-cancer drug therapies will be included in the listing 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

6.5.2. Study conduct and patient disposition

The following analyses will be performed based on the FAS 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 treated 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 administered in combination)
- Number and percentage of treated patients who discontinued study drug overall and by the main reason for discontinuation of study drug (separately for each study drug administered in combination)
- 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

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

In addition, 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 + 28 (in 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 may 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 formulae below should be applied to each study drug separately even when study drugs are administered in combination.

The derivations below are provided for the following study drugs (administered alone or in combination):

- Avelumab administered as a 1-hour IV infusion at a dose of 800 mg once every 2 weeks in 4-week cycles.
- Axitinib administered orally at a dose of 5 mg BID.

6.5.3.1. Exposure to avelumab

The dose level for avelumab is calculated as actual dose administered (mg).

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 + 28 – 1

Duration of exposure to avelumab (weeks) =

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

Cumulative dose is the sum of the actual doses of avelumab received.

Actual Dose Intensity (DI)

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

Relative Dose Intensity (RDI)

- Intended DI (mg/4-week cycle) = [intended cumulative dose per cycle] / [intended number of 4-weeks in a cycle] = [2 x 800 (mg)] / [1 (4-week cycle)] = 1600 (mg/4-week cycle)
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [1600 (mg/4-week cycle)]

6.5.3.2. Exposure to axitinib

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

Intended duration of treatment with axitinib (weeks) =

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

Duration of exposure to axitinib (weeks) =

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

Overall 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 study drug]

  = 100 × [overall cumulative dose] / [7 ×2 × 5 × duration of exposure to axitinib in weeks]

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 the number 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.

Number and percentage of patients with at least one dose escalation as well as a breakdown of dose escalations (1, 2, 3, ≥ 4) 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 for axitinib. 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, 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, then the total number of dose interruptions is 1).
- If an interruption occurs for more than one day, 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, 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.

For Cycle 1 Dose 1:

Dose Delay (days) = day of the first day of study drug -1

After Cycle 1 Dose 1:

> Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) – Planned days between two consecutive doses= Date of Dose x – Date of Dose (x-1) – 14.

Dose delays will be grouped into the following categories:

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

Concomitant medications will be summarized from the 'General Concomitant Medications' eCRF page.

Summary of concomitant 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 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.

6.5.5. Subsequent anti-cancer therapies

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

Anti-cancer drug treatment will be provided in a data listing with data retrieved from 'Follow-up Cancer Therapy' eCRF page.

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 Treatment' and 'Follow-up Anti-Cancer 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 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 study 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 study 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.

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
- TEAEs 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 leading to death by SOC and PT
- Related TEAEs leading to death 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 both interruption and dose reduction 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
- 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).

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

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

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

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

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

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

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

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

Summary of liver function tests will include the following categories. The number and percentage of patients with each of the following during the on-treatment period will be summarized by treatment 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 concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and (ALP ≤ 2×ULN or missing) will be provided.

Parameters with NCI-CTC grades available:

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

- The shift table will summarize baseline CTCAE grade versus the worst on-treatment CTCAE grade. The highest CTCAE grade during the on-treatment period is considered as the worst grade for the summary.

- The number and percentage of patients with newly occurring or worsening laboratory abnormalities during the on-treatment period will be summarized by worst grade on-treatment (Grade 1, 2, 3, 4, Grade 3/4 and any grade (Grades 1-4)).

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

- Hematology:

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

- Serum Chemistry:

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

6.6.5.2. Other laboratory parameters

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.

6.6.6. Vital signs

Weight will be recorded at screening and pre-dose Days 1 and 15 of each cycle, End of Treatment, and short-term follow-up (30, 60, and 90 days post treatment). Height will be measured at screening only.

Vital sign summaries will include all vital sign assessments from the on-treatment period.

All vital sign 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

Patients with qualitative ECG abnormalities will be listed for each patient and time point and the corresponding abnormality finding will be included in the listing.

6.6.8. MUGA/ECHO

In addition, LVEF% will be summarized as frequency (number and percentage) of patients with:

- a shift from baseline normal to at least one result below the institutional lower limit of normal during the on-treatment period
- ≥10-point decrease from baseline in LVEF% during the on-treatment period
- ≥10-point decrease from baseline in LVEF% to a post-baseline value < LLN during the on-treatment period
- ≥15-point decrease from baseline in LVEF% during the on-treatment period
- ≥15-point decrease from baseline in LVEF% to a post-baseline value < LLN during the on-treatment period

7. INTERIM ANALYSES

There is no formal interim analysis planned for this study.

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. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika; 26, 404-413, 1934

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

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

9. APPENDICES

Appendix 1. Immune-Related Adverse Events

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

Immune-related AEs (irAEs) will be programmatically identified as outlined in Table 12. Unless otherwise noted, this case definition is hierarchical, ie, each step is only checked for patients and events that have already met the prior step.

Table 12. 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'	Steps 3 and 4 will be checked concurrently. Step 5 will be checked if the criteria in Step 4 is met, irrespective of whether the Criteria in Step 3 is met.
4	AE treated with corticosteroids or other immunosuppressant therapy. For endocrinopathies only: AE required hormone replacement	Look in the conmed pages for AE identifiers that match the AEs from Step 3. For each of such AEs if A) OR B) OR C) below are met then the AE is in for the next step A) conmed ATC code is in (H02A, H02B, H02C, D07, A01AC, S01BA, S01BB, L04AA, L04AB, L04AC, L04AD, L04AX, A07EA) and AE PT is in any of the irAE clusters. B) conmed ATC code is in (H03A, H03B) and AE PT is in one of the irAE clusters associated with "Immune-mediated endocrinopathies" C) conmed ATC code is A10A and AE PT is in the irAE cluster associated with "Immune-mediated endocrinopathies: Type I Diabetes Mellitus"

5	A) No clear etiology (other than immune mediated etiology)	A) From the AE eCRF page. Is the AE clearly related to an etiology other than immune-mediated etiology? Yes / No If answer is Yes, check all that apply: • Underlying malignancy / progressive disease. • Other medical conditions. • Prior or concomitant medications / procedures. • Other. Specify.
	B) Histopathology / biopsy consistent with immune-mediated event	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)
	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]	

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

Appendix 2. Infusion Related Reactions

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

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

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

Infusion-related reactions (IRRs) will be programmatically identified as outlined in Table 13 and will be identified for IV drugs only.

Table 13. 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