Statistical Analysis Plan for Analysis of Efficacy and Safety

Clinical Trial Protocol Identification No.	EMR 100070-003
Title:	A Phase II, open-label, multicenter trial to investigate the clinical activity and safety of avelumab (MSB0010718C) in subjects with Merkel cell carcinoma
Trial Phase	Phase II
Clinical Trial Protocol Version	08 January 2016 /Version 8.0
Statistical Analysis Plan Author	PPD
Statistical Analysis Plan Date and Version	31 March 2016/Version 3.0
Statistical Analysis Plan Reviewers	PPD

[redacted]

[redacted]

1 Signature Page

Statistical Analysis Plan: EMR 100070-003

A Phase II, open-label, multicenter trial to investigate the clinical activity and safety of avelumab in subjects with Merkel cell carcinoma

PPD

PPD

Signature

Trial Biostatistician

PPD

PPD

Date of Signature

PPD

Signature

Project Statistician
Merck KGaA
Franfurter Strasse 250
64293 Darmstadt

PPD

Date of Signature

PPD

Signature

Medical Responsible
EMD Serono, Inc.
45A Middlesex Turnpike
Billerica, MA 01821, USA

PPD

Date of Signature

2 Table of Contents

3 List of Abbreviations and Definition of Terms

ACTH	Adrenocorticotropic hormone
ADR	Adverse drug reaction
AE(s)	Adverse event(s)
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
ANC	Absolute neutrophil count
ANCA	Antineutrophil cytoplasmic antibody
aPTT	Activated partial thromboplastin time
AST	Aspartate aminotransferase
BOR	Best overall response
CI	Confidence interval(s)
CK20	Cytokeratin 20
C_{min}	Minimum postdose (trough) concentration
CR	Complete response
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic case report form
CCI	[redacted]
FACT-M	Functional Assessment of Cancer Therapy – Melanoma
GCP	Good Clinical Practice
GGT	Gamma-glutamyl transferase
HAHA	Human antihuman antibody

ICH	International Conference on Harmonisation
IEC	Independent Ethics Committee
IERC	Independent Endpoint Review Committee
IMP	Investigational Medicinal Product
INR	International normalized ratio
irAE	Immune-related adverse event
IRB	Institutional Review Board
CCI [redacted]	
IRR	Infusion-related response
CCI [redacted]	
IV	Intravenous
IVRS	Interactive voice response system
LDH	Lactate dehydrogenase
MCC	Merkel cell carcinoma
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
CCI [redacted]	
MedDRA	Medical Dictionary for Regulatory Activities
ORR	Objective response rate
OS	Overall survival

PD	Progressive disease
PD-1	Programmed death 1
PD-L1	Programmed death ligand 1
PFS	Progression-free survival
PK	Pharmacokinetic(s)
PR	Partial response
RBC	Red blood cell
RECIST 1.1	Response Evaluation Criteria in Solid Tumors version 1.1
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SLD	Sum of the longest diameter
SMC	Safety Monitoring Committee
StdDev	Standard Deviation
TEAE	Treatment-emergent adverse event
TTP	Time to progression
ULN	Upper limit of normal
WBC	White blood cell

4 Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	03JUL2014	PPD	NA. The first version
2.0	23MAR2015	PPD	Details of changes are specified in Section 5.1 "Changes to Previous Version"
2.1	17NOV2015	PPD	Details of changes are specified in Section 5.1 "Changes to Previous Version"
3.0	31MAR2016	PPD	Details of changes are specified in Section 5.1 "Changes to Previous Version"

5 Purpose of the Statistical Analysis Plan

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specifications for the interim and final analysis of data collected for protocol EMR 100070-003. Results of the analyses described in this SAP will be included in the Clinical Study Report (CSR). Additionally, the planned analyses identified in this SAP will be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective SAP will be clearly identified in the CSR.

The SAP is based upon section 8 (Statistics) of the trial protocol and protocol amendments and is prepared in compliance with ICH E9. The first version (version 1.0) focuses on the detailed description of the primary and key secondary endpoints analysis and key safety endpoints analysis, whereas detailed specifications of the statistical analyses for other variables will be added in a later version of this SAP. Version 2.0 and 2.1 were created based on the protocol version 7.0 dated 26 February 2015. Additional updates in version 3.0 are specificied in Section 5.1, Changes to Previous Version.

5.1 Changes to Previous Version

Version 2.0

The following changes are made in version 2.0 of the SAP:

1. The futility analysis date was updated from at least 6 months to be at least 3 months after 20 subjects have been enrolled; the scope of the futility analysis was updated to include primary endpoint, secondary efficacy endpoints and safety analysis.
2. The compound name was changed from MSB0010718C to avelumab per Clinical Trial Protocol amendment 6 (26 FEB 2015);

3. Details of pharmacokinetics data analysis (Section 16.5), immunogenicity data analysis (Section 16.6) and brief description of CCI [redacted] were added in this version;
4. Brief description of subject health-related quality of life interview analysis was added in Section 16.4.4;
5. PD-L1 expression status was added in the subgroup definition;
6. The determination of date of tumor assessment result per IERC was updated to be detailed in the Imagine Review Charter;
7. Imputation algorithm for incomplete prior/concomitant medication dates was added in Section 11;
8. Dose reduction definition was updated in Section 15;
9. Immune-related adverse events (irAEs) was defined in Section 17.1; Appendix III was included to define irAEs per the pre-specified serach list of MedDRA preferred terms.
10. Appendix II, Programmed Major Protocol Deviations was updated to include two deviation criteria, subjects with dosing errors and subjects who did not receive mandatory premedications;
11. Medical responsible personnel of the study was updated;
12. Appendix II Programmed Major Protocol Deviations was updated per Protocol Amendnent 7.0 by medical responsible personnel;
13. Added a sensitivity analysis of primary endpoint to only include those subjects with measurable disease at baseline per RECIST 1.1 and by the IERC in Section 16.1;
14. Infusion Related Reaction was defined in Section 17.1 and summary of the analysis was added in Section 17.2.4;
15. The ITT and Safety analysis set were updated in Section 10; Section 8.1 Sequence of Analyses was updated accordingly.

Version 2.1

The following changes are made in version 2.1 of the SAP:

1. Section 8 Sample size calculation was updated with the actual enrollment of 88 subjects who received at least one dose of study treatment. The nominal p-values for the 2-stage group sequential testing strategy were updated accordingly.
2. Section 9 Changes to the Planned Analyses was updated to specify the actual number of subjects who received at least one dose of study drug would be included in the primary analysis;
3. Section 10 and Section 13.1, Australia was updated to be in Rest of the World subgroup of pooled region;
4. Section 10 Per Protocol Analysis Set was updated to include criteria “Have evaluable tumor assessments”;

5. Section 10 "Number of prior systemic therapies for metastatic disease" was added as a Subgroup Analysis Set;
6. CCI [redacted]
7. Section 11 definition of on-treatment period was updated; data analysis for re-initiated treatment was specified;
8. Section 11 visit window algorithm was added for by-visit summary of safety parameters;
9. CCI [redacted]
10. Section 13.3.2 Summary of prior anti-cancer therapy was updated to include additional summary by metastatic disease settings;
11. Section 16.1 definition of the confirmed BOR was updated to exclude assessments after start of subsequent anticancer therapy;
12. Section 16.2.3 censoring rule for overall survival was updated to be the last date known to be alive per M-S updated standard;
13. CCI [redacted]
14. Section 16.4.5 duration of Follow-up was updated;
15. Section 16.5 Pharmacokinetic Data was updated by M-S PK team;
16. Section 16.6 title was updated from Immunogenicity to Human Anti-human Antibody (HAHA). The contents of Section 16.6 was updated by M-S PK team;

CCI [redacted]

18. Section 17.1.1 Overall AE summary category was updated. AESI was removed, treatment emergent irAEs was added;
19. Section 17.2.3 Other Adverse Events was updated to remove AESI; AE summary of treatment emergent irAEs was added; the exploratory analysis of AE with PK paramaters was updated by M-S PK team;
20. CCI [redacted]
21. Section 17.3.1 liver function elevation crieteria was updated and eDISH plot was added per M-S updated standard;
22. Section 17.5.1 potentially clinically significant abnormalities criteria for ECG was updated per M-S updated standard; descriptive summary of ECG parameters change over time was added;
23. Appendix III Pre-specified Search List of MedDRA Preferred Term for Immune Related Adverse Events was removed from this SAP. A version-controlled search list will be available in Sponsor's MARVEL system.

Version 3.0

The following changes are made in version 3.0 of the SAP:

1. Cover page the list of statistical analysis plan reviewers was updated;
2. Section 9 Changes to the Planned Analyses in the Clinical Trial Protocol;
3. Section 10 Analysis Sets was updated with additional subgroup sets;
4. Section 11 General Specification for Statistical Analyses was updated for data handling rule after cut-off date per sponsor master SAP template;
5. Section 12.1 Disposition of Subjects and Discontinuations was updated with additional disposition summary;
6. Section 12.2 Protocol Deviations was updated with standard text per sponsor master SAP template;
7. Section 14.1 Prior and Concomitant Medications/Procedures was updated by adding a listing of premedication;
8. Section 14.2 Subsequent Anti-Cancer Therapies/Procedures was added;
9. Section 16.1 Primary Endpoint Analyses Table 2 was updated to use the original table as described in the paper by Eisenhauer, et al; additional summary for patients with BOR of NE was added;
10. Section 16.2.2 A clarification was made in the censoring rules for Progression Free Survival regarding the definition of last adequate tumor assessment.
11. Section 16.2.4 Response Status at 6 and 12 Months after Start of Study Treatment was revised back to the wording in SAP v1 to drop the condition “have evaluable tumor assessment at the time of the analysis”, as this condition was added incorrectly in the SAP v2.1; summary of durable response was added;
12. Section 16.3 Analysis on Subgroups was updated by adding analyses to assess the association between the primary and key secondary endpoints (BOR, PFS and OS) and each subgroup factor;
13. Section 16.4.1 Time to Response per RECIST 1.1 was added;
14. Section 16.4.5 Health-related Quality of Life was updated by adding line plot and figure plot;
15. Section 16.5.1 Descriptive PK Analysis was updated for drug concentration;
16. Section 16.5.2 Population Pharmacokinetic Analysis was updated to include the details of PopPK analysis in a separate Data Analysis Plan.

CCI [redacted]

18. Section 17.1 Adverse Events was updated for the definition of Infusion Related Reaction per the updated definition by the sponsor; definition of Adverse Events of Special Interest (AESI) was removed as this information was only collected on the eCRF page at the very beginning of the study;

19. Section 17.3.1 Hematology and Chemistry Parameters was updated to have CTCAE version 4.03; corrected calcium calculation was added; eDISH plot was updated to plot peak serum (ALT/AST) vs. peak total bilirubin.

6 Summary of Clinical Trial Features

Trial Objectives	**Primary objective** The primary objective of the trial is to assess the clinical activity of avelumab as determined by the objective response rate (ORR) according to Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) by an Independent Endpoint Review Committee (IERC) in subjects with metastatic Merkel cells carcinoma MCC after failing first-line chemotherapy. **Secondary objectives** Secondary objectives are as follows: • To assess the duration of response according to RECIST 1.1 • To assess the progression-free survival time (PFS) according to RECIST 1.1 • To assess the safety profile of avelumab in subjects with MCC • To assess the overall survival (OS) time • To assess response status according to RECIST 1.1 at 6 and 12 months after start of study treatment • To characterize the population PK of avelumab in subjects with MCC by sparse sampling • To evaluate the immunogenicity of avelumab and to correlate it to exposure **Exploratory objectives** Exploratory objectives are as follows: • CCI [redacted]

	CCI [redacted]
Primary endpoint(s)	The primary endpoint for the trial is the confirmed BOR, per RECIST 1.1, as determined by an IERC. Both CR and partial response (PR) must be confirmed by a second tumor assessment that will be performed preferably at the regularly scheduled 6-week assessment interval, but no sooner than 5 weeks after the initial documentation of CR or PR.
Secondary/ Exploratory endpoints	Secondary endpoints include • duration of response according to RECIST 1.1 as determined by an IERC, • PFS time according to RECIST 1.1 as determined by an IERC, • occurrence and severity of treatment-related AEs according to NCI-CTCAE v 4.0, • OS time, • response status according to RECIST 1.1 at 6 and 12 months after start of study treatment, • serum titers of anti-avelumab antibodies, and • population PK profile of avelumab (sparse sampling). Exploratory endpoints include • CCI [redacted]

	CCI [redacted]
Trial design and plan	This is a multicenter, international, single-arm, open-label, Phase II trial that will evaluate the efficacy and safety of avelumab in subjects with metastatic MCC. Subjects must have received at least 1 line of chemotherapy for the treatment of metastatic MCC. Up to 84 eligible subjects will receive avelumab at a dose of 10 mg/kg as a 1-hour intravenous (IV) infusion (-10 minutes / +20 minutes, that is, over 50 to 80 minutes) once every 2 weeks. Tumor measurements (including the assessment of skin lesions by physical examination) to determine response will be performed every 6 weeks and response to the treatment will be evaluated by RECIST 1.1. Treatment will continue until • therapeutic failure (subjects may stay on treatment beyond observation of progressive disease [PD] provided there is no significant clinical deterioration); • unacceptable toxicity; or • any criterion for withdrawal from the trial or the trial drug is fulfilled. Significant clinical deterioration is defined as • new symptoms or worsening of symptoms that cannot be managed by optimal supportive care or disease localization that require immediate medical or surgical intervention (for example, lesion close to the spine), and/or • change in Eastern Cooperative Oncology Group Performance Status (ECOG PS) to ≥ 3 that lasts more than 14 days. Decisions regarding medical management of subjects will be made by the Investigator; however, the primary and secondary endpoint determinations will be according to tumor assessments performed by the IERC. The date of the first observation of PD by RECIST 1.1 by the IERC will be used to determine the date of the PD as well as the duration of response in all subjects, including the subjects for which treatment was maintained beyond first determination of disease progression. Adverse events will be assessed throughout the trial period and evaluated using the National Cancer Institute (NCI) Common Technology Criteria version 4.0 (CTCAE v 4.0). CCI [redacted] There will be 1 interim analysis for futility after 20 subjects have been enrolled and observed for at least 3 months and 1 interim analysis for efficacy 6 months after 56 subjects have been enrolled. The primary analysis will be conducted 6 months after the accrual of the last subject and a further exploratory analysis will be conducted 12 months after the accrual of the last subject. Subject follow-up for progression and survival will continue

	until 1 year after the last subject receives the last dose of avelumab.
Planned number of subjects	Eighty-four subjects are planned to be enrolled in this trial.
Investigational Medicinal Product (s): dose/mode of administration/ dosing schedule	Avelumab will be administered as a 1-hour IV infusion (-10 minutes / +20 minutes, that is, over 50 to 80 minutes) at 10 mg/kg once every 2-week treatment cycle. In order to mitigate infusion-related reactions, subjects will receive pretreatment with H1 blockers and acetaminophen 30 to 60 minutes prior to every avelumab infusion. Steroids as premedication are not acceptable. A premedication regimen of 25 to 50 mg diphenhydramine and 650 mg acetaminophen (IV or oral equivalent) is recommended prior to each dose of study drug. This regimen may be modified based on local treatment standards and guidelines as appropriate. The dose of avelumab will be calculated based on the weight of the subject determined on the day of each drug administration. Immediate access to an intensive care unit or equivalent environment and appropriate medical therapy (including IV epinephrine, corticosteroids, antihistamines, bronchodilators, and oxygen) must be in place for use in the treatment of potential infusion-related reactions. Infusion of avelumab will be stopped in case of Grade ≥ 2 infusion-related, allergic, or hypersensitivity reactions (according to NCI-CTCAE v 4.0). Following avelumab infusions, subjects must be observed for 2 hours post infusion for potential infusion-related reactions. In the case of Grade 1 or 2 infusion reactions, the infusion rate should be decreased by 50%. If the subject has a second infusion-related reaction Grade ≥ 2 on the slower infusion rate, the infusion should be stopped and the subject should be removed from avelumab treatment.
Reference therapy(ies): dose/mode of administration/ dosing schedule	Not applicable.
Planned treatment duration per subject	Subjects will receive avelumab treatment until significant clinical deterioration, unacceptable toxicity, or any criterion for withdrawal from the trial or trial drug is fulfilled. Subjects who have experienced a confirmed complete response (CR) should be treated for a maximum of 12 months and a minimum of 6 months after confirmation, at the discretion of the investigator. If the investigator believes that a subject may benefit from treatment beyond 12 months, it may be permissible after discussion with the sponsor. In case a subject with a confirmed CR relapses after stopping treatment, but prior to the end of the

	trial, 1 re-initiation of treatment is allowed at the discretion of the Investigator and agreement of the Medical Monitor. In order to be eligible for retreatment, the subject must not have experienced any toxicity that led to treatment discontinuation of the initial avelumab therapy. Subjects who re-initiate treatment will stay on study and will be treated and monitored according to the protocol and the "until progression" schedule in the Schedule of Assessments.
Randomization and Blinding	Not applicable.

7 Sample Size/Randomization

The primary endpoint of the trial is the confirmed best overall response (BOR) according to RECIST 1.1, based on independent review of tumor assessments. The objective response rate (ORR) will be determined as the proportion of subjects with a confirmed BOR of PR or CR. The trial aims at demonstrating an ORR greater than 20% by means of an exact binomial test.

The planned total sample size is 84 subjects. There will be 1 interim analysis for futility after 20 subjects have been enrolled and observed for at least 3 months and 1 interim analysis for efficacy 6 months after 56 subjects have been enrolled. The primary analysis will be conducted 6 months after the accrual of the last subject. An exploratory analysis of secondary and exploratory endpoints will be conducted 12 months after the accrual of the last subject.

The following assumptions are made for the sample size calculation:

- ORR of 35%
- overall alpha = 0.025 (1-sided) for the test of the null hypothesis of an ORR $\leq$ 20%

The following analyses are planned:

- Futility: Enrollment will be stopped for futility if no response (confirmed or unconfirmed) is observed in the first 20 subjects after 3 months of follow-up.
- Efficacy: A two-stage group sequential testing approach will be applied for efficacy. The null hypothesis can be rejected if 20 subjects in the interim analysis after 56 subjects, or 25 subjects in the primary analysis after 84 subjects, show a confirmed PR or CR according to RECIST 1.1. The corresponding nominal p-values of the exact binomial test are 0.0045 and 0.0214, respectively. The resulting overall probability of reaching a positive result in the interim or primary analysis under the null hypothesis assumption of an ORR $\leq$ 20% is $\leq$ 0.0225, as derived from the binomial distribution according to Jennison and Turnbull (1); therefore the overall type I error rate is controlled at a level of 2.5% (one-sided).

Under the given assumptions, the power to reject the null hypothesis at the interim or the primary analysis is approximately 87%.

The actual enrolment of the trial was 88 subjects, and all 88 subjects received at least one dose of study drug. With the same two-stage group sequential testing approach, the null hypothesis can be rejected if 20 subjects in the interim analysis after 56 subjects, or 26 subjects in the primary analysis after 88 subjects, show a confirmed PR or CR according to RECIST 1.1. The corresponding nominal p-values of the exact binomial test are 0.0045 and 0.0211, respectively. The resulting overall probability of reaching a positive result in the interim or primary analysis under the null hypothesis assumption of an ORR $\leq$ 20% is $\leq$ 0.0223.

Justification of the assumption of a response rate of 35% to define the trial sample size

Data presented at the International Association for Study of Lung Cancer annual meeting in 2013 suggest the blocking of PD-L1 through the administration of anti-PD-L1 monoclonal antibody has antitumor activity in tumors whose micro-environment express PD-L1. Indeed, a response rate of 83% was reported (5/6) in patients where more than 10% of the immune-infiltrating cells were PD-L1 positive. In addition, the expression of PD-L1 at the surface of the tumor has also been correlated with clinical activity of anti-PD-L1. Robust expression of PD-L1 in the immune-infiltrating cells of MCC has been reported in the literature (1) and confirmed by the Sponsor.

On these grounds, the expression of PD-L1 by the immune infiltrating cells as well as the expression of PD-L1 at the surface of MCC constitute a very strong rationale for the evaluation of avelumab in that disease. Extrapolation of the response rate observed with the Genentech PD-L1 (up to 80% in PD-L1+ tumor cells and up to 80% in patients with more than 10% of the tumor infiltrating cells expressing PD-L1) suggest that a clinically significant overall response rate can be expected from anti PD-L1. Since a response rate of approximately 20% has been reported in several tumor types where only a fraction of the tumors express PD-L1 at the surface of the tumors or at the surface of immune infiltrating cells, it is considered that an expected response rate of 35% is a reasonable hypothesis.

8 Overview of Planned Analyses

This SAP covers the analyses for efficacy and safety based on the data cut-off dates for the interim and final analyses. Statistical analyses will be performed using cleaned eCRF data gained until a clinical cut-off date. It also includes detailed analysis for pharmacokinetics and immunogenicity data analysis.

For the interim futility analysis, the clinical cut-off date will be the date 3 months after start of study treatment of the 20th subject; for the interim efficacy analysis, the clinical cut-off date will be the date 6 months after start of study treatment of the 56th subject; for the primary analysis, the clinical cut-off date will be 6 months after start of study treatment of the last subject enrolled in the trial.

The Safety Monitoring Committee (SMC) will be responsible for reviewing the interim futility and efficacy analysis as well as the primary analysis. A SMC data report plan will be created separately to specify what analysis results will be distributed to SMC for review for each of the

interim analyses and the primary analysis. There will be no unplanned interim analyses, with the exception of the ongoing safety monitoring during the trial.

Further separate SAPs will cover the following analyses:

CCI [redacted]

8.1 Sequence of Analysis

The following analyses will be performed during this trial:

- Interim futility analysis: will be performed 3 months after start of study treatment of the 20th subject. Only the first 20 subjects at the clinical cut-off date will be included in the primary endpoint and secondary efficacy endpoint analysis; ; and all subjects who received at least one dose of treatment up to the cut-off date will be included for analysis of safety.
- Interim efficacy analysis: will be performed 6 months after start of study treatment of the 56th subject. The analysis will be performed on the ITT population and will include subjects with minimum follow up of 6 weeks, 12 weeks and 24 weeks beyond day 1.
- The primary analysis: will be performed 6 months after the accrual of the last subject. The primary analysis will be performed on the ITT population.
- An exploratory analysis will be conducted 12 months after the accrual of the last subject.

The date and time of the first study treatment will be used to select the first 20 and 56 subjects for the interim futility and interim efficacy analysis, respectively. There will be a partial database lock for both interim analyses and the primary analysis.

8.2 Interim Analysis

The interim futility analysis will include the primary endpoint analysis, secondary efficacy endpoint analysis, summary of baseline characteristics, treatment exposure and the safety analysis as outlined in this SAP. If no unconfirmed response according to RECIST 1.1 is seen in the first interim analysis, the enrollment will be stopped until the SMC makes a recommendation as to whether the trial should continue.

The interim efficacy analysis will comprise a full evaluation of all efficacy and safety endpoints as described in this SAP. If the efficacy goals are met at the interim efficacy analysis, enrollment will continue to the planned full number of subjects in order to collect further data on the primary and secondary endpoints.

8.3 Final Analysis

All planned analyses outlined in this SAP will be performed for the primary analysis. A partial database lock will be performed for the primary analysis and the last planned SMC meeting will

be held thereafter. The full DB lock will take place one year after the last subject receives the last dose of avelumab.

In addition, no database can be locked until this SAP has been approved.

Subject follow-up for progression and survival will continue until 1 year after the last subject receives the last dose of avelumab.

9 Changes to the Planned Analyses in the Clinical Trial Protocol

The statistical methods as described in the protocol of the trial were adopted. Additional statistical analyses and summary performed that are not included in the CTP are summarized in details in Section 5.1 Changes to Previous Version.

10 Analysis Sets

Screening Analysis Set

The screening analysis set includes all subjects who signed the informed consent.

Intent-to-Treat Analysis Set

For the interim futility and the interim efficacy analysis, the intention-to-treat (ITT) analysis set includes the first 20 and 56 subjects who received at least one dose of trial treatment, respectively. For the primary analysis, the ITT analysis set includes all subjects who received at least one dose of trial treatment.

The ITT analysis set will be used for all analyses of efficacy, and health-related quality of life and will be the primary population for decision-making.

Safety Analysis Set

The Safety analysis set includes all subjects who received at least one dose of trial treatment at the clinical cut-off date. The Safety set will be used for all safety analyses.

Per Protocol Analysis Set

The Per-Protocol (PP) analysis set is a subset of the ITT analysis set and includes all ITT subjects who meet all of the following criteria:

- Measurable disease per RECIST 1.1 and IERC assessment
- Distant metastatic disease
- Histologically proven MCC with confirmation of the diagnosis by immuno-histochemistry detection of CK20 (or other appropriate cytokeratin expression such as pancytokeratin,

AE1/AE3, or Cam5.2) according to the assessment documented in the "Disease History" eCRF page

- Have progressed after 1 line of chemotherapy that was administered for the treatment of distant metastatic MCC as defined in Section 5.3.1 Inclusion Criteria in the protocol
- Evaluable patients, defined as having at least one post-baseline tumor assessment with absence of non-assessable status

The PP analysis set will be used for additional sensitivity analyses for the primary and secondary efficacy endpoints. If the PP analysis set includes at least 95% of subjects in the ITT analysis set, additional efficacy analyses on the PP analysis set will be omitted as the differences in the results based upon these two analysis sets are expected to be negligible.

Additional Subgroup Analysis Sets

The below subgroup analyses may be performed on primary and key secondary efficacy endpoints. Since the study is not powered for any subgroup analysis, CCI [redacted].

The following subgroups will be defined:

- Age Group 1:
  - Age < 65 years (Reference)
  - Age ≥ 65 years
- Age Group 2:
  - Age ≤ Median (Reference)
  - Age > Median

CCI [redacted]

CCI [redacted]

- Gender
  - Male (Reference)
  - Female
- Pooled Region:
  - North America
  - Europe (Reference)
  - Rest of the World
- Time from initial diagnosis to study entry:
  - <= 1 year (Reference)
  - > 1 year and <= 2 years
  - > 2 years
- ECOG PS at baseline
  - ECOG PS 0 (Reference)
  - ECOG PS 1
- Disease Burden
  - Baseline SLD ≤ Median (Reference)
  - Baseline SLD > Median
- Site of primary tumor
  - Skin
  - No-skin (Reference)
- Presence of visceral metastases at baseline . Target or non-target lesions that are categorized as “metastatic” per investigator assessment are classified as visceral metastases.
  - Present
  - Absent (Reference)
- Number of prior systemic therapies for metastatic disease
  - 1 Prior Therapy (Reference)
  - ≥ 2 Prior Therapies
- Number of prior systemic therapies for metastatic disease or locally advanced therapies
  - 1 Prior Therapy (Reference)

- ≥ 2 Prior Therapies

11 General Specifications for Statistical Analyses

All efficacy analyses will be performed on the ITT population considering subjects as treated. Selected efficacy analyses will be repeated for the PP population only if the size of this population is less than 95% of the ITT population.

The Safety population will be considered for safety analyses.

Data handling after cut-off date:

Data after cut-off do not undergo the cleaning process and will not be displayed in any listings or used for summary statistics, statistical analyses or imputation. The only exceptions are the date of death and the date last known to be alive from the eCRF form "Subject Status / Survival Follow-up".

Pooling of centers:

Because of the high number of participating centers and the anticipated small number of subjects receiving trial treatment in each center, data will be pooled across centers. The factor center will not be considered for subgroup analyses.

Definition of baseline:

The last non-missing measurement prior to the first study drug administration will be used as the baseline measurement, if not otherwise specified. Additionally, ECG baseline will be derived from the visit where at least HR, PR, QRS, and QT are not missing. QTcF will be derived based on HR and QT. The average of the replicate measurements (prior infusion and post infusion) should be used in deriving baseline values for each parameter.

Definition of study day/treatment day:

Treatment day is defined relative to the first study drug administration.

Treatment day 1 defines the first study drug administration date, the day before is defined as treatment day -1 (no treatment day 0 is defined).

Definition of on-treatment period:

On-treatment period is defined as the time from the first study drug administration to the last drug administration date + 30 days or earliest date of subsequent anti-cancer drug therapy – 1 day, whichever occurs first, unless otherwise stated. If the earliest date of subsequent anticancer drug therapy was a partial date and only day was missing, it was imputed as the last day of the month. If both day and month were missing, no imputation was performed. The imputed date was used for defining on-treatment period as well as confirming CCI

Unscheduled visits:

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.

Data collected after re-initiated treatment:

Data collected after re-initiation of treatment will not be included for safety and efficacy analyses except for overall survival and disposition.

Visit Window:

The assignment of visit window is described in Table 1 for the purpose of by-visit analyses of safety data:

- Baseline will be derived as described above.
- Both scheduled and unscheduled assessments are included for visit windowing.
- No visit windowing will be performed at Discontinuation, End-of-Treatment, or Safety Follow-up visits for laboratory, vital sign, ECG data, and 2hr post dose assessment on Week 1 Day 1 for ECG data. Instead, the earliest non-missing observation among the unscheduled or scheduled assessments for each visit (Discontinuation, End-of-Treatment, and Safety Follow-up) will be used for the analysis. For 2hr post dose assessment on Week 1 Day 1 ECG data, the earliest non-missing observation on Week 1 Day 1 will be used for the analysis.
- If there are multiple assessments for any specified visit and some of them are from scheduled visits with non-missing assessment results, the assessment from scheduled visit that is closest to the planned study day will be used for analysis.
- If there are multiple assessments for any specified visit and none of them are from scheduled visits, the assessment with non-missing results and closest to the planned study day will be used for analysis.
- If there are two or more unscheduled assessments with non-missing results and the same distance to the planned study day, the assessment prior to the planned study day will be used in deriving visit window. For example, if the lab assessment was done on both study day -1 and 1, then the assessment on -1 study day will be used for visit windowing.
- For ECG assessment associated with study drug dose, only assessments where time point (prior to infusion or 2 hr after infusion) are not missing will be considered for the analysis.

Table 1 Visit Window Definition for Safety Assessment

Assigned Study Day (Inclusive)		Planned Study Day (AWTARGET)	Analysis Visit (N) (AVISITN)	Analysis Visit (AVISIT)	Assessment
From (AWLO)	To (AWHI)				
~	1		1	Baseline	Lab, Vital Sign, ECG

1	1	1	2	Week 1 Day 1*	ECG
5	11	8	3	Week 2 Day 5-11	Core Serum Chemistry
12	18	15	4	Week 3 Day 12-18	Lab, Vital Sign, ECG
19	25	22	5	Week 4 Day 19-25	Core Serum Chemistry
26	32	29	6	Week 5 Day 26-32	Lab, Vital Sign, ECG
33	39	36	7	Week 6 Day 33-39	Core Serum Chemistry
40	50	43	8	Week 7 Day 40-50	Lab, Vital Sign, ECG
51	64	57	10	Week 9 Day 51-64	Lab, Vital Sign, ECG
65	78	71	12	Week 11 Day 65-78	Lab, Vital Sign, ECG
79	92	85	14	Week 13 Day 79-92	Lab, Vital Sign, ECG
93	106	99	16	Week 15 Day 93-106	Lab, Vital Sign
107	120	113	18	Week 17 Day 107-120	Lab, Vital Sign
121	134	127	20	Week 19 Day 121-134	Lab, Vital Sign
107	148	127	20	Week 19 Day 107-148	ECG
135	148	141	22	Week 21 Day 135-148	Lab, Vital Sign
149	162	155	24	Week 23 Day 149-162	Lab, Vital Sign
163	176	169	26	Week 25 Day 163-176	Lab, Vital Sign
149	190	169	26	Week 25 Day 149-190	ECG
177	190	183	28	Week 27 Day 177-190	Lab, Vital Sign
191	204	197	30	Week 29 Day 191-204	Lab, Vital Sign
205	218	211	32	Week 31 Day 205-218	Lab, Vital Sign
191	232	211	32	Week 31 Day 191-232	ECG

* Only applies to 2 hr post dose.

Calculation of duration:

Duration will be calculated by the difference of start and stop date + 1 (e.g. survival time (days) = date of death – date of first study treatment + 1), unless otherwise specified.

Conversion of days to months or years:

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

Significance level:

The overall significance level is 2.5% one-sided. The confirmatory statistical test for the primary efficacy endpoint analysis is described in Section 16.1 along with procedures for controlling the overall type I error rate. All other statistical analyses performed on the secondary and other endpoints defined in this SAP are exploratory.

Confidence intervals will be two-sided with a confidence probability of 95%, if not otherwise specified.

Presentation of continuous and qualitative variables:

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

Qualitative variables will be summarized by counts and percentages. Unless otherwise stated 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.

All data recorded during the trial will be presented in individual data listings performed on the Safety analysis set.

Handling of missing data:

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

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

Missing statistics, e.g. when they cannot be calculated, should be presented as "nd". For example, if n=1, the measure of variability (StdDev) cannot be computed and should be presented as "nd".

Handling of incomplete dates:

Incomplete dates for disease history (initial diagnosis date, date of documented, locally advanced or metastatic disease diagnosis) 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.

Incomplete AE-related dates will be handled as follows:

- In case the onset date is completely missing or the onset is in the same month and year (if the day is missing) as the start of trial treatment then the onset date will be replaced by the minimum of start of trial treatment and AE resolution date.
- In all other cases the missing onset day or missing onset month will be replaced by 1.
- Incomplete stop dates 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 subject'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 date of cut-off outcome of AE is ongoing at cut-off.
- Further information after cut-off (like fatal outcome) might be taken from Safety database and included separately into CTR.

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 trial treatment.
- If the day of medication date is missing, but the month and year are equal to the start of trial treatment, then the medication date will be replaced by the start of trial treatment. For example, if the medication start date is --/JAN/2015, and trail treatment start date is 15/JAN/2015, then the imputed medication start date will be 15/JAN2015.
- If both the day and month of medication start date are missing but the start year is equal to the start of trial treatment, then the medication date will be replaced by the start of trial treatment. For example, if the medication start date is --/---/2014, and trial treatment start date is 19/NOV/2015, 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 subject'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.

For reporting conventions, mean and median should generally be displayed to 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. E.g. 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.

All statistical analyses will be performed using SAS® Version 9.2 or higher, or R (www.r-project.org), version 2.15.2 or higher.

12 Trial Subjects

The subsections in this section include specifications for reporting subject disposition and treatment/trial discontinuations. Additionally procedures for reporting protocol deviations are provided.

12.1 Disposition of Subjects and Discontinuations

The following will be summarized overall:

Subject disposition:

- Number of subjects screened (i.e. subjects who gave informed consent)
- Number of subjects per reason of non-eligibility at end of screening visits
- Number of subjects in the ITT population
- Number of subjects in the Safety population
- Number of subjects in the PP population
- Number of subjects still on treatment
- Number of subjects off-treatment overall and by the main reason for discontinuation
- Number of subjects who discontinued the treatment but are still in follow-up
- Number of subjects who discontinued the trial overall and by the main reason for discontinuation

In addition, the following will be summarized:

- Number and percentage of subjects by region region (Europe, North America, Latin America, Asia, Australia), by country within region and by center for each of the analysis set (Screening, ITT (Safety) and PP analysis set).

12.2 Protocol Deviations

Protocol deviations will be listed and summarized based on the ITT.

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

- subjects that are dosed on the study despite not satisfying the inclusion criteria;
- subjects that develop withdrawal criteria while on the study but are not withdrawn;
- subjects that receive the wrong treatment or an incorrect dose;
- subjects that receive an excluded concomitant medication;

- Deviation from GCP.

Important protocol deviations will be determined for all subjects by either medical review processes or programming based on the inclusion/exclusion criteria or other criteria presented in the protocol.

All important protocol deviations should be documented in CDISC datasets whether identified through sites monitoring, medical review or programming. Important Protocol Deviations to be identified by programming as well as all clinically Important Protocol Deviations are listed and described in Appendix I. They will be presented in the summary table and in a data listing.

13 Demographics and Other Baseline Characteristics

Analysis sets: ITT analysis set / Safety analysis set

13.1 Demographics

Demographic characteristics will be summarized as follows:

- Demographic characteristics
  - Gender: Male, Female
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Unknown
  - Ethnic origin: Hispanic/Latino (Yes/No), Japanese (Yes/No)
  - Age (years): summary statistics
  - Age categories : < 65 years, ≥ 65 years
    - 65-74, 75-84, ≥ 85
  - Height (cm)
  - Weight (kg)
  - Body Mass Index (BMI) (kg/m^2)
  - Pooled Region:
    - North America
    - Europe
    - Rest of the World
  - Geographic Region:
    - North America
    - Latin America
    - Western Europe
    - Eastern Europe
    - Australia
    - Asia
  - Eastern Cooperative Oncology Group (ECOG) Performance Status

Specifications for computation:

- Age [years]:
  - (date of given informed consent - date of birth + 1) / 365.25
  - In case of missing day of birth, day is replaced by 15
  - 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 purpose.

- BMI (kg/m^2) = Weight(kg)/[Height(m)]^2.

13.2 Medical History

Medical history will be coded using the most current available version of Medical Dictionary for Regulatory Activities (MedDRA). Medical history will be summarized as the numbers and percentages of subjects by MedDRA preferred term (PT) as event category and MedDRA primary system organ class (SOC) as summary category. Each subject 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.

13.3 Other Baseline Characteristics

13.3.1 Disease Characteristics

Information on disease characteristics collected on "Disease History" eCRF page will be summarized as follows:

- Site of primary tumor
- Tumor size
- Peripheral margin status
- Deep margin status
- Lymphovascular invasion
- Extracutaneous extension
- Time since initial diagnosis (months), defined as (date of the first study treatment – date of initial diagnosis)/30.4375
- Time since first metastatic disease (months), defined as (date of the first study treatment – date of the first occurrence of metastatic disease)/30.4375
- Time since first locally advanced disease (months), defined as (date of the first study treatment – the date of the first occurrence of locally advanced disease)/30.4375

- Time since last progression of disease prior to study entry (months), defined as (date of the first study treatment –date of the last progression of disease)/30.4375
- TNM classification at initial diagnosis
- TNM classification at study entry
- Additional clinically relevant factors: Depth (Breslow), Mitotic index, Tumor-infiltrating lymphocytes, tumor growth pattern, Presence of second malignancy, CCI [redacted]

Baseline characteristics with respect to ECOG, vital signs, physical examinations, ECG and hematology/biochemistry will be part of Section 17 (Safety Evaluation).

13.3.2 Prior Anti-Cancer Therapies

The prior anti-cancer therapies are collected under the "Prior Anti-Cancer Drug Therapies Details", "Prior Anti-Cancer Radiotherapy Details" and "Prior Anti-Cancer Surgeries Details" eCRF pages.

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

- Subjects with at least one type of prior anti-cancer treatment
- Subjects with at least one prior anti-cancer drug therapy for metastatic disease
- Subjects with at least one prior anti-cancer radiotherapy
- Subjects with at least one prior anti-cancer surgery

Prior anti-cancer drug therapy will be summarized as follows:

- Number of subjects with at least one prior anti-cancer drug therapy for metastatic disease
- Number of any prior anti-cancer therapy regimens for metastatic disease: missing / 1 / 2 / 3 / ≥4
- Number of prior anti-cancer therapy lines for locally advanced disease only: missing / 1 / 2 / 3 / ≥4. If the disease setting is locally advanced, it will be counted into therapy lines for locally advanced disease.
- Number of prior anti-cancer therapy lines for metastatic disease only: missing / 1 / 2 / 3 / ≥4. If the disease setting is metastatic, it will be counted into therapy lines for metastatic disease.
- Type of prior anti-cancer therapy: Chemotherapy for metastatic disease / Chemotherapy for non-metastatic disease / Hormonal therapy / Antibody therapy / Kinase inhibitors / Vaccines / Other
- Disease setting: Neo-Adjuvant / Adjuvant / Metastatic /Locally advanced
- Best response: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive Disease (PD) / Not assessable / Unknown / Not applicable. Best response is derived from the last treatment regimen

The prior anti-cancer drugs will also be extensively detailed with the number and percentage of subjects by the drug class and PT in a table. A subject 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.

The listings of prior anti-cancer treatments and procedures will also be provided: a) listing of prior anti-cancer drug therapies for non-metastatic and for metastatic disease, b) listing of prior anti-cancer radiotherapy and c) listing of prior anti-cancer surgeries. These will include the subject identification number, and all the relevant collected data-fields on the corresponding eCRF pages.

14 Previous or Concomitant Medications/Procedures

Analysis sets: ITT analysis set / Safety analysis set

14.1 Prior and Concomitant Medications/Procedures

Concomitant medications are medications, other than trial medications, which are taken by subjects any time during the on-treatment period. **Previous medications** are medications, other than trial medications and pre-medications for trial drug, which are taken and stopped before first administration of trial drug.

Prior and concomitant medications will be summarized from the “Concomitant Medications Details” eCRF page.

In cases where the date values do not allow unequivocal allocation of a medication to concomitant (as opposed to previous) medication the medication will be considered as concomitant medication.

Summary of prior and concomitant medications will include the number and percentage of subjects by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term based on the WHO-DD dictionary current version. A subject will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. If any prior or concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes. The summary tables will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used.

A listing of concomitant medications will be created with the relevant information collected on the “Concomitant Medications Details” eCRF page. All concurrent procedures, which were undertaken any time during the on-treatment period, will be summarized according to the CRF page “Concomitant Procedures Details”.

Number of subjects with concurrent procedures will be tabulated overall and by type of procedure (as classified by medical review)

A listing of concurrent procedures will be created with the relevant information collected on the "Concomitant Procedures Details" eCRF page.

In addition, a listing of premedication will be created with the relevant information collected on the "Premediation Details" eCRF page.

14.2 Subsequent Anti-Cancer Therapies/Procedures

Anti-cancer treatment after discontinuation will be provided in a data listing with data retrieved from "Anti-Cancer Treatment after Discontinuation Details", "Radiotherapy after Discontinuation", and "Surgery after Discontinuation" eCRF pages. The earliest date of start of new anti-cancer drug therapy will be used for the definition of the on-treatment period.

15 Treatment Compliance and Exposure

Analysis sets: ITT analysis set / Safety analysis set

All dosing calculations and summaries will be based on "avelumab Administration Details" CRFs pages.

Subjects will receive an IV infusion of avelumab at a dose of 10 mg/kg (over the duration of 1 hour) once every 2 weeks (one cycle). Analysis of exposure will be based on the calculated actual dose levels (total dose administered/weight, mg/kg). The last available weight of the subject on or prior to the day of dosing will be used.

In case of missing total dose, the dose level as entered in the eCRF will be used.

The duration of therapy (in weeks) during the study is defined as:

$$\text{duration} = \left(\frac{\text{date of last dose} - \text{date of first dose} + 14}{7}\right)$$

The cumulative dose (mg/kg) per subject in a time period is the sum of the actual dose levels that the subject received within that period (i.e. total dose administered (mg) / weight (kg)).

Each cycle is defined by a 2-week period. The dose intensity and the relative dose intensity will be calculated for each subject across all cycles. The dose intensity (mg/kg/cycle) per cycle is defined as

$$\text{dose intensity} = \left(\frac{\text{Cumulative dose}}{(\text{duration of therapy (in weeks)})/2}\right)$$

The relative dose intensity is defined as the actual dose intensity divided by the planned dose per cycle.

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

- Duration of therapy (weeks)
- Total number of infusions received
- Cumulative dose of therapy (mg/kg)
- Dose intensity (mg/kg/cycle)
- Relative dose intensity of therapy (%)

A listing of treatment exposure and compliance will also be created to summarize the information listed above for each subject.

Dose Reduction

A dose reduction is defined as actual non-zero dose < 90% of the planned dose. Number of subjects with at least one dose reduction as well as a breakdown of dose reductions (1 / 2 / 3 / ≥4) will be summarized.

Dose Delays

Delays will be derived based on infusion date and will be grouped into the following categories based on the deviation of the actual to the planned treatment administration day (relative to the previous treatment administration date):

- No delay (including 1-2 days delays)
- 3-6 days delay
- 7 or more days delay

For example, if one subject receives the study drug on day 1, then the next study drug administration date will be on day 15; however, if the subject receives the study drug at day 16 or 17, this is considered as 'no delay'.

The categorization is based on the maximum length of delay, i.e. the worst case of delay if subjects have multiple dose delays will be summarized.

A listing of study drug administration will be created with the information collected on the "Avelumab Administration Details" eCRF page.

16 Endpoint Evaluation

The subsections in this section include specifications for analyzing clinical trial endpoints specified in the Clinical Trial Protocol to meet the trial objectives, as well as any endpoints not identified in the Clinical Trial Protocol.

16.1 Primary Endpoint Analyses

Analysis sets: ITT analysis set, PP analysis set

The primary endpoint is the confirmed BOR to therapy in terms of either CR or PR according to RECIST 1.1 (Appendix I), as determined by an IERC. The tumor response will be based on the IERC assessment of overall response at each time point. In case of different dates of scans within the same tumor assessment visit, the earliest scan date should be used as the date of tumor assessment.

The confirmed BOR is defined as the best response obtained among all tumor assessment visits after start of trial treatment until documented disease progression, excluding assessments after start of subsequent anticancer therapy, taking requirements for confirmation into account as detailed below. The confirmed ORR will be defined as the proportion of subjects with a confirmed BOR of PR or CR.

A separate Imaging Data Management Plan and Data Transfer Plan will be created to summarize the details of the data structure and data delivery schedule of IERC assessment results.

The following requirement is taken into account for confirmation of response:

- PR or CR needs to be confirmed at a subsequent tumor assessment, preferably 6 weeks after the initial observation of response and according to the normal 6-week assessment schedule but no sooner than 5 weeks after the initial documentation of CR or PR.
- PR can be confirmed at an assessment later than the next assessment after the initial documentation of PR.
- The minimum duration for a BOR of SD is defined as at least 6 weeks after start of study treatment.

Table 2 summarizes the derivation rules described by Eisenhauer, et al. for the BOR when conformation from subsequent assessment is needed (5).

Table 2 Best overall response when confirmation of CR/PR is required

Overall response 1st time point	**Overall response subsequent time point**	**Best overall response (BOR)**
CR	CR	CR
CR	PR	SD, PD or PR[a]

Overall response 1st time point	Overall response subsequent time point	Best overall response (BOR)
CR	SD	SD provided minimum criteria for SD duration met; otherwise, PD
CR	PD	SD provided minimum criteria for SD duration met; otherwise, PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise, PD
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease,
NE = in-evaluable.

[a] If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the subject had PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.

The analysis of the primary endpoint will be conducted for the ITT analysis set. The test statistic of the exact binomial test will be calculated and compared with the thresholds given in Section 7 to determine whether the null hypothesis of an ORR $\leq$ 20% can be rejected at the interim analysis, or, if not rejected at the interim, at the primary analysis. The thresholds given in Section 7 were chosen based on the binomial distribution:

- The binomial probability of 20 or more responses in 56 subjects under a true response rate of 20% is 0.0045.
- In case of an ITT population of 88 patients, the binomial probability of 26 or more responses in 88 subjects under a true response rate of 20% is 0.0211.
- In the case of 88 patients, the probability of rejecting the null hypothesis in the interim or the primary analysis, given a true respone rate of 20%, is 0.0211, obtained via the exact calculation described in Jennison and Turnbull(1), Section 12.1.2, Formula 12.5-12.7. The associated confidence interval would not include 20%. If the ITT population would not be 88 patients, the calculation of the rejection boundaries would still be by Jennison and Turbull (1).

Therefore, the overall Type I error rate is controlled at a level of 2.5% (one-sided).

As some subjects included in the interim analysis will be followed up until the time of the primary analysis, some subjects' BOR might change from non-response at the interim analysis to response at the primary analysis. As the parameter of interest is the response at the time of the primary analysis, the actual type I error rate of this study might be slightly smaller than designed.

The Clopper-Pearson method will be used to calculate the two-sided CI for the ORR at both the interim and the primary analysis (exact confidence interval for a binomial proportion as computed by default by the FREQ procedure using the EXACT option). In order to account for the two-stage group sequential design as applied in this study, the repeated CI method according to Jennison and Turnbull (1), Section 12.1.4, will be used to construct the two-sided CI for the interim and the primary analysis. For the interim analysis, the nominal p-value threshold of the exact binomial test of 0.0045 will be applied to calculate the two-sided CI; for the primary analysis, a p-value threshold of 0.0205, obtained as the difference between the overall type I error rate of 0.025 and the interim p-value threshold of 0.0045, will be applied to calculate the two-sided CI. This approach results in a two-sided 99.1% CI and 95.9% CI at the interim and the primary analysis, respectively.

The repeated CIs constructed in this way are simultaneous CIs for the ORR, i.e. the probability that both CIs contain the true response rate Θ is at least 95% for all possible values of Θ. They are slightly conservative in relation to the exact group sequential test procedure: In the case of exactly 25 responses in 84 subjects, the null hypothesis will be rejected, however the respective 95.9% CI for the ORR is given by (0.199, 0.412) and does not exclude the value of 0.2. The nominal p-value from the exact binomial test will also be reported.

In addition, the unadjusted 95% CIs of the ORR will also be calculated with the Clopper-Pearson method at the interim and the primary analysis in order to provide comparable results based on the PP analysis and subgroup analysis.

In addition, the number and percentage of subjects with BOR of CR, PR, SD, PD, non-CR/non-PD, and NE will be tabulated. Patients with BOR of NE will be summarized by reason for having NE status. The following reasons will be used:

- No post-baseline assessments
- All post-baseline assessments have overall response NE
- New anticancer therapy started before first post-baseline assessment
- SD of insufficient duration (tumor assessment <6 weeks after start date without further evaluable tumor assessment)

The primary endpoint analysis will be repeated for the PP analysis set as a sensitivity analysis if the PP analysis set includes less than 95% of subjects in the ITT analysis set. The ORR will be calculated along with the two-sided 95% Clopper-Pearson CIs.

Accoring to Section 5.3.1 Inclusion Criteria of the clinical protocol, all subjects who are enrolled to the study to receive at least one dose of avelumab should have meansurable disease at baseline by RECIST 1.1 according to investigator assessments. However, it is possible that a subject in the ITT analysis set could be found to have no measurable disease at baseline per IERC assessment, and would not have a response in the primary endpoint analysis. Therefore the primary endpoint analysis will be repeated on subjects with measurable disease at baseline per RECIST 1.1 by the IERC as a sensitivity analysis. The ORR will be calculated along with the two-sided 95% Clopper-Pearson CIs.

The response at each scheduled tumor assessment and the BOR will be listed for each subject.

16.2 Secondary Endpoint Analyses

Analysis set: ITT analysis set, PP analysis set

The secondary endpoints include:

- Duration of response (IERC);
- Progression free survival (PFS) time according to RECIST 1.1 as determined by an IERC,
- Overall survival (OS) time;
- Response status according to RECIST 1.1 at 6 and 12 months after start of study treatment.

For imputing missing parts of dates for the secondary efficacy endpoints the missing day in a date will be imputed as the 15th of the month, if month and year are documented. If the imputation is earlier than the date of first study treatment, then the day of first study treatment is taken. In all other cases missing or incomplete dates will not be imputed.

16.2.1 Duration of Response (IERC)

The duration of response will be calculated for each subject with a confirmed response according to RECIST 1.1 as the time from first observation of response until first observation of documented disease progression or death when death occurs within 12 weeks of the last tumor assessment, whichever occurs first. Details on determination of the first disease progression date are specified in the IERC charter. For subjects with a confirmed response but neither documented disease progression nor death within 12 weeks after the last tumor assessment, as of the cut-off date for the analysis, the duration of response will be censored at the date of the last adequate tumor assessment.

Duration of response = (date of PD or death – date of objective response + 1)/30.4375 (months).

The Kaplan-Meier method will be used to estimate parameters for duration of response. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. In particular, the proportion of duration of responses at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (2) and CIs for the survival function estimates at the above defined time points will be derived using the log-log transformation according to Kalbfleisch and Prentice (3) (CONFTYPE = loglog default option in SAS PROC LIFETEST). The estimate of the standard error will be computed using Greenwood's formula.

16.2.2 Progression Free Survival (IERC)

The PFS time, according to the RECIST 1.1, will be defined as time from first administration of trial treatment until first observation of PD or death when death occurs within 12 weeks of the last tumor assessment or first administration of trial treatment (whichever is later). Details on determination of the first disease progression date are specified in the IERC charter.

Specific censoring rules are defined in Table 3 and Table 4. Last adequate tumor assessment is defined as the last tumor assessment result that is not "NE" or "NA".

PFS time (in months) is defined as:

(date of PD or death - date of the first dose of study treatment + 1)/30.4375 (months).

For subjects with both of the following conditions fulfilled, the general censoring rules apply as presented in Table 3.

- Baseline tumor assessment
- At least one adequate tumor assessment after start of treatment

Table 3 General Censoring Rules for Progression-free Survival

			Date of event / censoring	Censoring
PD according to RECIST as assessed by the IERC			Date of PD	No
No PD	Death within 12 weeks after last tumor assessment or date of first study treatment (whichever is later)	Yes	Date of death	No
		No	Date of last adequate tumor assessment	Yes

IERC=Independent Endpoint Review Committee, PD=progressive disease.

For subjects with no tumor assessment at baseline or no tumor response assessment post-baseline, the special censoring rules as presented in Table 4, which overrule the general censoring rules, apply.

Table 4 Special Case Censoring Rules for Progression-free Survival

			Date of event / censoring	Censoring
No tumor assessment at Baseline	Death within 12 weeks after first study treatment	Yes	Date of death	No
		No	Date of first study treatment	Yes
No tumor response assessment post Baseline	Death within 12 weeks after first study treatment	Yes	Date of death	No
		No	Date of first study treatment	Yes

The analysis of PFS will be performed using a Kaplan-Meier method with the same approach as for duration of response described in Section 16.2.1. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. In particular, the proportion of PFS at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs.

The PFS time will also be presented in a subject listing.

16.2.3 Overall Survival

The overall survival (OS) time will be defined as the time from first administration of trial treatment until the date of death.

For subjects who are still alive at the time of data analysis or who are lost to follow up, OS time will be censored at the last recorded date that the subject is known to be alive (date of last contact, last visit date, date of last trial treatment administration or date of last scan, whichever is the latest) as of the data cut-off date for the analysis.

OS = (date of death – date of the first dose of study treatment + 1)/30.4375 (months).

The analysis of OS time will be performed using a Kaplan-Meier method with the same approach as for duration of response described in Section 16.2.1. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. In particular, the proportion of overall survival at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs.

The OS time will also be presented in a subject listing.

16.2.4 Response Status at 6 and 12 Months after Start of Study treatment

The response status at 6 and 12 months after start of study treatment according to RECIST 1.1 (as determined by the IERC) will be determined. A subject will be considered to be in response at the given time point (6 or 12 months after start of study treatment) if the subject had a documented response (PR or CR) prior to that time point, and did not have any of the following events up to the given time point:

- death
- documented disease progression according to the RECIST 1.1,
- lost to follow-up
- withdraw consent

The proportion of subjects with response at 6 and 12 months after start of study treatment will be calculated as the proportion of subjects in response among all subjects that have started study treatment at least 6 (12) months prior to the time of the analysis, respectively. Clopper-Pearson 95% CIs will be reported as well.

In addition, durable response rate will be summarized for the ITT analysis set. Durable response rate (DRR) is defined as the proportion of subjects with an objective response in terms of CR or PR according to RECIST 1.1, as determined by the IERC, with a duration of at least 6 months. Clopper-Pearson 95% CIs will be reported for the DRR as well.

16.3 Analysis on Subgroups

Analysis on subgroups as defined in Section 10 will be performed for the primary and key secondary efficacy endpoints. In addition, the odds ratio and its 95% CI corresponding to the ORR of each subgroup relative to the reference subgroup will be calculated, and the p-value from Fisher's exact test will be provided to assess the association between each subgroup factor and ORR. Furthermore, to compare PFS time and OS time within each subgroup factor, the hazard ratio and its 95 % CI of each subgroup relative to the reference subgroup will be calculated using Cox proportional hazards model.

A forest plot will be created to graphically present ORR and its 95% CIs for the ITT analysis set and all the subgroups.

CCI [redacted]. In case of low number of subjects within a category (< 5 subjects, which is about 5% of the study population), subgroups will be pooled.

16.4 Other Endpoint Analyses

Analysis sets: ITT analysis set

16.4.1 Time to Response per RECIST 1.1

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

TTR (in weeks) = (data of first objective response – date of first dose of study treatment + 1)/7

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

CCI [redacted]

CCI

CCI

CCI

[redacted]

CCI

[redacted]

16.4.6 Duration of Follow-up

Kaplan-Meier analysis will be performed to estimate median time of duration of follow-up for PFS using the censoring rules (reverse censoring indicator) as specified in Table 5.

Table 5 Reverse Censoring Rules for Duration of Follow-up

Endpoint	Event		Date of event / censoring	Censoring
Duration of follow-up for PFS	First PD or Death	Yes	Time from first dose of study treatment to date of first PD or death	Yes
		No	Time from first dose of study treatment to last date known to be alive	No

Kaplan-Meier survival curves will be presented with median duration of follow-up and its two-sided 95% CI. In particular, the follow-up rate at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs.

CCI

[redacted]

16.5 Pharmacokinetic Data

The PK Analysis Set will consist of all subjects who receive at least one dose of Avelumab, and provide at least one measurable post-dose concentration.

16.5.1 Descriptive PK Analysis

Avelumab concentrations in serum will be listed in tables and descriptively summarized by treatment, day and nominal time using the number of non-missing observations (N), number of missing observations (NMiss) arithmetic mean (Mean), standard deviation (StdDev), coefficient of variation (CV%), standard error of the mean (SEM), minimum (Min), median (Median) and maximum (Max). The pre-dose samples will be considered as if they had been taken simultaneously with the administration. Additional tables listing trough concentrations (Cmin) per cohort/day/administration will also be provided, and summarized descriptively using the aforementioned statistics.

For the calculation of descriptive statistics, values as presented in the data listings will be used. Concentrations will be set to missing in summary tables if the value is reported as N.R. (No result). PK concentrations below the lower limit of quantification (<LLOQ) are taken as zero for descriptive statistics. Minimum and maximum will be presented to the same decimal precision as collected. Mean, median and GeoMean will be presented to one decimal place more than the precision of the data collected. SD and SEM will be presented to two decimal places more than

the precision of the data collected. The coefficient of variation (CV%), GeoCV and the 95% CI will be reported to 1 decimal place.

Drug concentration in serum for HAHA ever-positive subjects versus HAHA never-positive subjects will be descriptively summaried to evaluate the potential effect of HAHA on pharmacokinetics.

16.5.2 Population Pharmacokinetic Analysis

Sparse sampled PK profiles from study EMR 100070-003 will be analyzed jointly with data from CCI [redacted] by non-linear mixed effect approach, in order to describe the PK concentration time profile followed by multiple dose infusion of avelumab, to identify covariates explaining (part of) the between patient PK variability and to estimate the residual PK inter-individual variability. More details will be given in a separate Data Analysis Plan for Population Pharmacokinetic Analysis. The results will be reported separately.

16.5.2.1 Relation of Pharmacokinetics to Efficacy

The exposure-response analysis will include the subjects of the ITT population from clinical trial 003 Part A as defined above for the primary analysis. The patients will be classified as responder or non-responder based on BOR according to RECIST 1.1 (responder: CR and PR; non-responder: stable disease [SD], progressive disease [PD], non-evaluable [NE]). The results will be reported separately.

CCI [redacted]

The influence of the exposure parameter CCI [redacted] on response will be explored graphically by depicting response (no/yes) versus exposure and applying a spline function to the data.

The relation between exposure and response will then assessed using logistic regression.

CCI [redacted]

Based on this model, the probability of being responder will be depicted versus CCI [redacted]. A 95% confidence interval will the added to the graph.

In a next step, covariates will be inserted in the logistic regression model.

CCI

Significance of covariates CCI will be assessed using the Wald statistic (alpha=0.05). All covariates will be inserted in the logistic regression model simultaneously and insignificant covariates will be eliminated one by one (starting with the most insignificant covariate) until all remaining covariates are significant.

The odds ratios with 95% confidence intervals for the significant covariates will be calculated and graphically presented in a Forest plot.

CCI

CCI

All exposure-response calculations will be done using the software program R or other software.

16.6 Human Anti-human Antibody (HAHA)

Subjects will be categorized as either never positive or ever positive (a positive result at any time point, including baseline). Listings of all individual HAHA results from ever positive subjects will be prepared by time point.

Subjects with a positive HAHA result prior to treatment with Avelumab will be characterized as pre-existing. If the titer increases ≥8 fold while on Avelumab treatment, they will be additionally characterized as treatment-boosted. The summary table will include the total pre-existing incidence as well as the treatment-boosted incidence.

Subjects not positive prior to treatment with Avelumab and with at least one positive result in the HAHA assay will be characterized as treatment-emergent and further characterized for persistence of antibody response. Subjects will be categorized as transient positive (time between first and last positive result <16 weeks apart and a negative result at the most recent visit) or persistent positive (time between first and last positive result ≥16 weeks apart or a positive evaluation at the most recent visit). The summary table will include the total treatment-emergent incidence as well as the persistent and transient incidence. Table 6 below summarized the criteria of different HAHA result categories.

Table 6 Subjects Characterized based on HAHA Results

Category	Definition	Subjects at Risk (Denominator for Incidence)
Never positive	No positive results at any time point	Number of subjects with at least one valid result at any time point
Ever positive	At least one positive result at any time point	Number of subjects with at least one valid result at any timepoint
Pre-existing	A positive HAHA result prior to treatment with avelumab	Number of subjects with valid baseline result
Treatment boosted	A positive HAHA result prior to treatment with avelumab and the titer ≥ 8*baseline titer while on avelumab treatment	Number of subjects with valid baseline and at least one valid post-baseline result
Treatment emergent	Not positive prior to treatment with avelumab and with at least one positive post-baseline result	Number of subjects with at least one valid post-baseline result and without positive baseline results (including missing, NR)
Transient positive	If treatment emergent subjects have (a single positive evaluation, or duration between first and last positive result <16 weeks) and last assessment not positive.	Number of subjects with at least one valid post-baseline result and without positive baseline results (including missing, NR)
Persistent positive	If treatment emergent subjects have duration between first and last positive result ≥16 weeks or a positive evaluation at the last assessment	Number of subjects with at least one valid post-baseline result and without positive baseline results (including missing, NR)

If data permit (at least 5 ever positive or treatment emergent), subanalyses of immunogenicity responses on safety (TEAEs, TE irAEs, SAE, IRRs) and efficacy (PFS) will be performed using summary statistics for qualitative comparison only. This analysis will be conducted both by comparing ever positive versus never positive and by comparing treatment-emergent versus the rest of the subjects.

CCI

[redacted]

17 Safety Evaluation

Analysis sets: Safety analysis set

17.1 Adverse Events

Treatment emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period or if the worsening of an event is during the on-treatment period.

All analyses described in Section 17.1 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). A separate listing including AEs started after the on-treatment period will also be provided.

- **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 (i.e. no answer to the question "Relationship with study treatment").

- **Serious Adverse Events (SAE):** serious adverse events (as recorded on the AE eCRF page, Serious Adverse Event = Yes).
- **Adverse Events Leading to 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).
- **Infusion Related Reaction:** IRRs (as identified according to a pre-specified search list of MedDRA PTs, documented in a version-controlled repository maintained by the Sponsor and finalized for analysis of the current study data prior to DB lock) with onset on study drug dosing date (not prior to infusion of study drug) or the day following study drug infusion.
- **Immune Related Adverse Events (irAE):** irAEs are identified according to a pre-specified search list of MedDRA PTs, documented in a version-controlled repository maintained by the Sponsor and finalized for analysis prior to DB lock.

Unless otherwise stated adverse events will be displayed in terms of frequency tables: PT and primary SOC in alphabetical order. If an adverse event is reported for a given subject more than once during treatment, the worst severity and the worst relationship to trial treatment will be tabulated. Each subject will be counted only once within each PT or SOC and recording period. If a subject experiences more than one AE within a PT or SOC for the same recording period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity.

Adverse events related to trial treatment are those events with relationship missing, unknown or yes.

17.1.1 All Adverse Events

Adverse events will be summarized by worst severity (according to NCI-CTCAE version 4.0) per subject, 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 subject had 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 following summaries:
  - TEAEs
  - TEAEs, grade ≥ 3

- Related TEAEs
- Related TEAEs, grade ≥ 3
- TEAEs leading to permanent treatment discontinuation
- Related TEAEs leading to permanent treatment discontinuation
- Serious TEAEs
- Related serious TEAEs
- TEAEs leading to death
- Related TEAEs leading to death
- Treatment emergent irAEs
- Related treatment emergent irAEs
- Treatment emergent IRRs

- Related treatment emergent 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
- TEAEs by SOC and PT: displaying in separate columns the All TEAEs / Related TEAEs / Grade ≥3 TEAEs / Related Grade ≥3 TEAEs
- TEAEs excluding, SAEs with frequency ≥ 5% by SOC and PT

A listing of AEs including all relevant information will be provided. In addition, a listing of non-TEAEs will also be provided.

17.1.2 Adverse Events Leading to Treatment Discontinuation

The following overall frequency tables will be prepared for the adverse event actions that lead to treatment discontinuation:

- AEs leading to permanent treatment discontinuation by SOC and PT
- Related AEs leading to permanent treatment discontinuation by SOC and PT

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

17.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

17.2.1 Deaths

All deaths, deaths within 30 days after last dose of study drug, death within 60 days after first dose as well as the primary reason for death, will be tabulated based on information from the "Report of Subject Death" and "Survival Follow-Up" CRFs.

- Number of Deaths
- Number of Deaths within 30 days after last dose of avelumab
- Number of Deaths within 60 days after first dose of avelumab
- Primary Reason of Death
  - Disease progression
  - Adverse event related to study treatment
  - Adverse event not related to study treatment
  - Other
  - Unknown

In addition, date and cause of death will be provided in individual subject data listing together with selected dosing information (date of first / last administration, number of infusions received overall) 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 trial treatment
- flag for death within 60 days of first trial treatment

17.2.2 Serious Adverse Events

The following overall frequency tables will be prepared for serious adverse events (SAEs):

- Incidence of serious AEs by SOC and PT
- Incidence of related serious AEs by SOC and PT

The listings of SAEs will also be provided with the relevant information.

17.2.3 Other Significant Adverse Events

The following tables will be created for treatment emergent irAEs:

- The overall summary of treatment emergent irAEs table will include the following summaries:
    - Treatment emergent irAEs
    - Treatment emergent irAEs, grade ≥ 3
    - Related treatment emergent irAEs
    - Related treatment emergent irAEs, grade ≥ 3
    - Treatment emergent irAEs leading to permanent treatment discontinuation
    - Related treatment emergent irAEs leading to permanent treatment discontinuation
    - Serious treatment emergent irAEs
    - Related serious treatment emergent irAEs
    - Treatment emergent irAEs leading to death
    - Related treatment emergent irAEs leading to death
- irAEs by SOC and PT
- Related irAEs by SOC and PT

The listing of irAEs will also be provided with the relevant information. The following information will be provided for each irAE:

- The time from first study treatment to start of AE
- The time from most recent study treatment until start of AE
- The duration of the AE

In addition, a listing of irAEs with onset or worsening date after the on-treatment period will also be provided.

An exploratory analysis will be performed whether the AEs are correlated with the individual AUC as derived from population pharmacokinetics using data combined from this study and other studies, and reported separately.

17.2.4 Infusion Related Reaction

IRRs will be summarized by the follow variables:

- Number of subjects with at least one event by the worst toxicity grade (grade 1/ grade 2/ grade 3/ grade 4/ grade 5/ missing grade)
- Number of subjects with IRR leading to permanent treatment discontinuation
- Time related to first onset (infusion 1/ infusion 2/ infusion 3/ infusion 4 or later). The events should be assigned to the actual drug infusions that the subject received, not to the planned

dates. An IRR is assigned to a drug infusion if its onset is at the same date (but not before dosing) or the following day of drug infusion.

The listing of IRRs will also be provided with the relevant information.

17.3 Clinical Laboratory Evaluation

17.3.1 Hematology and Chemistry Parameters

Laboratory results will be classified according to the NCI-CTCAE criteria version 4.03. Additional laboratory results that are not part of NCI-CTCAE will be presented according to the categories: below normal limits, within normal limits and above normal limits (according to the laboratory normal ranges).

The worst grade during the on-treatment period will be summarized considering only subjects with post baseline laboratory samples: Laboratory tests by NCI-CTC grade (>=0, >=3, >=4).

Quantitative data will be examined for trends using descriptive statistics (mean, StdDev, median, Q1, Q3, minimum, and maximum) of actual values, absolute changes and percent changes from baseline for each visit over time based on visit windows specified in Section 11. Qualitative data based on reference ranges will be described according to the categories (i.e. Low, Normal, High).

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

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

Corrected Calcium (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4].

Alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total bilirubin 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. The possible Hy's Law case is identified as (ALT or AST) ≥ 3*ULN and total bilirubin ≥ 2*ULN during on-treatment period.

Parameters with NCI-CTC grades available:

The laboratory toxicities will be tabulated using descriptive statistics (count and percentage) during the on-treatment period.

- The summary of laboratory parameters by CTCAE grade table will include number and percentage of subjects with any grade, grade 3/4, and grade 4 laboratory abnormalities during the on-treatment period.

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

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

- 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 bilitubin increased, Cholesterol (cholesterol high), Creatinine (creatinine increased), Creatine Kinase (CPK increased), Potassium (hypokalemia/hyperkalemia), Sodium (hyponatremia/ hypernatremia), Magnesium (hypomagnesemia/hypermagnesemia), Calcium (hypocalcemia/ hypercalcemia), Glucose (hypoglycemia/hyperglycemia), Gamma Glutamyl Transferase (GGT) (GGT increased), Lipase (lipase increased), Phosphates (hypophosphatemia), Triglycerides (hypertriglyceridemia).

Parameters with NCI-CTC grades not available:

Hematology and chemistry evaluations which can't be graded per CTCAE criteria will be summarized as:

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

In this study, these apply to the following parameters:

- Hematology:

  Hematocrit, Red Blood Cell (RBC), Reticulocytes, Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin Concentration (MCHC).

- Serum Chemistry:

  Chloride, C-Reactive Protein, Lactate Dehydrogenase (LDH), Total Protein, Total Urea, Uric Acid.

Summary of liver function elevation will include the following categories:

- AST: ≥3×ULN / ≥5×ULN / ≥10×ULN / ≥20×ULN

- ALT: ≥3×ULN / ≥5×ULN / ≥10*ULN / ≥20×ULN
- ALT or AST: ≥3×ULN / ≥5×ULN / ≥10×ULN / ≥20×ULN
- Total bilirubin: ≥2×ULN
- ALT ≥ 3×ULN currently with total bilirubin ≥ 2×ULN
- AST ≥ 3×ULN currently with total bilirubin ≥ 2×ULN
- (ALT or AST) ≥ 3×ULN concurrently with total bilirubin ≥ 2×ULN
- (ALT or AST) ≥ 3×ULN concurrently with total bilirubin ≥ 2×ULN and ALP > 2×ULN
- (ALT or AST) ≥ 3×ULN concurrently with total bilirubin ≥ 2×ULN and (ALP ≤ 2×ULN or missing)

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, i.e a subject with an elevation of AST ≥10×ULN will also appear in the categories ≥5×ULN and ≥3×ULN. Liver function elevation will be summarized using frequency and percentage.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created to graphically display

- 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).

In addition, a listing of all TBILI, ALT, AST and ALP values for patients with a post-baseline TBILI ≥ 2×ULN, ALT≥ 3×ULN or AST ≥ 3×ULN will be provided.

Subjects without post baseline laboratory samples will be excluded from analyses with respect to values after the baseline.

17.3.2 Other Laboratory Parameters

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

- Coagulation: activated partial thromboplastin time (aPTT) and prothrombin time (INR).
- Urinalysis: all urinalysis parameters
- Other parameters: hormone, and immunology parameters
- Pregnancy test

The listings of laboratory results will be provided for all laboratory parameters. The listings will be sorted by parameters and assessment dates or visits for each subject. Laboratory values that

are outside the normal range will also be flagged in the data listings, along with corresponding normal ranges.

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

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

17.4 Vital Signs

The maximum changes of vital sign measurements from screening/baseline will be grouped as follows:

Table 7 Vital Sign Parameters Change from Baseline Category

Parameters	Categories of Change from Baseline
Body temperature increase < 37°C, 37-<38°C, 38-<39°C, 39-<40°C, ≥ 40°C	< 1°C , 1-<2°C , 2-<3°C, ≥ 3 °C
Weight increase	<10%, ≥ 10%
Weight decrease	<10%, ≥ 10%
Heart rate increase <100 bpm ; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
Heart rate decrease <100 bpm ; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
SBP increase <140 mmHg; ≥ 140 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
SBP decrease <140 mmHg; ≥ 140 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP increase <90 mmHg; ≥ 90 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP decrease <90 mmHg; ≥ 90 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
Respiration rate increase <20 bpm ; ≥ 20 bpm	≤5 bpm, >5 – 10 bpm, >10 bpm
Respiration rate decrease <20 bpm ; ≥ 20 bpm	≤5 bpm, >5 – 10 bpm, >10 bpm

All vital sign parameters will be examined for trends using descriptive statistics (mean, StdDev, median, Q1, Q3, minimum, and maximum) of actual values, absolute changes and percent changes from baseline for each visit over time based on visit windows specified in Section 11.

For each patient the worst on-treatment value will be calculated. Missing values will define a separate category.

The following summaries will be prepared for vital sign parameters as grouped above considering only subjects with post baseline values:

- Maximal Shifts (changes in categories)
- Listing of highest change per subject

An additional subject data listing will present all changes from baseline reported in the highest categories.

17.5 Other Safety or Tolerability Evaluations

17.5.1 ECG

The incidence and percentage of subjects with potentially clinically significant abnormalities (PCSA) for 12-lead Electrocardiogram (ECG) parameters will be summarized during the on-treatment period. The PCSA criteria are provided in Table below.

Table 8 Potentially Clinically Significant Abnormalities criteria for ECG

Test	Potentially Clinically Significant Abnormalities (PCSA) Criteria
Heart Rate (HR)	≤ 50 bpm and decrease from baseline ≥ 20 bpm ≥ 120 bpm and increase from baseline ≥ 20 bpm
PR Interval	≥ 220 ms and increase from baseline ≥ 20 ms
QRS	≥ 120 ms
QTcF and QTcB absolute	Interval >450 msec and interval ≤ 480 msec Interval >480 msec and interval ≤ 500 msec Interval >500 msec
QTcF and QTcB change from baseline	Increase from baseline > 30 msec and ≤ 60 msec Increase from baseline > 60 msec

QTc will be corrected based on Fridericia's formula for QTcF and Bazett's formula for QTcB (QTcF= QT/ $\sqrt[3]{RR}$ and QTcB= QT/ $\sqrt[2]{RR}$) where RR=60/HR. Baseline QTcF and QTcB will be derived from the visit that other ECG parameters are flagged as baseline. If there are multiple assessments at the same visit and time point, the average will be calculated for each parameter and used for the analysis.

Quantitative data for HR, PR, QRS, QTcF and QTcB will also be examined for trends using descriptive statistics (n, missing, mean, StdDev, median, Q1, Q3, minimum, maximum, and confidence interval) of actual values with 95%CI, absolute changes from baseline to each visit with 90%CI based on visit windows specified in Section 11.

Listings of 12-lead ECGs will be provided with all relevant information and derived variables.

17.5.2 ECOG Performance Status

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

18 References

1. Jennison C, Turnbull BW. In: Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, 2000.
2. Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982;38:29–41.
3. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons 1980.
4. Beal S. L. Ways to fit a PK model with some data below the quantification limit. J Pharmacokin Pharmacodyn. 28: 481 – 504 (2001).
5. Eisenhauer, E.A, Therasse, P., Bogaerts. J. et. al. (2009): New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer 45, 228-247.

19 Appendices

Appendix I RECIST 1.1

Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228-47.

Appendix II Programmed Clinically Important Protocol Deviations

Inclusion and exclusion criteria are referred according to Protocol version 7.0 dated 26 February 2015

	Category of Protocol Deviation	Description of Protocol Deviation	Deviation Code	Protocol Section	Proposed check / comment
Inclusion criteria: **For the subject to be eligible for inclusion, each criterion must be checked 'YES':**					
Signed written informed consent.	Informed Consent Criteria	Subject did not meet inclusion criterion 1.	PDEV01	Section 5.3.1	Medical Review Required
Male or female subjects aged >=18 years	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 2	PDEV02	Section 5.3.1	Medical Review Required
Histologically proven MCC	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 3	PDEV03	Section 5.3.1	Medical Review Required
Disease must be measurable with at least 1 unidimensional measurable lesion by RECIST 1.1 (including skin lesions)	Eligibility and Ebtry Criteria	Subject did not meet inclusion criterion 7.	PDEV04	Section 5.3.1	Medical Review Required
Effective contraception for both male and female subjects if the risk of conception exists	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 11	PDEV05	Section 5.3.1	Medical Review Required
Exclusion criteria: **For the subject to be eligible for inclusion, each criterion must be checked 'NO':**					
Concurrent treatment with a non-permitted drug.	Eligibility and Entry Criteria	Subject met exclusion criterion 2	PDEV06	Section 5.3.2	Medical Review Required
Prior therapy with any antibody/drug targeting T cell co-regulatory proteins	Eligibility and Entry Criteria	Subject met exclusion criterion 3	PDEV07	Section 5.3.2	Medical Review Required

	Category of Protocol Deviation	Description of Protocol Deviation	Deviation Code	Protocol Section	Proposed check / comment
Concurrent anticancer treatment (for example, cytoreductive therapy, radiotherapy [with the exception of palliative bone-directed radiotherapy, or radiotherapy administered to non-target superficial skin lesions], immune therapy, or cytokine therapy except for erythropoietin)	Eligibility and Entry Criteria	Subject met exclusion criterion 4.	PDEV08	Section 5.3.2	Medical review required
Previous malignant disease (other than MCC) within the last 5 years with the exception of basal or squamous cell carcinoma of the skin or cervical carcinoma in situ	Eligibility and Entry Criteria	Subject met exclusion criterion 8	PDEV09	Section 5.3.2	Medical review required
Pregnancy or lactation	Eligibility and Entry Criteria	Subject met exclusion criterion 14.	PDEV10	Section 5.3.2	Medical review required
Prohibited Medication Criteria					
Non-permitted concomitant medication during the study	Prohibited Medications	Subjects took Immunotherapy, immunosuppressive drugs (that is, chemotherapy or systemic corticosteroids except for short-term treatment of allergic reactions or for the treatment of irAEs), or other experimental pharmaceutical products	PDEV11	Section 6.5.2	Medical review required
Other criteria					

	Category of Protocol Deviation	Description of Protocol Deviation	Deviation Code	Protocol Section	Proposed check / comment
Subjects that developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subject became pregnant, but continued on the study.	PDEV12	Section 5.5.2	
Subjects who developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subjects received treatment despite confirmed progression and significant clinical deterioration. Subjects had ECOG>=3, did not resolved to <=2 by day 14 of next cycle, and continued on the study.	PDEV13	Section 5.5.2 and 5.5.3	Medical review required
Subjects who developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subject developed grade 4 ADR, but continued on the study.	PDEV14	Section 5.5.2 and 5.5.3	Medical review required
Subjects who developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subject developed grade 3 ADR, but continued on the study.	PDEV15	Section 5.1.4.2	Medical review required
Subjects dosing error (≥ +/- 10 % assigned dose)	Other	Dosing error	PDEV16	Section 6.2	List if relative dose intensity ≥1.10 or ≤ 0.90
Subjects on avelumab treatment did not receive premedication	IP compliance	Subjects on avelumab treatment did not take mandatory premedication	PDEV17	Section 6.2.	Premedication data collected on eCRF "Premedication details" page
Any other protocol deviation which is deemed to be significant but has not been pre-specified in this table	any	any	PDEV99	NA	Medical Review required

Statistical Analysis Plan for Analysis of Efficacy and Safety

Clinical Trial Protocol Identification No.	EMR100070-003 Part B
Title:	A Phase II, open-label, multicenter trial to investigate the clinical activity and safety of avelumab (MSB0010718C) in subjects with Merkel cell carcinoma
Trial Phase	Phase II
Clinical Trial Protocol Version	17 May 2017 / Version 11.0
Statistical Analysis Plan Author	PPD
Statistical Analysis Plan Date and Version	16 November 2017 / Version 3.0
Statistical Analysis Plan Reviewers	PPD

[redacted]

[redacted]

1 Signature Page

Statistical Analysis Plan: EMR100070-003 Part B

A Phase II, open-label, multicenter trial to investigate the clinical activity and safety of avelumab in subjects with Merkel cell carcinoma

PPD

PPD

Signature

Trial Biostatistician

PPD

PPD

Date of Signature

PPD

Signature

Project Statistician
Merck KGaA
Franfurter Strasse 250
64293 Darmstadt

PPD

Date of Signature

PPD

PPD

Signature

Medical Responsible
EMD Serono, Inc.
45A Middlesex Turnpike
Billerica, MA 01821

PPD

Date of Signature

[redacted]

2 Table of Contents

3 List of Abbreviations and Definition of Terms

ADA	Antidrug antibody
ADR	Adverse drug reaction
AE(s)	Adverse event(s)
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
ANC	Absolute neutrophil count
aPTT	Activated partial thromboplastin time
AST	Aspartate aminotransferase
BOR	Best overall response
CI	Confidence interval(s)
CK20	Cytokeratin 20
C_{min}	Minimum postdose (trough) concentration
CR	Complete response
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DR	Duration of response
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic case report form
EEA	European Economic Area
CCI [redacted]	[redacted]
FACT-M	Functional Assessment of Cancer Therapy – Melanoma
FAS	Full Analysis Set
GCP	Good Clinical Practice
GGT	Gamma-glutamyl transferase

ICF	Informed Consent Form
ICH	International Conference on Harmonisation
IERC	Independent Endpoint Review Committee
INR	International normalized ratio
irAE	Immune-related adverse event
CCI	[redacted]
IRR	Infusion-related reaction
CCI	[redacted]
IV	Intravenous
LDH	Lactate dehydrogenase
MCC	Merkel cell carcinoma
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
CCI	[redacted]
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PD-1	Programmed death 1

CCI	[redacted]
PFS	Progression-free survival
PK	Pharmacokinetic(s)
PR	Partial response
RBC	Red blood cell
RECIST 1.1	Response Evaluation Criteria in Solid Tumors version 1.1
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SLD	Sum of target lesion diameters
SMC	Safety Monitoring Committee
TEAE	Treatment-emergent adverse event
TTP	Time to progression
ULN	Upper limit of normal
US	United States
WBC	White blood cell

4 Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	18 March 2016	PPD	NA. The first version
2.0	20 October 2016	PPD	1. Updated in accordance with Protocol v10 (20-Oct-2016) 2. Modified subgroup analysis sets in Section 10 to match Part A SAP 3. Updated irAE definition in accordance with Part A; added Appendix II 4. Minor editorial updates
3.0	16 November 2017	PPD	1. Updated in accordance with Protocol v11 (17-May-2017) 2. Changed CRO name from 'PPD' to 'PPD' 3. Updated language to use 'study' instead of 'trial' to be consistent with current Harmonized SAP (except in text copied directly from Protocol) 4. Updated Per Protocol Analysis Set definition to match current list of Clinically Important PDs 5. Removed references to replicate ECG measurements 6. Australia and/or Latin America will now be included as additional pooled geographical regions if there are >= 10 subjects in FAS (instead of > 10% of FAS) 7. Updated subgroup categories of time from initial diagnosis to study entry to: <= 3 months, > 3 months and <= 1 year, > 1 year 8. Added subgroup for time from occurrence of metastatic disease to study entry 9. Updated definition of on-treatment period to end at start of new anti-cancer *drug* therapy (instead of any anti-cancer therapy), when applicable 10. Added specifications for handling summary statistics over time 11. Removed ECG assessments from visit windowing 12. Re-assigned AVISITN values in visit windowing 13. Changed 'randomized treatment' to 'study treatment' as this is a non-randomized study 14. Removed from AE data imputation rules: 'If stop date of AE is after date of cut-off outcome of AE is ongoing at cut-off.' 15. Added date imputation rule for subsequent anti-cancer therapy 16. Updated Geographic Region values and added EEA in Demographics section

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
			17. Removed reference to listing of prior therapies for metastatic disease (n/a in Part B); consolidated planned summaries of prior systemic anti-cancer drug therapies 18. Updated terminology of 'Dose Reduction' to 'Partial Doses' 19. Clarified handling of tumor assessments on the same day as start of new anti-cancer therapy in BOR & PFS derivations 20. Updated 'in-evaluable' to 'non-evaluable' to match RECIST terminology 21. Added lists of OS and PFS censoring reasons 22. Added details of concordance rate calculation between IERC and investigator assessed BOR 23. Added bullet points to clarify CCI confirmation; updated footnote (b) in Table 5 accordingly 24. Removed scatterplots from HRQoL analyses 25. Added description of handling multiple consecutive occurrences of the same AE 26. Added Immunogenicity section (17.5.3) and Immunogenicity subgroup analysis section (17.2.4); removed section 16.6 (superseded) 27. Added indication for 'Clinically Important' PDs in Appendix I 28. Minor editorial updates

5 Purpose of the Statistical Analysis Plan

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specifications for the interim and the primary analysis of data collected for Part B of the protocol EMR100070-003. Results of the analyses described in this SAP will be included in the Clinical Study Report (CSR). Additionally, the planned analyses identified in this SAP will be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective SAP will be clearly identified in the CSR.

The SAP is based upon section 8 (Statistics) of the study protocol and protocol amendments and is prepared in compliance with ICH E9. The detailed description of all statistical analysis for Part A of the protocol is included in a separate SAP (version 3.0 dated March 31, 2016). This version of the SAP only includes the detailed description of the efficacy and safety analysis for Part B of the protocol.

6 Summary of Clinical Trial Features

Trial Objectives	**Primary objective** The primary objective of Part B of the trial is to evaluate the clinical activity of avelumab as first-line treatment for metastatic or distally recurrent MCC as determined by the durable response rate (DRR) according to RECIST 1.1 by an IERC. **Secondary objectives** Secondary objectives of Part B of the trial are as follows: • To assess the OS time • To assess ORR according to RECIST 1.1 • To assess DR according to RECIST 1.1 • To assess PFS according to RECIST 1.1 • To assess the safety profile of avelumab in subjects with MCC • To assess response status according to RECIST 1.1 at 6 and 12 months after start of study treatment • To characterize the population PK of avelumab in subjects with MCC by sparse sampling • To evaluate the immunogenicity of avelumab and to correlate it to exposure **Exploratory objectives** CCI [redacted]
Primary endpoint(s)	The primary endpoint of Part B of the trial is durable response, defined as objective response (CR or PR) according to RECIST 1.1, and as determined by an IERC, with a duration of at least 6 months.

Secondary/ Exploratory endpoints	Secondary endpoints of Part B of the trial include • confirmed BOR per RECIST 1.1 as determined by an IERC, • DR according to RECIST 1.1 as determined by an IERC, • PFS time according to RECIST 1.1 as determined by an IERC, • occurrence and severity of treatment-related AEs according to NCI-CTCAE v 4.0, • OS time, • response status according to RECIST 1.1 at 6 and 12 months after start of study treatment, • serum titers of anti-avelumab antibodies, and • population PK profile of avelumab (sparse sampling). Exploratory endpoints of Part B of the trial include CCI [redacted]
Trial design and plan	This is a multicenter, international, single-arm, open-label, Phase II trial in two parts that will evaluate the efficacy and safety of avelumab in subjects with metastatic MCC. For Part B of the trial, subjects must be treatment naïve to systemic therapy in the metastatic setting. For Part B of the trial, up to 112 eligible subjects are planned to be enrolled and will receive avelumab at a dose of 10 mg/kg as a 1-hour intravenous (IV) infusion (-10 minutes / +20 minutes, that is, over 50 to 80 minutes) once every 2 weeks. Tumor measurements (including the assessment of skin lesions by physical examination) to determine response will be performed every 6 weeks until 12 months after the first study drug administration, then every 12 weeks, and response to the treatment will be evaluated by RECIST 1.1. Treatment will continue until • therapeutic failure (subjects may stay on treatment beyond observation of progressive disease [PD] provided there is no significant clinical deterioration); • unacceptable toxicity; or • any criterion for withdrawal from the trial or the trial drug is fulfilled.

	Significant clinical deterioration is defined as • new symptoms or worsening of symptoms that cannot be managed by optimal supportive care or disease localization that require immediate medical or surgical intervention (for example, lesion close to the spine), and/or • change in Eastern Cooperative Oncology Group Performance Status (ECOG PS) to ≥ 3 that lasts more than 14 days. Decisions regarding medical management of subjects will be made by the Investigator; however, the primary and secondary endpoint determinations will be according to tumor assessments performed by the IERC. The date of the first observation of PD by RECIST 1.1 by the IERC will be used to determine the date of the PD as well as the DR in all subjects, including the subjects for which treatment was maintained beyond first determination of disease progression. Adverse events will be assessed throughout the trial period and evaluated using the National Cancer Institute (NCI) Common Technology Criteria version 4.0 (CTCAE v 4.0). CCI [redacted] For Part B of the trial there will be 1 interim analysis at 3 months after the accrual of the 25th subject, with additional interim analyses possible. The primary analysis will be conducted 15 months after the accrual of the last subject. Subject follow-up for progression and survival will continue until 5 years after the last subject receives the last dose of avelumab or the last subject dies, whichever occurs first. Under some circumstances, subjects may not be followed for 5 years for survival in this study, for example, subjects may be offered to enroll into a rollover study, or the Sponsor may terminate the study early.
Planned number of subjects	One hundred and twelve subjects are planned to be enrolled in Part B of this study.
Investigational Medicinal Product (s): dose/mode of administration/ dosing schedule	Avelumab will be administered as a 1-hour IV infusion (-10 minutes / +20 minutes, that is, over 50 to 80 minutes) at 10 mg/kg once every 2-week treatment cycle. In order to mitigate infusion-related reactions, subjects will receive pretreatment with antihistamines and acetaminophen prior to the first four avelumab infusions. Premedication should be administered for subsequent avelumab doses based upon clinical judgment and presence/severity of prior infusion reactions. Steroids as premedication are not acceptable. This regimen may be modified based on local treatment standards and guidelines as appropriate. Following avelumab infusions, subjects must

	be observed for 2 hours post-infusion for potential infusion-related reactions for the first four infusions. Starting with the fifth infusion, observation should be based upon clinical judgment and presence/severity of prior infusion reactions. The dose of avelumab will be calculated based on the weight of the subject determined within 3 days of each drug administration. Immediate access to an intensive care unit or equivalent environment and appropriate medical therapy (including IV epinephrine, corticosteroids, antihistamines, bronchodilators, and oxygen) must be in place for use in the treatment of potential infusion-related reactions. Infusion of avelumab will be stopped in case of Grade ≥ 2 infusion-related, allergic, or hypersensitivity reactions (according to NCI-CTCAE v 4.0). Following avelumab infusions, subjects must be observed for 2 hours post infusion for potential infusion-related reactions. In the case of Grade 1 or 2 infusion reactions, the infusion rate should be decreased by 50%. If the subject has a second infusion-related reaction Grade ≥ 2 on the slower infusion rate, the infusion should be stopped and the subject should be removed from avelumab treatment.
Reference therapy(ies): dose/mode of administration/ dosing schedule	Not applicable.
Planned treatment duration per subject	Subjects will receive avelumab treatment until significant clinical deterioration, unacceptable toxicity, or any criterion for withdrawal from the trial or trial drug is fulfilled. Part B of the trial: Subjects who have experienced a confirmed CR should be treated for a minimum of 12 months after confirmation, at the discretion of the Investigator. If the Investigator believes that a subject may benefit from treatment beyond 12 months, it may be permissible after discussion with the sponsor. In case a subject with a confirmed CR relapses after stopping treatment, but prior to the end of the trial, 1 re-initiation of treatment is allowed at the discretion of the Investigator and agreement of the trial Medical Monitor. In order to be eligible for retreatment, the subject must not have experienced any toxicity that led to treatment discontinuation of the initial avelumab therapy. Subjects who re-initiate treatment will stay on study and will be treated and monitored according to the protocol and the "until progression" schedule in the Schedule of Assessments.
Randomization and Blinding	Not applicable.

7 Sample Size/Randomization

The primary endpoint for Part B of this study is durable response, defined as objective response (CR or PR) according to RECIST 1.1 and as determined by an IERC, with a duration of at least 6 months. The planned total sample size for Part B is 112 subjects for addressing the primary objective, relevant subgroup analyses, consistency, and further safety assessments. Assuming a true DRR of 45%, the probability to observe lower bound of the exact 95% confidence interval (CI) above 20% would be > 99% and above 30% would be 90%.

A historical reference for the DRR under first-line chemotherapy is based on P. Nghiem (personal communication), and Iyer 2016, which showed 11 of 62 patients meeting the definition of 6 months durable response, i.e. a DRR of 17.7% (95% CI: 9.2, 29.5). Part B of the study aims to estimate the DRR with a sufficient level of precision. For example, with 112 subjects, an observed DRR of 28.6% would lead to an exact (Clopper-Pearson) 95% CI of (20.4%; 37.9%), which would exclude the observed DRR in Iyer et al, an observed DRR of 40.2% would lead to a 95% CI of (31.0%; 49.9%), which would exclude the upper bound of the 95% CI of Iyer et al, and an observed DRR of 44.6% would lead to an exact 95% CI of (35.2%; 54.3%).

The two-sided 95% Clopper-Pearson confidence intervals (CIs) for different observed values for the DRR are provided in Table 1.

Table 1 Confidence Intervals for Different Durable Response Rates

N	DRR (%)	n	LBCI (%)	UBCI (%)
112	28.6	32	20.4	37.9
112	40.2	45	31.0	49.9
112	44.6	50	35.2	54.3

N, n = number of subjects; DRR = durable response rate; LBCI = lower boundary of the confidence interval; UBCI = upper boundary of the confidence interval.

8 Overview of Planned Analyses

This SAP covers the analyses for efficacy and safety based on the data cut-off dates for the interim and the primary analyses for Part B of the study. Statistical analyses will be performed using cleaned eCRF data as well as external data, including tumor assessment results by the Independent Endpoint Review Committee (IERC), quality of life (QoL) data, CCI [redacted] data.

There is one exploratory interim analysis and one primary analysis planned for Part B of this study. Additional interim analyses may be added as needed.

All data will be included up to a clinical cut-off date corresponding to prespecified minimum duration of follow-up. For the interim analysis, the clinical cut-off date will be the date 3 months after the 25th subject receives the first dose of study treatment; for the primary analysis, the clinical cut-off date will be 15 months after the last subject enrolled receives the first dose of study treatment.

8.1 Sequence of Analysis

The following analyses will be performed during Part B of this study:

- Interim analysis: will be performed 3 months after start of study treatment of the 25th subject.
- Primary analysis: will be performed 15 months after the start of study treatment of the last subject. All subjects enrolled in the study who received at least 1 dose of study treatment will be included in the primary analysis.

The date of the first dose of study treatment of the 25th subject enrolled will be used to determine the data cut-off date for the interim analysis; the date of the first dose of study treatment of the last subject enrolled will be used to determine the data cut-off date for the primary analysis.

8.2 Interim Analysis

As there will not be sufficent data to perform meaningful analysis for the primary endpoint durable response and secondary endpoint OS time at the time of the specified interim analysis, these two efficacy endpoint analyses will not be performed. All other efficacy and safety endpoint analyses will be performed as described in this SAP for the specified interim analysis. Additional interim analyses may be added. If sufficient data are available at an additional interim analysis, then analysis of the primary endpoint, durable response, as well as on OS, may be performed.

8.3 Final Analysis

All planned analyses outlined in this SAP will be performed for the primary analysis. A partial database lock will be performed for the interim and primary analyses. The full DB lock will take place 5 years after the last subject receives the last dose of avelumab, or when the last subject dies, whichever occurs first.

In addition, no database can be locked until this SAP has been approved.

Subjects follow-up for progression and survival will continue until 5 years after the last subject receives the last dose of avelumab, or when the last subject dies, whichever occurs first. Under some circumstances, subjects may not be followed for 5 years for survival in this study, for example, subjects may be offered to enroll into a rollover study, or the Sponsor may terminate the study early.

9 Changes to the Planned Analyses in the Clinical Trial Protocol

The statistical methods as described in the protocol for Part B of this study were adopted.

10 Analysis Sets

Screening Analysis Set

The screening analysis set includes all subjects who signed the informed consent.

Full Analysis Set / Safety Analysis Set

Full analysis set (FAS) / Safety analysis set will include all subjects who received at least one dose of study drug.

Per Protocol Analysis Set

The Per-Protocol (PP) analysis set is a subset of the FAS and includes all FAS subjects who meet all of the following criteria that could impact the key objectives of the study:

- Measurable disease per RECIST 1.1 and IERC assessment
- Distant metastatic disease
- No prior anti-cancer therapy for metastatic disease
- No prior therapy with any antibody/drug targeting T-cell coregulatory proteins such as anti-PD-1, CCI , or anticytotoxic T-lymphocyte antigen-4 (CTLA-4) antibody
- Histologically proven MCC with confirmation of the diagnosis by immuno-histochemistry detection of CK20 (or other appropriate cytokeratin expression such as pancytokeratin, AE1/AE3, or Cam5.2) according to the assessment documented in the "Disease History" eCRF page.
- No active central nervous system (CNS) metastases; subjects with a history of treated CNS metastases (by surgery or radiation therapy) must have fully recovered from treatment, demonstrated no progression for at least 2 months, and not require continued steroid therapy.
- Subject has a tumor assessment > 5 weeks after first dose of study treatment (unless PD or death is observed before that time, in which case the subject will not be excluded from the PP analysis set).

The PP analysis set will be used for additional sensitivity analyses for the primary and secondary efficacy endpoints. If the PP analysis set includes at least 95% of subjects in the safety analysis set, additional efficacy analyses on the PP analysis set will be omitted as the differences in the results based upon these two analysis sets are expected to be negligible.

Subgroup Analysis Sets

The below subgroup analyses will be performed on primary and key secondary efficacy endpoints. Since the study is not powered for any subgroup analysis, CCI .

The following subgroups will be defined:

- Age Group 1:
  - Age < 65 years (Reference)
  - Age ≥ 65 years
- Age Group 2:
  - Age ≤ Median (Reference)
  - Age > Median
- Gender
  - Male (Reference)
  - Female
- Race
  - Caucasian / White (Reference)
  - Asian
  - Black/ African American
  - Other
- Pooled Geographical Region:
  - North America
  - Europe (Reference)
  - Asia
  - Rest of the World (Australia and/or Latin America will be included as additional pooled geographical regions if including >= 10 subjects in the FAS)

CCI

[redacted]

CCI

- Time from initial diagnosis to study entry
  - <= 3 months (Reference)
  - > 3 months and <= 1 year
  - > 1 year
- Time from first occurrence of metastatic disease to study entry
  - <= 3 months (Reference)
  - > 3 months and <= 1 year
  - > 1 year
- ECOG PS at baseline
  - ECOG PS 0 (Reference)
  - ECOG PS 1
- Disease Burden by IERC assessment
  - Baseline SLD ≤ Median (Reference)
  - Baseline SLD > Median
- Presence of visceral metastases at baseline. Target or non-target lesions that are categorized as neither skin (including soft tissue and eye) nor lymph node lesions per IERC assessment are classified as visceral metastases.
  - Present
  - Absent (Reference)
- Number of prior systemic chemotherapies
  - 0 Prior Therapy (Reference)
  - ≥ 1 Prior Therapies

- Site of primary tumor (representing known versus unknown primary status)
  - Skin
  - Non-Skin (Reference)
- Lymph node disease only at Baseline per IERC assessment
  - Yes (Reference)
  - No
- Baseline CD8 T-cell density (expressed in % CD8+ cells and number of CD8+ cells/mm^2)
  - < Median (Reference)
  - ≥ Median

11 General Specifications for Statistical Analyses

Data handling after 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.

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 subjects treated at each center.

Presentation of continuous and qualitative variables:

Continuous variables will be summarized using descriptive statistics i.e., number of non-missing values and number of missing values, [i.e., 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 stated 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 subjects still present in the study at that visit, unless otherwise specified.

Definition of baseline:

The last available assessment prior to the start of study treatment is defined as baseline measurement for both 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.

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

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

Baseline for HR and QTc assessments will be derived from the visit where both HR and QT are not missing.

Definition of study day/treatment day:

Treatment day is defined relative to the date of first study drug administration.

Treatment day 1 is defined as date of first study drug administration, the day before is defined as treatment day - 1 (no treatment day 0 is defined).

Definition of 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 date of new anti-cancer drug therapy after start date is derived as outlined in Section 14.2.

Standard derivations and reporting conventions:

The following conversion factors will be used to convert days into weeks, months or years:

1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

Demographics and physical measurements:

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

  The integer part of the calculated age will be used for reporting purposes.
- BMI (kg/m^2) = weight (kg)/[height (m)]^2.
- BSA (m^2) = ([height (cm) × weight (kg)] / 3600)$^{0.5}$

For reporting conventions, mean and median should generally be displayed to 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. E.g. 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.

Unscheduled visits:

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 and 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.

Summary statistics over time:

For descriptive statistics over time by nominal visit or time point for QoL, laboratory assessment, and vital sign data, only those planned visits/time points that have at least 5 subjects will be included in the summary tables and figures. The exception is the Discontinuation and End-of-Treatment visits, which will be included in the summary statistics regardless of the number of subjects who completed the visit.

Data collected after re-initiated treatment:

Data collected after re-initiation of treatment will only be included in the data listing. It will not be included for safety and efficacy summary analyses except for overall survival and disposition.

Visit window:

The assignment of visit window is described in Table 2 for the purpose of by-visit analyses of laboratory assessment and vital sign data:

- Baseline will be derived as described above.
- Both scheduled and unscheduled assessments are included for visit windowing.
- No visit windowing will be performed at Discontinuation, End-of-Treatment, or Safety Follow-up visits for laboratory and vital sign data. Instead, the earliest non-missing observation among the unscheduled or scheduled assessments for each visit (Discontinuation, End-of-Treatment, and Safety Follow-up) will be used for the analysis.
- If there are multiple assessments for any specified visit and some of them are from scheduled visits with non-missing assessment results, the assessment from scheduled visit that is closest to the planned study day will be used for analysis.
- If there are multiple assessments for any specified visit and none of them are from scheduled visits, the assessment with non-missing results and closest to the planned study day will be used for analysis.

- If there are two or more unscheduled assessments with non-missing results and the same distance to the planned study day, the assessment prior to the planned study day will be used in deriving visit window. For example, if the lab assessment was done on both study day -1 and 1, then the assessment on -1 study day will be used for visit windowing.

Table 2 Visit Window Definition for Safety Assessment

Assigned Study Day (Inclusive)		Planned Study Day (AWTARGET)	Analysis Visit (N) (AVISITN)	Analysis Visit (AVISIT)	Assessment
From (AWLO)	To (AWHI)				
~	1	1	1	Baseline	Lab, Vital Sign
5	11	8	2	Week 2 Day 5-11	
12	18	15	3	Week 3 Day 12-18	Lab, Vital Sign
19	25	22	4	Week 4 Day 19-25	Core Serum Chemistry
26	32	29	5	Week 5 Day 26-32	Lab, Vital Sign
33	39	36	6	Week 6 Day 33-39	Core Serum Chemistry
40	50	43	7	Week 7 Day 40-50	Lab, Vital Sign
51	64	57	9	Week 9 Day 51-64	Lab, Vital Sign
65	78	71	11	Week 11 Day 65-78	Lab, Vital Sign
79	92	85	13	Week 13 Day 79-92	Lab, Vital Sign
93	106	99	15	Week 15 Day 93-106	Lab, Vital Sign
107	120	113	17	Week 17 Day 107-120	Lab, Vital Sign
121	134	127	19	Week 19 Day 121-134	Lab, Vital Sign
135	148	141	21	Week 21 Day 135-148	Lab, Vital Sign
149	162	155	23	Week 23 Day 149-162	Lab, Vital Sign
163	176	169	25	Week 25 Day 163-176	Lab, Vital Sign
177	190	183	27	Week 27 Day 177-190	Lab, Vital Sign
191	204	197	29	Week 29 Day 191-204	Lab, Vital Sign
205	218	211	31	Week 31 Day 205-218	Lab, Vital Sign

Further windows can be calculated based on the following general algorithm starting from the planned study day in Week X:

Planned study day in Week X = (X – 1) * 7 + 1

Assigned study days (inclusive) = (X – 6) days to (X + 7) days

Missing data and imputation rules:

Unless otherwise specified in this SAP, all data will be evaluated as reported and no imputation of missing values will be done.

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

Missing statistics, e.g. 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'.

Partial dates will be imputed as follows:

Disease history

Incomplete dates for disease history (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.

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 subject'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.

Prior/concomitant medication

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 randomized treatment.
- If the day of medication date is missing, but the month and year are equal to the start of randomized treatment, then the medication date will be replaced by the start of randomized treatment. For example, if the medication start date is --/JAN/2015, and randomized 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 randomized treatment, then the medication date will be replaced by the start of randomized treatment. For example, if the medication start date is --/---/2014, and randomized 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 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 subject'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.

Subsequent anti-cancer therapy

Incomplete subsequent anti-cancer therapy start dates will be imputed as follows:

- If only day is missing, it will be imputed as the last day of the month unless the end date of subsequent anti-cancer therapy is before that date. In that case, the incomplete anti-cancer therapy start date will be imputed as the end date of the anti-cancer therapy.
- If both day and month are missing, no imputation will be performed.

Exposure

- In case the study drug start date is missing, no imputation will be performed.
- In case the last date of study drug is incomplete the date of last administration of study drug will be taken from the treatment termination eCRF pages.

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 subject should be considered to be ongoing and use the last dosing date on or prior to 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:

  = 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

Date of last contact

The last contact date will be the latest complete date among the following, up to the cut-off date for the analysis:

- All subject 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
- Date last known to be alive collected on the eCRF form "Subject Status / Survival Follow-up" (only used if status is 'alive')
- Study drug start and end dates
- Date of discontinuation from the "Study Termination" eCRF page (do not use if reason for discontinuation is lost to follow-up)

Death date

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

- If the date is completely missing it will be imputed as the day after the date of last contact.
- If the day or month is missing, death will be imputed to the maximum of the full (non-imputed) date of last contact (excluding the date of death) 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

If the day is missing from the date of last contact it will be imputed to 1st day of the month and year of last contact only if derived from the survival page.

Tumor assessments

The following specifications apply to the determination of tumor assessment date for investigator assessment. For tumor assessments by the IERC, the date of assessment is determined by the IERC as specified in the Charter.

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

If there are multiple scan dates associated with an evaluation, i.e., 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 selected as the earliest of all investigation dates (e.g., X-ray, CT-scan, photography).

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 assessment are not available, this assessment will not be used for any calculations.

For IERC-based tumor assessments, the process is a double radiology review with adjudication if needed, and a single oncologist review. The records containing the adjudicated data from the radiologists are used for analysis. In cases where adjudication was not performed (i.e., the two radiologists were in agreement on the timepoint overall response), the record from the radiologist designated as 'Reader 1' was used for analysis. This decision applies to the algorithms used to identify subjects with visceral metastases at baseline and lymph node disease at baseline, in addition to all efficacy analyses.

All statistical analyses will be performed using SAS® Version 9.2 or higher.

12 Trial Subjects

The subsections in this section include specifications for reporting subject disposition and treatment/study discontinuations. Procedures for reporting protocol deviations are also provided.

12.1 Disposition of Subjects and Discontinuations

The following will be summarized overall.

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

- Total number of subjects screened overall
- Number of subjects who discontinued from the study prior to treatment overall and grouped by the main reason (e.g. the failed specific inclusion or exclusion criteria, withdrawal of consent)
- Number and percentage of treated subjects in the following populations:
  - FAS (Safety analysis set)
  - PP analysis set
- Number and percentage of subjects still on treatment
- Number and percentage of subjects off-treatment overall and by the main reason for discontinuation
- Number and percentage of subjects who discontinued the treatment but are still in follow-up
- Number and percentage of subjects who discontinued the study overall and by the main reason for discontinuation
- Number and percentage of subjects who re-initiated avelumab treatment and discontinued treatment after re-initiation

In addition, the following will be summarized:

- Number and percentage of treated subjects overall, by region (Europe, EEA (required by EudraCT), North America, Latin America, Asia, Australia), by country within region
- Number and percentage of treated subjects by center.
- Number and percentage of subjects by region region (Europe, EEA (required by EudraCT), North America, Latin America, Asia, Australia), by country within region and by center for each of the analysis set (Screening, FAS (Safety) and PP analysis set).

12.2 Protocol Deviations

Protocol deviations will be listed and summarized based on the FAS.

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

- Subjects that are dosed on the study despite not satisfying all inclusion criteria, or satisfying any exclusion criteria;
- Subjects that develop withdrawal criteria while on the study but are not withdrawn;
- Subjects that receive an incorrect dose (with the exception of those who receive a lower dose than planned due to dose interruption resulting from an AE at the time of infusion);
- Subjects that receive an excluded concomitant medication;

- Deviation from GCP.

Important protocol deviations will be determined for all subjects by either medical review processes or programming based on the inclusion/exclusion criteria or other criteria presented in the protocol.

All important protocol deviations should be documented in CDISC datasets whether identified through sites monitoring, medical review or programming. Important protocol deviations to be identified by programming, as well as all clinically important protocol deviations, are listed and described in Appendix I.

13 Demographics and Other Baseline Characteristics

Analysis sets: FAS / Safety Analysis Set

13.1 Demographics

Demographic characteristics will be summarized as follows:

- Demographic characteristics
  - Gender: Male, Female
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Unknown
  - Ethnic origin: Hispanic/Latino (Yes/No), Japanese (Yes/No)
  - Age (years): summary statistics
  - Age categories:
    - < 65 years, ≥ 65 years
    - < 65, 65 – < 75, 75 – < 85, ≥ 85 years
  - Pooled Geographical Region:
    - North America
    - Europe
    - Asia
    - Rest of the World (Australasia, Middle East, Africa, and/or Latin America will be included as additional pooled geographical regions if including >= 10 subjects in the FAS)
  - Geographic Region:
    - North America
    - Latin America
    - Western Europe
    - Eastern Europe/Middle East

  - Africa
  - Australasia
  - Asia
- EEA: Yes or No
- Eastern Cooperative Oncology Group (ECOG) Performance Status

- Physical measurements
  - Height (cm)
  - Weight (kg)
  - Body Mass Index (BMI) (kg/m^2)

Site codes will be used for the determination of the subject's geographic region.

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

13.2 Medical History

Medical history will be coded using the most current available version of Medical Dictionary for Regulatory Activities (MedDRA). Medical history will be summarized as the numbers and percentages of subjects by MedDRA preferred term (PT) as event category and MedDRA primary system organ class (SOC) as summary category. Each subject 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.

13.3 Other Baseline Characteristics

13.3.1 Disease Characteristics

Information on disease characteristics collected on "Disease History" eCRF page will be summarized as follows:

- Site of primary tumor
- Tumor size
- Time from initial diagnosis to study entry
- CCI [redacted]
- Peripheral margin status
- Deep margin status

- Lymphovascular invasion
- Extracutaneous extension
- Time since first metastatic disease (months), defined as (date of the first study treatment – date of the first occurrence of metastatic disease)/30.4375
- Time since first locally advanced disease (months), defined as (date of the first study treatment – the date of the first occurrence of locally advanced disease)/30.4375
- Time since last progression of disease prior to study entry (months), defined as (date of the first study treatment –date of the last progression of disease)/30.4375
- TNM classification at initial diagnosis
- TNM classification at study entry
- Additional clinically relevant factors: Depth (Breslow), Mitotic index, Tumor-infiltrating lymphocytes, tumor growth pattern, Presence of second malignancy

Baseline characteristics with respect to ECOG, vital signs, physical examinations, ECG and hematology/biochemistry will be part of Section 17 (Safety Evaluation).

13.3.2 Prior Anti-Cancer Therapies

The prior anti-cancer therapies are collected under the "Prior Anti-Cancer Drug Therapies Details", "Prior Anti-Cancer Radiotherapy Details" and "Prior Anti-Cancer Surgeries Details" eCRF pages.

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

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

Prior systemic anti-cancer drug therapy will be summarized as follows:

- Number of prior systemic anti-cancer therapy: missing / 0 / 1 / 2 / >2
- Duration of prior chemotherapy
- Type of prior anti-cancer therapy: Chemotherapy / Hormonal therapy / Antibody therapy / Kinase inhibitors / Vaccines / Other
- Number of prior systemic anti-cancer therapies by disease setting: Neo-Adjuvant / Adjuvant / Locally advanced / Metastatic
- Best response: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive Disease (PD) / Not assessable / Unknown / Not applicable. Best response is derived from the last treatment regimen

The prior anti-cancer drugs will also be extensively detailed with the number and percentage of subjects by the drug class and PT in a table. A subject 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.

The following listings of prior anti-cancer treatments and procedures will also be provided: a) listing of prior anti-cancer drug therapies for non-metastatic disease, b) listing of prior anti-cancer radiotherapy and c) listing of prior anti-cancer surgeries. These will include the subject identification number, and all the relevant collected data-fields on the corresponding eCRF pages.

14 Previous or Concomitant Medications/Procedures

Analysis sets: FAS / Safety analysis set

14.1 Prior and Concomitant Medications/Procedures

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

Prior and concomitant medications will be summarized from the "Concomitant Medications Details" eCRF page. Pre-medications for study drug will also be summarized separately.

In cases where date values do not allow unequivocal allocation of a medication to concomitant (as opposed to previous) medication, the medication will be considered as concomitant medication.

Summary of prior and concomitant medications will include the number and percentage of subjects by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term. A subject will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. If any prior or concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes.

The summary tables will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. In case any specific medication does not have ATC classification level 2 coded term, it will be summarized under "Unavailable ATC classification" category.

A listing of concomitant medications will be created with the relevant information collected on the "Concomitant Medications Details" eCRF page.

All concurrent procedures collected on the eCRF page "Concomitant Procedures Details" will be presented in the listing only.

In addition, a listing of premedication will be created with the relevant information collected on the "Premedication Details" eCRF page.

14.2 Subsequent Anti-Cancer Therapies/Procedures

Subsequent anti-cancer treatment will be provided in a data listing with data retrieved from "Anti-Cancer Treatment after Discontinuation Details", "Radiotherapy after Discontinuation", and "Surgery after Discontinuation" eCRF pages. The earliest date of start of new anti-cancer *drug* therapy will be used for the definition of the on-treatment period; the earliest date of start of any new anti-cancer therapy will be used for censoring for efficacy analyses.

15 Treatment Compliance and Exposure

Analysis sets: FAS / Safety analysis set

All dosing calculations and summaries will be based on "Avelumab Administration Details" eCRF pages.

Subjects will receive an IV infusion of avelumab at a dose of 10 mg/kg (over the duration of 1 hour [-10 minutes / +20 minutes, that is, over 50 to 80 minutes]) once every 2 weeks (one cycle). Analysis of exposure will be based on the calculated actual dose levels (total dose administered/weight, mg/kg). The last available weight of the subject within 3 days prior to the day of dosing will be used.

In case of missing total dose, the dose level as entered in the eCRF will be used.

The duration of therapy (in weeks) during the study is defined as:

$$\text{duration} = \left(\frac{\text{date of last dose} - \text{date of first dose} + 14}{7}\right)$$

The cumulative dose (mg/kg) per subject in a time period is the sum of the actual dose levels that the subject received within that period (i.e. total dose administered (mg) / weight (kg)).

Each cycle is defined by a 2-week period. The dose intensity and the relative dose intensity will be calculated for each subject across all cycles. The dose intensity (mg/kg/cycle) per cycle is defined as

$$\text{dose intensity} = \left(\frac{\text{Cumulative dose}}{(\text{duration of therapy (in weeks)})/2}\right)$$

The relative dose intensity is defined as the actual dose intensity divided by the planned dose per cycle.

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

- Duration of therapy (weeks)
- Total number of infusions received
- Cumulative dose of therapy (mg/kg)
- Dose intensity (mg/kg/cycle)
- Relative dose intensity of therapy (%)

A listing of treatment exposure and compliance will also be created to summarize the information listed above for each subject.

Partial Doses

According to the protocol, preplanned dose reductions are not allowed for avelumab; however, interruptions in delivering the planned dose that resulted in an actual non-zero dose < 90% of the planned dose are defined as partial doses within an administration for avelumab. The number and percentage of subjects with at least one partial dose as well as a breakdown of partial doses per subject (1 / 2 / 3 / ≥4) will be summarized. In addition, the number and percentage of subjects with at least one infusion rate reduction of 50% or more will be summarized.

Dose Delays

Delays will be derived based on infusion date and will be grouped into the following categories based on the deviation of the actual to the planned treatment administration day (relative to the previous treatment administration date):

- No delay (including 1-2 days delays)
- 3-6 days delay
- 7-13 days delay
- 14-20 days delay
- 21-27 days delay
- ≥ 28 days delay

For example, if one subject receives the study drug on Day 1, then the next study drug administration date will be on Day 15; however, if the subject receives the study drug at Day 16 or 17, this is considered as 'no delay'.

Number of subjects with delayed infusions and maximum length of delay, i.e. the worst case of delay if subjects have multiple dose delays will be summarized.

A listing of study drug administration will be created with the information collected on the "Avelumab Administration Details" eCRF page.

16 Endpoint Evaluation

In the interim analysis, the efficacy endpoints analysis will be conducted for each of the following subsets of FAS:

- Subjects in FAS with ≥ 6 weeks duration of follow-up
- Subjects in FAS with ≥ 3 months duration of follow-up
- Subjects in FAS with ≥ 6 months duration of follow-up

The primary endpoint and secondary endpoint analysis on FAS will be repeated for the PP analysis set as a sensitivity analysis if the PP analysis set includes less than 95% of subjects in the FAS. This sensitivity analysis will only be performed for subjects with ≥ 6 months duration of follow-up.

16.1 Primary Endpoint Analyses

Analysis sets: FAS / PP analysis set

The primary endpoint for Part B of this study is durable response, defined as objective response in terms of CR or PR according to RECIST 1.1, as determined by an IERC, with a duration of at least 6 months. The duration of response (DR) will be calculated as specified in Section 16.2.3.

DRR is defined as the proportion of subjects with an objective response in terms of CR or PR according to RECIST 1.1, as determined by an IERC, with a duration of at least 6 months, among all subjects in the FAS.

For subjects with a confirmed response but neither documented disease progression nor death within 12 weeks after the last tumor assessment as of the cut-off date for the analysis, the DR will be censored at the date of the last tumor assessment. If the censored DR is less than 6 months, the subject will be treated as failure in the analysis of DRR. Otherwise, the subject will be considered as success and will be included in the numerator in the calculation of DRR.

The following requirement is taken into account for confirmation of response: PR or CR needs to be confirmed at a subsequent tumor assessment, preferably 6 weeks after the initial observation of response and according to the normal 6-week assessment schedule (or a subsequent assessment for PR), but no sooner than 5 weeks after the initial documentation of CR or PR.

DRR will be calculated along with the two-sided 95% CI using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option). Assuming a true DRR of 45%, the probability to observe lower bound of the exact 95% confidence interval (CI) above 20% would be > 99% and above 30% would be 90%.

The primary endpoint analysis on FAS will be repeated for the PP analysis set as a sensitivity analysis if the PP analysis set includes less than 95% of subjects in the FAS.

16.2 Secondary Endpoint Analyses

Analysis set: FAS / PP analysis set

The secondary endpoints of Part B of the study include:

- Overall survival (OS) time
- Best overall response (BOR) according to RECIST 1.1 as determined by an IERC
- DR according to RECIST 1.1 as determined by an IERC
- Progression free survival (PFS) time according to RECIST 1.1 as determined by an IERC
- Response status according to RECIST 1.1 and as determined by an IERC at 6 and 12 months after start of study treatment.

16.2.1 Overall Survival

The overall survival (OS) time is defined as the time from first administration of study treatment until the date of death. The OS time is considered a key secondary endpoint.

For subjects who are still alive at the time of data analysis or who are lost to follow up, OS time will be censored at the last recorded date that the subject is known to be alive as specified in Section 11 as of the data cut-off date for the analysis.

OS (months) = (date of death or censoring – date of first dose of study treatment + 1)/30.4375

The analysis of OS time will be performed using a a Kaplan-Meier method with the same approach as for duration of response described in Section 16.2.2. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. In particular, the proportion of overall survival at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs.

Frequency (number and percentage) of subjects with an event (death) and censoring reasons will be presented. Censoring reasons are as follows, listed in a hierarchical manner:

- Lost to follow-up
- Withdrawal of consent
- Ongoing in the study without an event and ≥ 16 weeks since last known alive
- Ongoing in the study without an event

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

16.2.2 Best Overall Response and Objective Response

The confirmed BOR is defined as the best confirmed response obtained among all tumor assessment visits after start of study treatment until documented disease progression, taking requirements for confirmation into account as detailed in Section 16.1.1. In addition, the minimum duration for a BOR of SD is defined as at least 6 weeks after start of study treatment. A BOR of PD requires that progression is observed ≤ 12 weeks after first dose of study treatment (and not qualifying for CR, PR or SD).

Only tumor assessments performed before the start of any further anti-cancer treatment will be considered in the assessment of BOR. Clinical deterioration will not be considered as documented disease progression. If a tumor assessment was performed on the same day as the start of a new anti-cancer therapy, it is assumed that the tumor assessment was performed prior to the start of the new anti-cancer therapy; therefore, this tumor assessment is included in the assessment of BOR.

The tumor response at each assessment visit will be determined according to RECIST 1.1 and as adjudicated by an IERC. In case of different dates of scans within the same tumor assessment visit, the date of assessment is determined by the IERC as specified in the Charter.

The objective response rate (ORR) is defined as the proportion of subjects having reached a confirmed BOR of CR or PR according to RECIST 1.1 and as determined by the IERC.

Table summarizes the derivation rules described by Eisenhauer, et al. for the BOR when confirmation from subsequent assessment is needed (5).

Table 3 Best Overall Response When Confirmation of CR/PR Is Required

Overall response 1st time point	Overall response subsequent time point	Best overall response (BOR)
CR	CR	CR
CR	PR	SD, PD or PR[a]
CR	SD	SD provided minimum criteria for SD duration met; otherwise, PD
CR	PD	SD provided minimum criteria for SD duration met; otherwise, PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise, PD
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, NE = non-evaluable.

[a] If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the subject had PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.

The confirmed ORR will be calculated along with the two-sided 95% CI using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option).

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

- No baseline assessment (IERC identifies neither any target nor any non-target lesions)
- No post-baseline assessments due to death
- No post-baseline assessments due to other reasons
- All post-baseline assessments have overall response NE
- New anti-cancer therapy started before first post-baseline assessment
- SD of insufficient duration (<6 weeks after first dose of study treatment without further evaluable tumor assessment)
- PD too late (>12 weeks after first dose of study treatment without any evaluable tumor assessment in between)

Special and rare cases where BOR is NE due to both early SD and late PD will be classified as 'SD too early'.

16.2.3 Duration of Response

DR is defined, for subjects with a confirmed objective response according to RECIST 1.1, as the time from first documentation of objective response (CR or PR) to the date of first documentation of objective PD or death due to any cause. If a subject 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 described in Section 16.2.4 for PFS.

DR (months) = (date of event or censoring – date of first objective response + 1)/30.4375

The Kaplan-Meier method will be used to estimate parameters for duration of response. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. In particular, the proportion of duration of responses at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (2) and CIs for the survival function estimates at the above defined time points will be derived using the log-log transformation according to Kalbfleisch and Prentice (3) (CONFTYPE = loglog default option in SAS PROC LIFETEST). The estimate of the standard error will be computed using Greenwood's formula.

16.2.4 Progression Free Survival

The PFS time, according to the RECIST 1.1, is defined as time from first administration of study treatment to the date of the first dcoumation of objective PD or death due to any cause, whichever occurs first.

PFS time (in months) is defined as:

(date of event or censoring - date of first dose of study treatment + 1)/30.4375

PFS data will be censored on the date of the last adequate tumor assessment for subjects who do not have an event (PD or death), for subjects who start a new anti-cancer therapy prior to an event or for subjects with an event after two or more missing tumor assessments. Subjects who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored on the date of the first study treatment unless death occurred on or before the time of the second planned tumor assessment in which case death will be considered an event. Details on determination of the first disease progression date are specified in the IERC charter.

If a tumor assessment was performed on the same day as the start of a new anti-cancer therapy, it will be assumed that the tumor assessment was performed prior to the start of the new anti-cancer therapy; therefore, this tumor assessment will be included in the derivation of the PFS time.

The censoring and event date options to be considered for the PFS and DR analysis are presented in Table 4. Last adequate tumor assessment is defined as tumor assessment result that is not “NE” or “NA”.

Table 4 Outcome and Event Dates for PFS and DR Analyses

Scenario	Date of event / censoring	Outcome
No baseline assessment	Date of first study treatment	Censored [a]
Progression or death ≤ xx weeks after last tumor assessment or ≤ xx weeks after treatment start date, where xx is 12 weeks through 12 months from start of treatment, and 24 weeks thereafter.	Date of progression or death	Event
Progression or death > xx weeks after the last tumor assessment, where xx is 12 weeks through 12 months from start of treatment, and 24 weeks thereafter.	Date of last adequate tumor assessment	Censored
No progression	Date of last adequate tumor assessment	Censored
Treatment discontinuation due to “Disease progression” without documented progression	Not applicable	Information collected on treatment discontinuation page is ignored since ooutcome should be derived based on documented progression only. General censoring rule is applied.
New anti-cancer therapy given	Date of last adequate tumor assessment before anti-cancer therapy is given	Censored

[a] However if the subject dies ≤ 12 weeks after treatment start date the death is an event with date on death date.

The analysis of PFS will be performed using a Kaplan-Meier method with the same approach as for duration of response described in Section 16.2.3. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. In particular, the proportion of PFS at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs.

Frequency (number and percentage) of subjects with each event type (PD or death) and censoring reasons will be presented. Censoring reasons are as follows, presented in a hierarchical manner:

- No baseline assessment
- No post baseline assessments
- Start of new anti-cancer therapy
- Event after 2 or more missing assessments
- Withdrawal of consent to continuing tumor assessment
- Lost to follow-up
- Ongoing in the study without an event

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

16.2.5 Response Status at 6 and 12 Months after Start of Study Treatment

The response status at 6 and 12 months after start of study treatment according to RECIST 1.1 and as determined by the IERC will be determined. A subject will be considered to be in response at the given time point (6 or 12 months after start of study treatment) if the subject had a documented response (PR or CR) prior to that time point, and did not have any of the following events up to the given time point:

- death
- documented disease progression according to RECIST 1.1
- lost to follow-up
- withdraw consent

The proportion of subjects with response at 6 and 12 months after start of study treatment will be calculated as the proportion of subjects in response among all subjects that have started study treatment at least 6 (12) months prior to the time of the analysis, respectively. Clopper-Pearson 95% CIs will be reported as well.

16.3 Analysis on Subgroups

Analysis on subgroups as defined in Section 10 will be performed for the primary and key secondary efficacy endpoints. In addition, Fisher's exact test will be performed to assess the association between each subgroup and ORR. Furthermore, to compare PFS time and OS time within each subgroup, the hazard ratio and its 95% CI of each subgroup will be calculated using Cox proportional hazards model.

A forest plot will be created to graphically present ORR and its 95% CIs for the FAS and all the subgroups.

CCI [redacted]. In case of low number of subjects within a category (< 11 subjects, which is about 10% of the study population), subgroups will be pooled.

16.4 Other Endpoint Analyses

Analysis sets: FAS

16.4.1 Time to Response per RECIST 1.1

Time to response (TTR) is defined, for subjects with an objective response, as the time from the start of first study treatment to the first documentation of objective response (CR or PR) according to RECIST 1.1, as determined by an IERC which is subsequently confirmed.

TTR (in weeks) = (date of first objective response – date of first dose of study treatment + 1)/7

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

CCI [redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

16.5 Pharmacokinetic Data

Analysis sets: Safety analysis set

The PK analysis set will consist of all subjects who receive at least one dose of Avelumab, and provide at least one measurable post-dose concentration.

16.5.1 Descriptive PK Analysis

Avelumab concentrations in serum will be listed in tables and descriptively summarized by treatment, day and nominal time using the number of non-missing observations (N), number of missing observations (NMiss) arithmetic mean (Mean), standard deviation (SD), coefficient of variation (CV%), standard error of the mean (SEM), minimum (Min), median (Median) and maximum (Max). The pre-dose samples will be considered as if they had been taken simultaneously with the administration. Additional tables listing trough concentrations (Cmin) per cohort/day/administration will also be provided, and summarized descriptively using the aforementioned statistics.

For the calculation of descriptive statistics, values as presented in the data listings will be used. Concentrations will be set to missing in summary tables if the value is reported as N.R. (No result). PK concentrations below the lower limit of quantification (<LLOQ) are taken as zero for descriptive statistics. Minimum and maximum will be presented to the same decimal precision as collected. Mean, median and GeoMean will be presented to one decimal place more than the precision of the data collected. SD and SEM will be presented to two decimal places more than the precision of the data collected. The coefficient of variation (CV%), GeoCV and the 95% CI will be reported to 1 decimal place.

Drug concentration in serum for ADA ever-positive subjects versus ADA never-positive subjects will be descriptively summaried to evaluate the potential effect of ADA on pharmacokinetics.

16.5.2 Population Pharmacokinetic Analysis

Sparse sampled PK profiles from study EMR100070-003 Part B will be analyzed jointly with data CCI and other emergent studies with evaluable PK data by non-linear mixed effect approach, in order to describe the PK concentration time profile followed by multiple dose infusion of avelumab, to identify covariates explaining (part of) the between patient PK variability and to estimate the residual PK inter-individual variability. More details will be given in a separate Data Analysis Plan for Poplation Pharmacokinetic Analysis. The results will be reported separately.

16.5.3 Relation of Pharmacokinetics to Efficacy

The exposure-response analysis will include FAS subjects for primary analysis from Part B of the study.

Each subject will be classified as responder or non-responder based on BOR according to RECIST 1.1 (responder: CR and PR; non-responder: stable disease [SD], progressive disease [PD], non-evaluable [NE]). The results will be reported separately.

CCI

The influence of the exposure parameter CCI on response will be explored graphically by depicting response (no/yes) versus exposure and applying a spline function to the data.

The relationship between exposure and response will then assessed using logistic regression.

CCI

CCI [redacted]

Based on this model, the probability of being responder will be depicted versus CCI [redacted]. A 95% confidence interval will the added to the graph.

In a next step, covariates will be inserted in the logistic regression model.

CCI [redacted]

Significance of covariates CCI [redacted] will be assessed using the Wald statistic (alpha=0.05). All covariates will be inserted in the logistic regression model simultaneously and insignificant covariates will be eliminated one by one (starting with the most insignificant covariate) until all remaining covariates are significant.

The odds ratios with 95% confidence intervals for the significant covariates will be calculated and graphically presented in a Forest plot.

Covariates to be tested include, but are not limited to, the following:

CCI [redacted]

CCI [redacted]

All exposure-response calculations will be done using the software program R or other software.

CCI [redacted]

[redacted]

[redacted]

[redacted]

17 Safety Evaluation

Analysis sets: Safety analysis set

17.1 Adverse Events

Treatment emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period or if the worsening of an event is during the on-treatment period.

All analyses described in Section 17.1 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). A separate listing including AEs started after the on-treatment period will also be provided.

- **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 (i.e. no answer to the question "Relationship with study treatment").
- **Serious Adverse Events (SAE):** serious adverse events (as recorded on the AE eCRF page, Serious Adverse Event = Yes).
- **Adverse Events Leading to 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):** immune related adverse events according to case definition classified by medical review as specified in Appendix II.
- **Infusion Related Reactions (IRR)**: IRRs as specified in Appendix II.

Unless otherwise stated adverse events will be displayed in terms of frequency tables: PT and primary System Organ Class (SOC) by decreasing frequency. If an adverse event is reported for a given subject more than once during treatment, the worst severity and the worst relationship to study treatment will be tabulated. Each subject will be counted only once within each PT or SOC and recording period. If a subject experiences more than one AE within a PT or SOC for the same recording period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity.

Changes in toxicity grade, seriousness or outcome of AEs are recorded as separate entries in the eCRF with associated end and start dates (start date equals end date of previous entry). Such entries reporting the same event in such immediately consecutive periods will be considered as one event in the analysis in case of an improvement in toxicity grade. These events will be kept as separate records in the database in order to maintain the full detailed history of the events. The start date of the initial record in the sequence is taken as start date of the entire event. Duration of the AE and the TEAE flag will be adjusted accordingly in the analysis.

17.1.1 All Adverse Events

Adverse events will be summarized by worst severity (according to NCI-CTCAE version 4.0) per subject, 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 subject had 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 subjects with each of the following summaries:
  - TEAEs
  - TEAEs, grade ≥ 3
  - Related TEAEs
  - Related TEAEs, grade ≥ 3
  - TEAEs leading to permanent treatment discontinuation
  - Related TEAEs leading to permanent treatment discontinuation
  - Serious TEAEs
  - Related serious TEAEs
  - TEAEs leading to death

- Related TEAEs leading to death
- Treatment emergent irAEs
- Related treatment emergent irAEs
- Treatment emergent IRRs
- Related treatment emergent 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
- TEAEs by SOC and PT: displaying in separate columns the All TEAEs / Related TEAEs / Grade ≥3 TEAEs / Related Grade ≥3 TEAEs
- TEAEs excluding SAEs, with frequency ≥ 5% by SOC and PT

A listing of AEs including all relevant information will be provided. In addition, a listing of non-TEAEs will also be provided.

17.1.2 Adverse Events Leading to Treatment Discontinuation

The following overall frequency tables will be prepared for the adverse event actions that lead to treatment discontinuation:

- AEs leading to permanent treatment discontinuation by SOC and PT
- Related AEs leading to permanent treatment discontinuation by SOC and PT

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

17.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

17.2.1 Deaths

All deaths, deaths within 30 days after last dose of study drug, death within 60 days after first dose as well as the primary reason for death, will be tabulated based on information from the "Report of Subject Death" and "Survival Follow-Up" eCRFs.

- Number of Deaths
- Number of Deaths within 30 days after last dose of avelumab
- Number of Deaths within 60 days after first dose of avelumab

- Primary Reason of Death
  - Disease progression
  - Adverse event related to study treatment
  - Adverse event not related to study treatment
  - Other
  - Unknown

In addition, date and cause of death will be provided in individual subject data listing together with selected dosing information (date of first / last administration, number of infusions received overall) 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 study treatment
- flag for death within 60 days of first study treatment

17.2.2 Serious Adverse Events

The following overall frequency tables will be prepared for serious adverse events (SAEs):

- Incidence of SAEs by SOC and PT
- Incidence of related SAEs by SOC and PT

The listings of SAEs will also be provided with the relevant information.

17.2.3 Other Significant Adverse Events

The following tables will be created for treatment emergent irAEs:

- The overall summary of irAEs table will include the following categories:
  - All irAEs
  - irAEs leading to death
  - Related irAEs leading to death
  - irAEs
  - irAEs, Grade ≥ 3
  - Related irAEs
  - Related irAEs, Grade ≥ 3
  - irAEs leading to permanent treatment discontinuation
  - Related irAEs leading to permanent treatment discontinuation

  - irAEs by worst grade
  - Related irAEs by worst grade
- irAEs leading to death, by subcategory and PT
- Related irAEs leading to death, by subcategory and PT
- irAEs by subcategory and PT
- irAEs, grade ≥ 3, by subcategory and PT
- Related irAEs by subcategory and PT
- Related irAEs, grade ≥ 3, by subcategory and PT
- irAEs leading to permanent treatment discontinuation by subcategory and PT
- Related irAEs leading to permanent treatment discontinuation by subcategory and PT
- irAEs by subcategory and PT and worst grade
- Related irAEs by subcategory and PT and worst grade

The listing of irAEs will also be provided with the relevant information. The following information will be provided for each irAE:

- The time from first study treatment to start of AE
- The time from most recent study treatment until start of AE
- The duration of the AE

In addition, a listing of irAEs with onset after the on-treatment period will also be provided.

The frequency (number and percentage) of subjects with each of the following will be presented for treatment emergent IRRs:

- IRRs leading to death, by PT
- Related IRRs leading to death, by PT
- IRRs, by PT
- IRRs, Grade ≥ 3, by PT
- Related IRRs, by PT
- Related IRRs, Grade ≥ 3, by PT
- IRRs leading to permanent treatment discontinuation, by PT
- Related IRRs leading to permanent treatment discontinuation, by PT
- Serious IRRs, by PT
- Related 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.

CCI [redacted]

17.2.4 Immunogenicity Subgroup Analysis of Adverse Events

A listing of immunogenicity data and TEAEs will be provided containing subject ID, age, gender, study treatment start and stop date, all dates with positive ADA result, all dates with positive nAb results, preferred term of TEAE, TEAE start and stop date, CTCAE toxicity grade, and flags for SAE, irAE, IRR, and/or reason for permanent treatment discontinuation, as well as ADA status group.

Only subjects who are pre-existing positive, transient treatment-emergent positive, or persistent treatment-emergent positive will be listed.

17.3 Clinical Laboratory Evaluation

17.3.1 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 (e.g., hypokalemia Grade 1 and Grade 2 are only distinguished by a non-numerical qualifier and therefore Grade 2 will not be derived). Additional laboratory results that are not part of NCI-CTCAE will be presented according to the categories: below normal limit, within normal limits and above normal limit (according to the laboratory normal ranges).

Quantitative data will be summarized using simple descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each nominal visit over time (unscheduled measurements would therefore not be included in these summaries as described in Section 11). End of Treatment visit will be summarized separately. The changes computed will be the differences from baseline. Qualitative data based on reference ranges will be described according to the categories (i.e. Low, Normal, High).

Abnormalities classified according to NCI-CTCAE toxicity grading version 4.03 will be described using the worst grade. For those parameters which are graded with two toxicities such as potassium (hypokalemia/hyperkalemia), the toxicities will be summarized separately. Low direction toxicity (e.g. hypokalemia) grades at baseline and post baseline will be set to 0 when the variables are derived for summarizing high direction toxicity (e.g. 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:
  o derived absolute count does not meet Grade 2-4 criteria, and
    - % value < % LLN value, and
    - derived absolute count ≥ 800/mm3
- Neutrophil count decreased:
  o 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. Corrected Calcium is calculated from Albumin and Calcium as follows

Corrected Calcium (mmol/L) = Calcium (mmol/L) + 0.02 [40 - Albumin (g/L)].

Liver function tests: Alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total bilirubin 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 subjects with each of the following during the on-treatment period will be summarized:

- ALT ≥ 3×ULN, ALT ≥ 5×ULN, ALT ≥ 10×ULN, ALT ≥ 20×ULN
- AST ≥ 3×ULN, AST ≥ 5xULN, 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, i.e., a subject 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 and percentage.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created to graphically display

- 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).

In addition, a listing of all TBILI, ALT, AST and ALP values for subjects with a post-baseline TBILI ≥ 2×ULN, ALT≥ 3×ULN or AST ≥ 3×ULN will be provided.

Parameters with NCI-CTC grades available:

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

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

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

- 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 bilitubin increased, Cholesterol (cholesterol high), Creatinine (creatinine increased), Creatine Kinase (CPK increased), Potassium (hypokalemia/hyperkalemia), Sodium (hyponatremia/ hypernatremia), Magnesium (hypomagnesemia/hypermagnesemia), Calcium (hypocalcemia/ hypercalcemia), Glucose (hypoglycemia/hyperglycemia), Gamma Glutamyl Transferase (GGT) (GGT increased), Lipase (lipase increased), Phosphates (hypophosphatemia), Triglycerides (hypertriglyceridemia).

Parameters with NCI-CTC grades not available:

Hematology and chemistry evaluations which can't be graded per CTCAE criteria will be summarized as:

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

In this study, these apply to the following parameters:

- Hematology:

  Hematocrit, Red Blood Cell (RBC), Reticulocytes, Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin Concentration (MCHC).

- Serum Chemistry:

  Chloride, C-Reactive Protein, Lactate Dehydrogenase (LDH), Total Protein, Total Urea, Uric Acid.

Subjects without post baseline laboratory samples will be excluded from analyses with respect to values after the baseline.

17.3.2 Other Laboratory Parameters

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

- Coagulation: activated partial thromboplastin time (aPTT) and prothrombin time (INR).
- Urinalysis: all urinalysis parameters
- Other parameters: hormone, and immunology parameters
- Pregnancy test

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

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

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

17.4 Vital Signs

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

All vital sign parameters will be summarized using descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each visit over time based on visit windows specified in Section 11. The changes computed will be the differences from baseline.

For each subject the maximum on-treatment increase/decrease from baseline will be calculated and categorized by vital sign measurement (see Table 7). Missing values will define a separate category.

Table 7 Vital Sign Parameters Change from Baseline Category

Parameters	Categories of Change from Baseline
Body temperature increase < 37°C, 37-<38°C, 38-<39°C, 39-<40°C, ≥ 40°C	< 1°C , 1-<2°C , 2-<3°C, ≥ 3 °C
Weight increase	<10%, ≥ 10%
Weight decrease	<10%, ≥ 10%
Heart rate increase <100 bpm ; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
Heart rate decrease <100 bpm ; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
SBP increase <140 mmHg; ≥ 140 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
SBP decrease <140 mmHg; ≥ 140 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP increase <90 mmHg; ≥ 90 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP decrease <90 mmHg; ≥ 90 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
Respiration rate increase <20 bpm ; ≥ 20 bpm	≤5 bpm, >5 – 10 bpm, >10 bpm
Respiration rate decrease <20 bpm ; ≥ 20 bpm	≤5 bpm, >5 – 10 bpm, >10 bpm

bpm = beats per minute for heart rate and breaths per minute for respiration rate;
DBP=diastolic blood pressure; SBP=systolic blood pressure

The following summaries will be prepared for vital sign parameters as grouped above considering only subjects with post baseline values:

- Maximal Shifts (changes in categories)
- Listing of highest change per subject

An additional subject data listing will present all changes from baseline reported in the highest categories.

17.5 Other Safety or Tolerability Evaluations

17.5.1 ECG

The incidence and percentage of subjects with potentially clinically significant abnormalities (PCSA) for 12-lead Electrocardiogram (ECG) parameters will be summarized during the on-treatment period. The PCSA criteria are provided in Table 8 below.

Table 8 Potentially Clinically Significant Abnormalities criteria for ECG

Test	Potentially Clinically Significant Abnormalities (PCSA) Criteria
Heart Rate (HR)	≤ 50 bpm and decrease from baseline ≥ 20 bpm ≥ 120 bpm and increase from baseline ≥ 20 bpm
PR Interval	≥ 220 ms and increase from baseline ≥ 20 ms
QRS	≥ 120 ms
QTcF and QTcB absolute	Interval > 450 msec Interval > 480 msec Interval > 500 msec
QTcF and QTcB change from baseline	Increase from baseline > 30 msec Increase from baseline > 60 msec

QTc will be corrected based on Fridericia's formula for QTcF and Bazett's formula for QTcB (QTcF= QT/ $\sqrt[3]{RR}$ and QTcB= QT/ $\sqrt[2]{RR}$) where RR=60/HR. Baseline QTcF and QTcB will be derived from the visit that other ECG parameters are flagged as baseline. If there are multiple assessments at the same visit and time point, the average will be calculated for each parameter and used for the analysis.

Quantitative data for HR, PR, QRS, QTcF and QTcB will also be examined for trends using descriptive statistics (n, missing, mean, SD, median, Q1, Q3, minimum, maximum, and confidence interval) of actual values with 95% CI, absolute changes from baseline to each visit with 90% CI based on visit windows specified in Section 11.

Listings of 12-lead ECGs will be provided with all relevant information and derived variables.

17.5.2 ECOG Performance Status

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

17.5.3 Immunogenicity

ADA was assessed before the trial treatment start, and on Days 15, 29, 43, and 85 prior to the start of infusion, then every 6 weeks until Week 25, and every 12 weeks thereafter. ADA was also assessed at the End-of-Treatment and Safety Follow-up visits. If the sample is positive for ADA, it will be re-analyzed to determine the titer and nAb. The ADA results will be derived based on the algorithm in Table 9. Subjects will be characterized into different ADA categories based on the criteria in

Table 10.

Table 9 Algorithm for the Derivation of ADA Results

Sample Screen Result	Confirmatory	Titer	ADA Result
Negative	NA	NA	Negative
NR	NA	NA	NR
Positive	Negative	NA	Negative
Positive	NR	NA	NR
Positive	Positive	Number	Number
Positive	Positive	NR	Positive-TNR

NR = no result, NA = not applicable, TNR = titer no result.
Negative, number, or positive-TNR are valid results, while number and positive-TNR are considered as positive.

Table 10 Subjects Characterized based on ADA Results

Category	Definition	Subject at Risk (Denominator for Incidence)
Never positive	No positive results at any time point	Number of subjects with at least one valid result at any time point
Ever positive	At least one positive result at any time point	Number of subjects with at least one valid result at any time point
Pre-existing	A positive ADA result prior to treatment with avelumab	Number of subjects with valid baseline result
Treatment boosted	A positive ADA result prior to treatment with avelumab and the titer ≥ 8*baseline titer while on avelumab treatment	Number of subjects with valid baseline and at least one valid post-baseline result
Treatment emergent	Not positive prior to treatment with avelumab and with at least one positive post-baseline result	Number of subjects with at least one valid post-baseline result and without positive baseline results (including missing, NR)
Transient positive	If treatment emergent subjects have (a single positive evaluation, or duration between first and last positive result <16 weeks) and last assessment not positive.	Number of subjects with at least one valid post-baseline result and without positive baseline results (including missing, NR)

Category	Definition	Subject at Risk (Denominator for Incidence)
Persistent positive	If treatment emergent subjects have duration between first and last positive result ≥16 weeks or a positive evaluation at the last assessment	Number of subjects with at least one valid post-baseline result and without positive baseline results (including missing, NR)

Samples with a reportable ADA titer will also be tested in the neutralizing antibody (nAb) assay. NAb results are positive or negative in a single assay and only derived when not performed because ADA was negative (see Table 11). Subjects will be characterized into different nAb categories based on the criteria in Table . For nAb, treatment-boosted is not applicable since no titer result is available.

Table 11 Algorithm for the Derivation of nAb Results

ADA Confirmatory Result	nAb Result	Derived nAb Result
Negative	NA	Negative
NR	NA	NR
NA	NA	Negative
Positive	NR	NR
Positive	Positive	Positive
Positive	Negative	Negative

ADA = antidrug antibody, NA = not applicable, nAb = neutralizing antibody, NR = no result.

Table 12 Subjects Characterized based on nAb Results

Category	Definition	Subject at Risk (Denominator for Incidence)
Never positive	No nAb positive results at any time point	Number of subjects with at least one valid ADA result at any time point
Ever positive	At least one nAb positive result at any time point	Number of subjects with at least one valid ADA result at any time point
Pre-existing	A positive nAb result prior to treatment with avelumab	Number of subjects with valid ADA baseline result
Treatment emergent	Not nAb positive prior to treatment with avelumab and with at least one nAb positive post-baseline result	Number of subjects with at least one ADA valid post-baseline result and without nAb positive baseline results (including missing, NR)
Transient positive	If treatment emergent subjects have (a single nAb positive evaluation, or duration between first and last nAb positive result <16 weeks) and last ADA assessment not nAb positive.	Number of subjects with at least one ADA valid post-baseline result and without nAb positive baseline results (including missing, NR)
Persistent positive	If treatment emergent subjects have duration between first and last nAb positive result ≥16 weeks	Number of subjects with at least one ADA valid post-baseline result and without nAb

Category	Definition	Subject at Risk (Denominator for Incidence)
	or a nAb positive evaluation at the last ADA assessment	positive baseline results (including missing, NR)

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

The frequency and percentage of each ADA and nAb category will be presented in tables. Listings of ADA and nAb results from ADA ever positive subjects will be prepared.

For the ADA ever-positive subjects, a listing will be prepared with subject ID, start and stop of treatment, date of onset, time to onset (weeks since treatment start) and last date of ADA positive results, as well as date of onset, time to onset and last date of nAb positive results, BOR and BOR date, DOR, PFS time or censoring time and reason for censoring, and OS time or censoring time and reason for censoring.

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

18 References

1. Iyer JG, Blom A, Doumani R, et al. Response rates and durability of chemotherapy among 62 patients with metastatic Merkel cell carcinoma. Cancer Med. 2016 Jul 19 [Epub ahead of print].
2. Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982;38:29–41.
3. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons 1980.
4. Eisenhauer, E.A, Therasse, P., Bogaerts. J. et. al. (2009): New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer 45, 228-247.
5. Beal S. L. Ways to fit a PK model with some data below the quantification limit. J Pharmacokin Pharmacodyn. 28: 481 – 504 (2001).

19 Appendices

Appendix I Important Protocol Deviations (Identified by Medical Review or Programmed)

Inclusion and exclusion criteria are referred according to Protocol version 11.0 dated 17-May-2017

	Category of Protocol Deviation	Description of Protocol Deviation	Deviation Code	Clinically Important	Protocol Section	Proposed check / comment
Inclusion criteria: **For the subject to be eligible for inclusion, each criterion must be checked 'YES':**						
Signed written informed consent.	Informed Consent Criteria	Subject did not meet inclusion criterion 1.	PDEV01		Section 5.3.1	Medical review required
Male or female subjects aged >=18 years	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 2.	PDEV02		Section 5.3.1	Medical review required
Histologically proven MCC	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 3.	PDEV03	Yes	Section 5.3.1	Medical review required
Subjects must not have received any prior systemic treatment for metasatic MCC. Prior chemotherapy treatment in the adjuvant setting (no clinically detectable disease; no metastatic disease) is allowable if the end of treatment occurred at least 6 months prior to study start	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 3d.	PDEV18	Yes	Section 5.3.1	Medical review required
Disease must be measurable with at least 1 unidimensional measurable lesion by RECIST 1.1 (including skin lesions)	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 7.	PDEV04	Yes	Section 5.3.1	Medical review required
Effective contraception for both male and female subjects if the risk of conception exists	Eligibility and Entry Criteria	Subject did not meet inclusion criterion 11.	PDEV05		Section 5.3.1	Medical review required

	Category of Protocol Deviation	Description of Protocol Deviation	Deviation Code	Clinically Important	Protocol Section	Proposed check / comment
Exclusion criteria: **For the subject to be eligible for inclusion, each criterion must be checked 'NO':**						
Concurrent treatment with a non-permitted drug	Eligibility and Entry Criteria	Subject met exclusion criterion 2.	PDEV06		Section 5.3.2	Medical review required
Prior therapy with any antibody/drug targeting T cell co-regulatory proteins	Eligibility and Entry Criteria	Subject met exclusion criterion 3.	PDEV07	Yes	Section 5.3.2	Medical review required
Concurrent anti-cancer treatment (for example, cytoreductive therapy, radiotherapy [with the exception of palliative bone-directed radiotherapy, or radiotherapy administered to non-target superficial skin lesions], immune therapy, or cytokine therapy except for erythropoietin)	Eligibility and Entry Criteria	Subject met exclusion criterion 4.	PDEV08		Section 5.3.2	Medical review required
Subjects with active central nervous system (CNS) metastases are excluded. Subjects with a history of treated CNS metastases (by surgery or radiation therapy) are not eligible unless they have fully recovered from treatment, demonstrated no progression for at least 2 months, and do not require continued steroid therapy	Eligibility and Entry Criteria	Subject met exclusion criterion 7.	PDEV19	Yes	Section 5.3.2	Medical review required
Previous malignant disease (other than MCC) within the last 5 years with the exception of basal or squamous cell carcinoma of the skin or cervical carcinoma in situ	Eligibility and Entry Criteria	Subject met exclusion criterion 8.	PDEV09		Section 5.3.2	Medical review required
Pregnancy or lactation	Eligibility and Entry Criteria	Subject met exclusion criterion 14.	PDEV10		Section 5.3.2	Medical review required

	Category of Protocol Deviation	Description of Protocol Deviation	Deviation Code	Clinically Important	Protocol Section	Proposed check / comment
Prohibited Medication Criteria						
Non-permitted concomitant medication during the study	Prohibited Medications	Subject took Immunotherapy, immunosuppressive drugs (that is, chemotherapy or systemic corticosteroids except for short-term treatment of allergic reactions or for the treatment of irAEs), or other experimental pharmaceutical products.	PDEV11		Section 6.5.2	Medical review required
Other criteria						
Subjects that developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subject became pregnant, but continued on the study.	PDEV12		Section 5.5.2	
Subjects who developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subjects received treatment despite confirmed progression and significant clinical deterioration. Subjects had ECOG>=3, did not resolved to <=2 by day 14 of next cycle, and continued on the study.	PDEV13		Section 5.5.2 and 5.5.3	Medical review required
Subjects who developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subject developed grade 4 ADR, but continued on the study.	PDEV14		Section 5.5.2 and 5.5.3	Medical review required
Subjects who developed withdrawal criteria whilst on the study but were not withdrawn;	Other Criteria	Subject developed grade 3 ADR, but continued on the study.	PDEV15		Section 5.1.4.2	Medical review required
Subjects dosing error (≥ +/- 10 % assigned dose)	Other	Dosing error	PDEV16		Section 6.2	List if relative dose intensity ≥1.10 or ≤ 0.90
Subjects on avelumab treatment did not receive premedication	IP compliance	Subjects on avelumab treatment did not take mandatory premedication	PDEV17		Section 6.2	Medical review required

	Category of Protocol Deviation	Description of Protocol Deviation	Deviation Code	Clinically Important	Protocol Section	Proposed check / comment
Any other protocol deviation which is deemed to be significant but has not been pre-specified in this table	any	any	PDEV99		NA	Medical review required

Appendix II Description of criteria for Adverse Events of Special Interest

A two-level approach is pursued to analyze potential immune-related adverse events (AEs):

1. A MedDRA Preferred Term (PT) query is used for each event category (i.e., immune-mediated rash, colitis, pneumonitis, hepatitis, nephritis and renal dysfunction, endocrinopathies and other immune-mediated adverse reactions).

2. AEs identified by the MedDRA PT queries will then be medically reviewed using pre-defined case definitions for immune-mediated adverse reactions.

Level 1:

To identify potentially immune-mediated AEs, MedDRA PT queries will be used to search for AEs of interest in the clinical database. The relevant event categories and corresponding MedDRA PT queries are documented in a version-controlled repository maintained by the Sponsor and finalized for analysis of the current study data prior to DB lock.

The respective MedDRA PT queries will be used to identify AEs reported in the current study that are potentially immune-mediated events.

Level 2:

In a second level (medical review), the potential immune-mediated AEs identified from the search performed at Level 1, will be reviewed by qualified medical personnel to determine whether the AE meets the criteria (case definition) for an immune-mediated adverse reaction based on the following algorithm:

Table 13 Algorithm for immune-related adverse reactions

Criteria	Description
Onset	AE onset after 1st avelumab administration until up to 90 days after last dose
Duration	AE does not spontaneously resolve (i.e., without corticosteroids/ immunosuppressant treatment) within 7 days after onset
Immunosuppressive therapy	AE treated with corticosteroid or other immunosuppressant therapy. *For endocrinopathies only*: AE required hormone replacement* and /or (corticosteroid or other immunosuppressive therapy)
Etiology	No other clear etiology or Histopathology/biopsy consistent with immune-mediated event
All criteria listed in the left column need to be fulfilled for an event to meet the case definition of immune-mediated reaction.	
*Hormone replacement will be evaluated for specific endocrinopthy disorders only as follows: • Thyroid disorders (HLT): Thyroid therapy (ATC codes (H03A, H03B)) • Diabetes mellitus (including hyperglycemia): Insulin (ATC code A10A)	

Infusion related reactions are identified based on a list of MedDRA PTs and criteria on the timely relationship according to Table 14.

Table 14 Criteria for infusion related reactions

Infusion related reactions	**Reactions - Considered when onset is on the day of avelumab infusion (during or after the infusion) or the day after the avelumab infusion (irrespective of resolution date):** • Infusion related reaction • Drug hypersensitivity • Anaphylactic reaction • Hypersensitivity • Type 1 hypersensitivity **Signs and Symptoms - occurring on the day of avelumab infusion (during or after the infusion) and resolved with end date within 2 days after onset** • Pyrexia • Chills • Flushing • Hypotension • Dyspnea • Wheezing • Back pain • Abdominal pain • Urticaria