# Statistical Analysis Plan

| | |
|---|---|
| **Clinical Trial Protocol Identification No.** | EMR 100070-008 |
| **Title:** | A Phase III open-label, multicenter trial of avelumab (MSB0010718C) as a third-line treatment of unresectable, recurrent, or metastatic gastric or gastroesophageal junction adenocarcinoma |
| **Trial Phase** | Phase III |
| **Investigational Medicinal Product(s)** | MSB0010718C Avelumab |
| **Clinical Trial Protocol Version** | 30 May 2017/Version 7.0 |
| **Statistical Analysis Plan Author** | PPD Statistical Scientist |
| **Statistical Analysis Plan Date and Version** | 12 October 2017 /Version 4.0 |
| **Statistical Analysis Plan Reviewers** | PPD PPD<br>PPD<br>PPD PPD<br>PPD PPD PPD<br>PPD PPD<br>PPD , PPD<br>PPD<br>PPD<br>PPD<br>PPD |

[redacted]



[redacted]

# 1 Signature Page

**Statistical Analysis Plan:** EMR 100070-008

A Phase III open-label, multicenter trial of avelumab (MSB0010718C) as a third-line treatment of unresectable, recurrent, or metastatic gastric or gastroesophageal junction adenocarcinoma

PPD

PPD

Date of Signature PPD

Trial Statistician

PPD

PPD

PPD

Signature

PPD

Date of Signature PPD

Project Statistician
EMD Serono Inc.
45A Middlesex Turnpike
Billerica, MA 01821

PPD

PPD

Signature

PPD

Date of Signature

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

# 2 Table of Contents

# 3 List of Abbreviations and Definition of Terms

| | |
|---|---|
| ADA | Anti-drug Antibody |
| AE | Adverse Event |
| ALK | Anaplastic Lymphoma kinase |
| ALT | Alanine Aminotransferase |
| ANCOVA | Analysis of Covariance |
| AST | Aspartate Aminotransferase |
| ATC | Anatomical Therapeutic Chemical |
| BMI | Body Mass Index |
| BOR | Best Overall Response |
| BSA | Body Surface Area |
| BSC | Best Supportive Care |
| CI | Confidence Interval |
| CIPD | Clinically Important Protocol Deviations |
| CMH | Cochran-Mantel-Haenszel |
| CR | Complete Response |
| CRF | Case Report Form |
| CSR | Clinical Study Report |
| CT | Computed Tomogrophy |
| CTP | Clinical Trial Protocol |
| DCR | Disease Control Rate |
| DR | Duration of Response |
| ECG | Electrocardiogram |
| ECOG PS | Eastern Cooperative Oncology Group Performance Status |
| eCRF | Electronic Case Report Form |
| eDISH | Evaluation of Drug-Induced Serious Hepatotoxicity |
| EORTC | European Organization for Research and Treatment of Cancer |
| EQ-5D-5L | European Quality of Life (EuroQoL) - 5 Dimensions - 5 Levels |
| EU | European Union |
| FAS | Full Analysis Set |
| HR | Hazard Ratio |
| HRQoL | Health-related Quality of Life |

| | |
|---|---|
| IC | Immune Cell |
| ICF | Informed Consent Form |
| ICH | International Conference on Harmonization |
| IDMC | Independent Data Monitoring Committee |
| IP | Investigational Product |
| IPD | Important Protocol Deviation |
| IRC | Independent Review Committee |
| irAE | Immune-related Adverse Event |
| IRR | Infusion-related Reaction |
| ITT | Intention-to-Treat |
| IV | Intravenous |
| IWRS | Interactive Web Response System |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MRI | Magnetic Resonance Imaging |
| nAb | Neutralizing Antibody |
| NCI-CTCAE | National Cancer Institute – Common Terminology Criteria for Adverse Events |
| nd | Not Done |
| NE | Non-evaluable |
| ORR | Objective Response Rate |
| OS | Overall Survival |
| PCSA | Potentially Clinically Significant Abnormalities |
| PD | Progressive Disease |
| PT | Preferred Term |
| Pd | Pharmacodynamics |
| PFS | Progression-Free Survival |
| PGt | Pharmacogenetics |
| PK | Pharmacokinetics |
| PP | Per-Protocol |
| PR | Partial Response |
| Q1 | First Quartile |
| Q3 | Third Quartile |

| | |
|---|---|
| QoL | Quality of Life |
| RECIST | Response Evaluation Criteria in Solid Tumors |
| SAE | Serious Adverse Event |
| SAP | Statistical Analysis Plan |
| SD | Stable Disease |
| StDev | Standard Deviation |
| SDTM | Study Data Tabulation Model |
| SOC | System Organ Class |
| TC | Tumor Cell |
| TEAE | Treatment Emergent Adverse Event |
| VAS | Visual Analogue Scale |

# 4 Modification History

| Unique Identifier for SAP Version | Date of SAP Version | Author | Changes from the Previous Version |
|---|---|---|---|
| 1.0 | 20 November 2015 | PPD | NA. The first version. |
| 2.0 | 27 May 2016 | PPD | Details of changes are specified in Section 5.1 "Changes to Previous Version" |
| 3.0 | 31 July 2017 | PPD | Details of changes are specified in Section 5.1 "Changes to Previous Version" |
| 4.0 | 12 October 2017 | 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 final analysis of data collected for Clinical Trial Protocol (CTP) EMR 100070-008. 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 may be included in any regulatory submissions and/or 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 CTP and is prepared in compliance with ICH E9. The first version (version 1.0) was created based on the protocol version 3.0 dated 22 October 2015. Version 2.0 was created based on the protocol version 5.0 dated 29 April 2016. Additional updates in version 2.0 and version 3.0 are specified in Section 5.1, Changes to Previous Version.

## 5.1 Changes to Previous Version

**Version 4.0**

The following changes are made in version 4.0 of the SAP:

1. Section 9 Changes to the Planned Analyses in the Clinical Trial Protocol: Details were updated to specify the changes that are deviated from what are in the CTP;
2. Section 10 Analysis Sets: Added rationale of including subject who received BSC only in the safety analysis set; the definition of HRQoL Analysis Set was updated to clarify post-baseline HRQoL assessment;
3. Section 11 General Specification for Statistical Analyses: On-treatment period was updated by adding definition for subjects who received BSC only;
4. Section 17.3.1 Hematology and Chemistry Parameters: Added boxplot for changes from baseline.

**Version 3.0**

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

1. Section 7 Sample Size/Randomization was updated to include the condition of a minimum of 6 months follow-up since the last subject randomized for the primary analysis per protocol version 7;
2. Section 8 Sequence of Analysis was updated to include the condition of a minimum of 6 months follow-up since the last subject randomized for the primary analysis per protocol version 7. Data cut-off date for the primary analysis was updated;
3. Section 10 Analysis Set was updated to include HRQoL analysis set, CCI [redacted] additional subgroup analysis set was added;
4. Section 11 General Specification for Statistical Analyses was updated for definition of baseline and on-treatment period; imputation rule for partial missing subsequent anti-cancer therapy date was added;
5. Section 12.1 Disposition of Subjects and Discontinuations was updated for region and disposition category;
6. Section 12.2 Protocol Deviations was updated with more details on how to summarize important protocol deviations and clinically important protocol deviations;
7. Section 13.1 Demographics was updated for geographic region category;
8. Section 13.3.2 Prior Anti-Cancer Therapies was updated to add summary of number of prior therapies;

C [redacted]
C

10. Section 14.1 Prior and Concomitant Medications/Procedures was updated to include summary table for pre-medications and concomitant procedures;
11. Section 14.2 Subsequent Anti-Cancer Therapies/Procedures was updated to include more details on the summary analysis;
12. Section 15 Treatment Compliance and Exposure was updated to add clarification for the definition of relative dose intensity;
13. Section 16.1.2 Sensitivity Analyses of Primary Endpoint was updated to add additional sensitivity analysis of OS. Details of RMST analysis was added;
14. Section 16.1.4 Time of Follow-up for OS was updated to specify time point of follow-up rate estimates;
15. Section 16.2.1 Progression-free Survival Table 3 was updated for the censoring rules. Multivariate Cox regression analysis on PFS was added;
16. Section 16.2.2. Best Overall Response was updated to include additional details for confirmation of response and summary of BOR = NE; multivariable logistic regression analysis was added as a sensitivity analysis;
17. Section 16.2.3 Time of Follow-up for PFS was added;

18. Section 16.2.4 Subgroup Analysis of Secondary Efficacy Endpoints was added;
19. Section 16.2.5 Health-Related Quality of Life was updated to include completion and compliance summary, mixed-effects model repeated measures analysis and more details on summary of change from baseline;
20. Section 16.3.1 Duration of Response was updated to include DR analysis according to investigator assessment;

CCI [redacted]

24. Section 17.1 Adverse Events definition of irAE and IRR was updated; added details of pooling the same AE with different toxicity grade, outcome or seriousness recorded as different entries on the eCRF;
25. Section 17.2.3 Other Significant Adverse Events was updated to replace SOC by Cluster;

CCI [redacted]

27. Section 17.3.1 Hematology and Chemistry Parameters was updated to specify the denominator for worst-grade during the on-treatment period analysis;
28. Section 17.5.1. ECG was updated to specify the denominator of ECG PCSA summary table;

CCI [redacted]

30. Appendix I Programmed Clinically Important Protocol Deviations was updated per the updated protocol deviation category per Medical team input;
31. Appendix II was added to include the screenshot of EORTC QLQ-C30 version 3.0 scoring system;
32. Appendix III was added to include the screenshot of EORTC QLQ-C30-STO22 Gastric Cancer Module scoring system;
33. Appendix IV was added to include "Description of the Case Review for Assessment of Immune-Related AEs and IRR".

**Version 2.0**

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

1. The term "Intention-to-Treat (ITT) analysis set" was updated as "Full analysis set (FAS)" throughout this document (the definition remains the same);
2. Section 10 Analysis Set was updated for the per-protocol analysis set; additional subgroup analysis sets were added;

3. Section 11 General Specifications for Statistical Analyses was updated;
4. Section 12.1 Disposition of Subjects and Discontinuations was updated;
5. Section 13.1 Demographics was updated;
6. Section 13.3.1 Disease Characteristics was updated to delete PD-L1 assay status it is not collected on the "Disease History" eCRF page;
7. Section 13.3.2 Prior Anti-Cancer Therapies was updated on the sorting order of summary output;
8. Section 13.3.3 PD-L1 Expression Status at Baseline was added;
9. Section 14.1 Prior and Concomitant Medications/Procedures, definition of prior medications was updated and sorting order of the summary output was updated;
10. Section 16.1.1 Primary Efficacy Analysis of Overall Survival was updated to add additional time points in estimating survival probability;
11. Section 16.1.2 Sensitivity Analyses was updated to include Restricted Mean Survival Time (RMST) as a method for evaluating the validity of Cox proportional hazard model assumptions;
12. Section 16.1.3 was added for exploratory analyses of primary endpoint;
13. Section 16.2.1 Progression-Free Survival was updated to include Table 2 for censoring rule;
14. Section 16.2.2 Best Overall Response was updated to include summary of subjects with BOR; the derivation rule for the BOR was updated to include subjects who only have non-target lesions at baseline;
15. Section 16.2.3 Subject-reported Outcomes/Quality of Life was updated to include the graphical summary of QoL data; symptom scales of QLQ-C30 was added;
16. Section 16.3.1 Duration of Response was updated for the definition of duration of response;
17. Section 16.3.2 Time to Response was updated;
18. Section 16.3.3 Tumor Shrinkage was updated;
19. Section 16.3.4 Disease Control Rate was updated to included Non-CR/Non-PD in the numerator of disease control rate calculation for subjects who have only non-target lesions at baseline;

CCI [redacted]

21. Section 17.1 Adverse Events, definition of infusion-related reactions was updated;
22. Section 17.1.1 All Adverse Events, additional categories were added to the overall AE summary;
23. Section 17.2.4 Immunogenicity Subgroup Analysis of Adverse Events was added;
24. Section 17.3.1 Hematology and Chemistry Parameters was updated;

25. Appendix I Required Full Laboratory Safety Tests was deleted from the SAP as these tests have been clearly identified in the CTP;
26. Programmed Clinically Important Protocol Deviations were added as Appendix I per sponsor SOP.

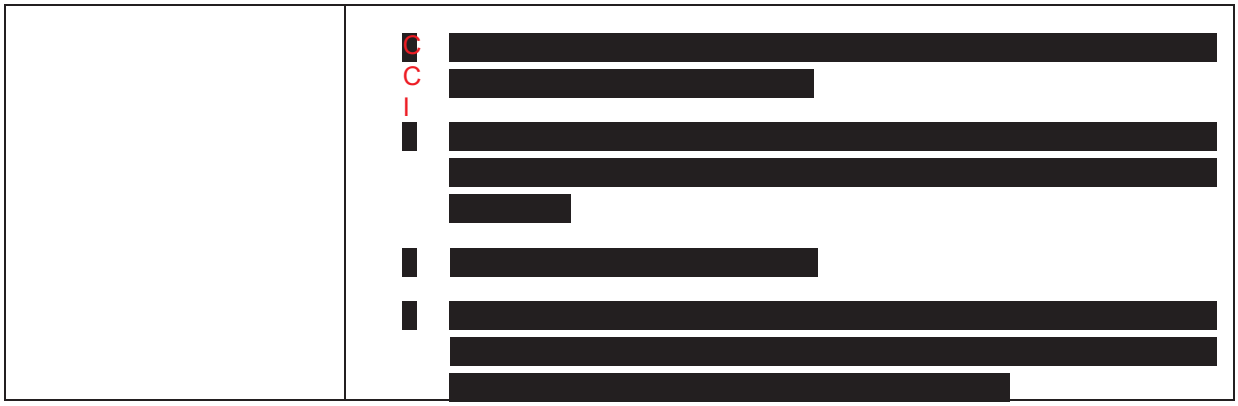
# 6 Summary of Clinical Trial Features

| | |
|---|---|
| **Trial objectives** | **Primary objective**<br>The primary objective is to demonstrate superiority with regard to overall survival (OS) of avelumab plus best supportive care (BSC) versus physician's choice (chosen from a pre-specified list of therapeutic options) plus BSC.<br>**Secondary objectives**<br>Secondary objectives are as follows:<br>• To compare avelumab plus BSC versus physician's choice plus BSC in regard to the following:<br>  • Progression-free survival (PFS) based on an Independent Review Committee (IRC) assessment<br>  • Objective response rate (ORR) based on IRC assessment<br>  • Subject-reported outcomes/quality of life (QoL) using the European Quality of Life (EuroQOL) 5-dimensions and 5-levels questionnaire (EQ-5D-5L), and the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 and module QLQ-STO22.<br>• To determine the safety and tolerability of avelumab.<br>**Exploratory objectives**<br>Exploratory objectives are as follows:<br>• To determine duration of response of avelumab plus BSC versus physician's choice plus BSC based on IRC assessment<br>• To determine time to response of avelumab plus BSC versus physician's choice plus BSC based on IRC assessment<br>• To evaluate tumor shrinkage in target lesions at each time point from Baseline based on IRC assessment<br>• To evaluate the disease control rate (DCR)<br>[redacted] |

| | |
|---|---|
| | [redacted] |
| **Trial design and plan** | This is a multicenter, international, randomized, open-label Phase III trial in subjects with unresectable, recurrent, locally advanced or metastatic adenocarcinoma of the stomach or of the gastroesophageal junction who have failed or relapsed from 2 prior chemotherapeutic regimens administered for the treatment of unresectable, recurrent, locally advanced or metastatic disease.<br><br>All subjects in both arms will receive BSC as background therapy. Subjects will receive BSC with either avelumab or physician's choice chemotherapy from among a prespecified list of therapeutic options or BSC alone with no active therapy.<br><br>Approximately 330 subjects will be randomized, stratified by region (Asia versus non-Asia), in a 1:1 ratio to receive BSC with either avelumab at a dose of 10 mg/kg as a 1-hour intravenous (IV) infusion once every 2 weeks or physician's choice chemotherapy from the following options:<br><br>• Paclitaxel as monotherapy (80 mg/m$^2$ on Days 1, 8, and 15 of a 4-week treatment cycle)<br><br>• Irinotecan as monotherapy (150 mg/m$^2$ on Days 1 and 15 of a 4-week treatment cycle)<br><br>• Subjects who are not deemed eligible to receive paclitaxel or irinotecan at the dose and schedule specified above will receive BSC once every 3 weeks.<br><br>Subjects will return to the clinic at regular intervals for assessments. Tumor measurements by computed tomography (CT) scan or magnetic resonance imaging (MRI) will be performed every 6 weeks for the first 12 months and every 12 weeks thereafter to determine response to treatment. A central imaging laboratory will be used to read and interpret all CT/MRI data. Response will be evaluated using Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1) and as adjudicated by a blinded IRC. Treatment will continue until: |

| | |
|---|---|
| | • Disease progression<br>• Significant clinical deterioration (clinical progression)<br>• Unacceptable toxicity, or<br>• Any criterion for withdrawal from the trial or study treatment is fulfilled.<br><br>For subjects receiving avelumab plus BSC, treatment may continue past the initial determination of disease progression per RECIST v1.1 as long the following criteria are met:<br><br>• No new symptoms or worsening of previous symptoms<br>• Tolerance of study treatment<br>• Stable Eastern Cooperative Oncology Group performance status (ECOG PS)<br>• Treatment beyond progression will not delay an imminent intervention to prevent serious complications of disease progression (for example, central nervous system metastases).<br><br>Subjects receiving avelumab plus BSC who have experienced a complete response (CR) should continue to receive avelumab for at least 12 months after confirmation of response and/or until disease progression or unacceptable toxicity. 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. To be eligible for re-treatment, 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 trial and will be treated and monitored according to the protocol and the "until progression" schedule in the Schedules of Assessments (see Table 1 in the clinical protocol).<br><br>Subjects receiving physician's choice chemotherapy plus BSC or BSC alone will receive study treatment until progressive disease (PD) per RECIST v1.1, significant clinical deterioration (clinical progression), unacceptable toxicity, or if any criterion for withdrawal from the trial or study treatment is fulfilled. Subjects receiving physician's choice plus BSC will not be offered to cross over to avelumab plus BSC.<br><br>Investigators must specify which of the physician's choice treatment regimens will be selected prior to randomization.<br><br>Assessments will be made by the Investigators for the purpose of subject management, but the disease response determinations, |

| | |
|---|---|
| | including PD assessments associated with the trial endpoints, will be supported by tumor assessments performed by an IRC.<br>Subjects will attend clinic visits at regular intervals to receive study treatment and for efficacy and safety assessments.<br>The primary endpoint of the trial is OS, defined as the time (in months) from randomization to the date of death, regardless of the actual cause of the subject's death.<br>Safety endpoints include adverse events (AEs) assessed throughout the trial and evaluated using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) v4.03, physical examination findings, clinical laboratory assessments, vital signs, and electrocardiograms (ECGs). |
| **Planned number of subjects** | Approximately 330 subjects are planned to be enrolled. |
| **Primary endpoint** | The primary endpoint for the trial is OS, defined as the time (in months) from randomization to the date of death, regardless of the actual cause of the subject's death. |
| **Secondary/Exploratory endpoints** | Secondary endpoints include:<br>• PFS according to RECIST 1.1 and as adjudicated by the IRC<br>• BOR according to RECIST 1.1 and as adjudicated by the IRC<br>• Subject-reported outcomes / quality of life (QoL) as assessed by the EQ-5D-5L, the EORTC QLQ-C30, and the EORTC Gastric Cancer Module QLQ-STO22<br>• Safety endpoints (including AEs, physical examination findings, clinical laboratory assessments, vital signs, ECG parameters, and ECOG PS)<br>Exploratory endpoints include:<br>• Duration of response<br>• Time to response<br>• Tumor shrinkage in target lesions at each time point from baseline<br>[redacted]<br>• Disease Control Rate<br>[redacted] |




# 7 Sample Size/Randomization

The sample size of this study is driven by the primary endpoint – OS, therefore the study is event-driven and the primary analysis will take place after 256 deaths have occurred and after a minimum of 6 months follow-up since the last subject was randomized (whichever is later) to ensure sufficient follow-up for all subjects for the primary analysis of OS. With 256 OS events, the study provides 90% power to demonstrate an improvement of 2 months median OS time (i.e., median OS time of 6 months in the avelumab + BSC arm versus 4 months in the control arm, which is equivalent to a reduction of 33% in hazard rate, i.e., hazard ratio (HR) of avelumab + BSC versus control = 0.67, at the one-sided 2.5% overall significance level. Assuming a randomization ratio of 1:1 (avelumab + BSC versus control), an exponential distribution of OS time, a uniform distribution of subject accrual period over 15 months, an additional 6 months minimum follow-up period for all subjects and an expected 5% overall drop-out rate, the study should enroll approximately 330 subjects (n=165 subjects/arm) to achieve 256 OS events in total.

The study is a randomized, active-controlled (based on physician's choice, including paclitaxel or irinotecan + BSC), and open- label study. Eligible subjects will be randomized into either the avelumab + BSC arm or control arm (physician's choice, including paclitaxel or irinotecan + BSC) in a 1:1 ratio using a stratified, fixed block design. The randomization will be stratified by region (Asia vs. non-Asia). The purpose of the stratification is to ensure a balance of the treatment arms within each region. Randomization will occur using the interactive web response system (IWRS) as described in CTP Section 6.3 upon completion of the Screening/Baseline procedures and determination of subject eligibility.

# 8 Overview of Planned Analyses

This SAP covers the analyses for efficacy and safety based on the data cut-off date for the primary analysis. Statistical analyses will be performed using cleaned eCRF data as well as data collected by the external vendor, including CCI patient quality of life data and tumor assessment results adjudicated by the IRC. All data will be included up to a clinical cut-off date which is determined by the requested number of events. The cut-off date is determined once a data extract (before data base lock) is available which indicates that the required number of events have occurred.

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

A separate SAP will cover the analyses for periodic safety reviews conducted by the IDMC.

Separate or supplemental analysis plans might be written to cover:

## 8.1 Sequence of Analysis

The following analyses will be performed during this trial:

- A primary analysis will be performed after approximately 256 events (deaths) have occurred and after a minimum of 6 months follow-up since the last subject was randomized, whichever is later. The actual data cut-off date for the primary analysis will be prospectively determined based on the event projection and the requirement of minimum of 6-month follow-up since the last subject randomized. The actual number of events may be slightly different from 256 at the data cut-off date for the primary analysis.
- There will be ongoing periodic safety reviews by the IDMC. Details will be provided in the IDMC charter and the IDMC SAP (see Section 8.2 as well).

## 8.2 Interim Analysis

No interim analysis for efficacy is planned for this study. An IDMC will be formed and will be responsible for periodic safety evaluations of the trial. The IDMC will be presented with patient disposition, patient background, baseline disease and demographic information, along with safety information as described in the IDMC SAP. An independent statistical provider will perform such partially blinded analyses to support the IDMC. Results from these periodic safety analyses will be transmitted from this independent statistical provider to the IDMC only. Details of the IDMC mission, composition, and operations and of maintaining the blinding of the study are provided in the IDMC charter.

## 8.3 Primary Analysis

The primary analysis will take place after the planned number of OS events (deaths) has occurred and after a minimum of 6 months follow-up since the last subject was randomized, whichever is later.

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.

A data review meeting will be held prior to partial database lock at the time of the primary analysis. In addition, no database can be locked until this SAP has been approved.

# 9 Changes to the Planned Analyses in the Clinical Trial Protocol

The statistical methods as described in the protocol version 7.0 dated 30 May 2017 were adopted.

The definition of Safety Analysis Set is updated to include subjects who received BSC only.

The definition of on-treatment period was updated to include subjects who received BSC only.

# 10 Analysis Sets

**Screening Analysis Set**

The Screening analysis set includes all subjects who signed the Informed Consent Form (ICF).

**Full Analysis Set / Intention-to-Treat (ITT)**

The Full Analysis Set (FAS) / Intention-to-Treat (ITT) analysis set will include all subjects who were randomized to study treatment. Analyses performed on the FAS will take into account subjects' allocation to treatment arms as randomized. The FAS will be the primary analysis set for all primary and secondary efficacy endpoints.

**Per-Protocol Analysis Set**

The Per-Protocol (PP) analysis set is a subset of the FAS and will include subjects who do not meet any of the following criteria that could impact the key objectives of the study. Subjects who meet any of the following criteria will be excluded from the PP analysis set:

- Subjects randomized to one treatment, but received the incorrect treatment.
- Subjects without histologically confirmed unresectable, recurrent, locally advanced or metastatic, adenocarcinoma of the stomach or the gastroesophageal junction, as per inclusion criterion #4.
- Subjects did not receive 2 prior courses of systemic treatment for unresectable recurrent, locally advanced or metastatic gastric cancer, and/or did not progress after the second line, as per inclusion criterion #6.
- Subjects received "prior therapy with any antibody or drug targeting T-cell coregulatory proteins (immune checkpoints), such as PD-1, PD L1, or cytotoxic T-lymphocyte antigen-4", as per exclusion criterion #1.
- Subjects received concurrent anticancer treatment, as per exclusion criterion 2.
- Subjects who met any of the exclusion criteria #4, #6, #9.

### Safety Analysis Set

Safety analysis set will include all subjects who were administered at least one dose of study treatment or those who were randomized in the control arm but only received BSC. Subjects will be classified according to the treatment assigned at randomization unless the incorrect treatment was received throughout the dosing period in which case subjects will be classified according to the first study treatment received.

The rationale that subjects only received BSC are included in the safety analysis set is because they were randomized and they received the intended medication.

### HRQoL Analysis Set

The HRQoL analysis set is a subset of the FAS and includes all FAS subjects who meet all of the following criteria:

- Have one baseline HRQoL assessment
- Have at least one post-baseline HRQoL questionnaire completed

Baseline HRQoL assessment is defined in Section 11, "Definition of baseline". Post-baseline assessment is defined as HRQoL questionnaires completed after the first dose of study treatment for subjects who received avelumab or physician's choice chemotherapy treatment, and questionnaires completed after date of randomization for subjects on BSC only.

CCI [redacted]

[redacted]

[redacted]

[redacted]

Table 1 summarizes how each of the analysis sets will be used.

**Table 1 Statistical Analyses by Analysis Set**

| Analysis | Analysis Set | | | | |
|---|---|---|---|---|---|
| | FAS | PP Analysis Set | Safety Analysis Set | HRQoL Analysis Set | CCI [redacted] |
| Demographic and Baseline Characteristics | ✓ | | ✓ | | |
| Prior and Concomitant Therapies/Procedures | ✓ | | | | |
| Compliance and Exposure | | | ✓ | | |

| Analysis | Analysis Set | | | | |
|---|---|---|---|---|---|
| | FAS | PP Analysis Set | Safety Analysis Set | HRQoL Analysis Set | CCI |
| Efficacy: Primary Endpoint (OS) | ✓ | ✓ | | | |
| Efficacy: Key Secondary Endpoints (PFS, BOR) | ✓ | ✓ | | | |
| Efficacy: Other Secondary Endpoints (QoL) | | | | ✓ | |
| CCI | | | | | |
| Safety Endpoints | | | ✓ | | |
| PK | | | | | |

### Subgroup Analysis Set

Subgroup analyses will be performed on primary and the key secondary efficacy endpoints (i.e. PFS and BOR) based on the subgroups as defined below.

To include baseline variables into Cox's proportional hazards model for OS and PFS multivariate analysis, and logistic regression model for BOR multivariate analysis, the following parameterization is to be used. The final parameterization will be updated and fixed at the Data Review Meeting at the latest and documented in an amendment to this SAP if different from the following definition.

In the case of a low number of subjects within a category ( < 19 subjects, which is about 5% of the randomized subjects), the categories may be pooled when meaningful. The subgroup analysis will not be performed on any subgroup category "Missing".

- PD-L1 assay status at cut-off value as follows:
  - < 1% (Reference) vs. >=1%
  - < 5% (Reference) vs. >=5%
  - < 25% (Reference) vs. >=25%
- Age Group 1
  - Age < 65 years (Reference)
  - Age ≥ 65 years
- Age Group 2
  - Age < 75 years (Reference)
  - Age ≥ 75 years
- Gender

- Male (Reference)
- Female

- Race
  - Caucasian / White (Reference)
  - Asian
  - Black/ African American
  - Other
- Pooled Geographical Region:
  - Asia
  - Non-Asia (Reference)
- Country
  - Japan
- ECOG PS at baseline
  - ECOG PS 0 (Reference)
  - ECOG PS 1
- Site of primary tumor
  - Stomach
  - Gastro-esophageal Junction (GEJ) (Reference)
- Type of prior anti-cancer treatment
  - Platinum-based regimen (reference)
  - Ramucirumab-based regimen
  - HER-2 directed regimen
  - Other regimens

Detailed grouping of type of prior anti-cancer treatments will be determined before database lock after a comprehensive medical review on all such therapy in the database.

# 11 General Specifications for Statistical Analyses

Unless otherwise stated all table summaries will be presented by treatment arm. All statistical analyses will be performed using SAS® Version 9.2 or higher.

**Data handling after cut-off date:**

Data after cut-off does not undergo the cleaning process.

Data obtained after a cut-off 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 trial 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 randomized at each center.

**Significance level:**

The overall significance level is 2.5% one-sided. The confirmatory statistical tests for the primary and key secondary efficacy endpoint analyses are described in Sections 16.1 and 16.2 along with procedures for controlling the overall type I error rate. The statistical tests performed on the primary and key secondary efficacy endpoints in comparing treatment arms will be one-sided with a significance level of 2.5%. The statistical tests to compare treatment arms on other secondary, exploratory, and safety analyses will be two-sided with a significance level of 5%.

**Presentation of continuous and qualitative variables:**

Continuous variables will be summarized using descriptive statistics i.e., number of non-missing values, mean, median, standard deviation (StDev), minimum, maximum, first quartile (Q1), and third quartile (Q3). For reporting conventions, mean, median, Q1, and Q2 will generally be displayed with one more decimal place than the raw data and StDev will 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.

Qualitative variables will be summarized by counts and percentages based on the number of subjects in the analysis set of interest, unless otherwise specified. In general, percentages will be reported to 1 decimal place unless greater precision is deemed appropriate.

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:**

For the safety analysis, baseline is defined as the last measurement before the first dose of study treatment. Baseline for heart rate and QTc assessments will be derived from the visit where both heart rate and QT are not missing.

For the efficacy analysis, baseline is defined as the last measurement prior to randomization. If such a value is missing, the last measurement prior to the first study drug administration will be used as the baseline measurement for the efficacy analysis except for analyses of tumor assessment data where the baseline assessment would be considered missing.

If an assessment is planned to be performed prior to the randomization/first dose of study treatment in the protocol and the assessment is performed on the same day as the randomization/first dose of study treatment, it will be assumed that it was performed prior to randomization/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- randomization/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.

**Definition of duration:**

Duration will be calculated by the difference of start and stop date + 1, if not otherwise specified. For example, survival time (days) = date of death – date of randomization + 1.

The time since an event (e.g. time since first diagnosis) will be calculated as reference date minus date of event.

**Definition of study day/treatment day:**

Study day / Treatment day are defined relative to the date of randomization / start of treatment. Study day 1 corresponds to the day of randomization and Study day -1 corresponds to the day before randomization (i.e., no Study day 0 is defined). Treatment day will be defined similarly with Study day 1 corresponding to the date of first administration of study treatment.

**Definition of on-treatment period:**

For subjects who received avelumab or physician's choice chemotherapy treatment, the 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).

For subjects in the control arm who only received BSC, the on-treatment period is defined as the time from the date of randomization through minimum (30 days + end date of BSC treatment period, start day of new anti-cancer drug therapy – 1 day), while end date of BSC treatment period is defined as the date collected on the "End of Assessments Visit" eCRF page.

The date of new anti-cancer drug therapy 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)$^{0.5}$

**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, StDev, 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 data and safety endpoints (laboratory, ECG and vital signs), only those scheduled visits/time points per-protocol that have at least 5 subjects in both treatment arms will be included in the summary tables and figures. The exception is Discontinuation and End-of-Treatment visit which will be included in the summary statistics despite the number of subjects who completed such visit.

**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 (i.e. when they cannot be calculated) will be presented as “nd” (i.e. not done). For example, if n=1, the measure of variability cannot be computed and will 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 randomized treatment.
- If only the day part of the AE onset date is missing, but the month and year are equal to the start of randomized treatment, then the AE onset date will be replaced by the start of randomized treatment. For example, if the AE onset date is --/JAN/2015, and randomized treatment start date is 15/JAN/2015, then the imputed AE onset date will be 15/JAN2015.
- If both the day and month of the AE onset date are missing but the onset year is equal to the start of randomized treatment, then the onset date will be replaced by the start of randomized treatment. For example, if AE onset date is --/---/2014, and randomized 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/JAN2015.

- 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 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.

**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 the subsequent anti-cancer therapy is before the last day of the month. 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**

- If the study medication start date is missing, it is assumed that the first dose of study medication is given at the randomization date. The randomization date will replace incomplete dates of the first dose of study medications.
- 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:

- 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
- Randomization date
- 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 missing it will be imputed as day after date of last contact from the CRF survival page
- 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**

All investigation dates (e.g. X-ray, CT scan) 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 the same day, the choice of date of assessment could impact the date of progression and/or date of response. If there are multiple scan dates associated with an evaluation, the earliest of the scan dates associated with the evaluation will be used as the date of assessment.

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

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

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

# 12 Trial Subjects

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

## 12.1 Disposition of Subjects and Discontinuations

The following will be summarized overall and by treatment arm (as randomized) for all randomized subjects:

- Number of subjects screened (overall only)
- Number of subjects who failed screening overall and grouped by the main reason
- Number and percentage of subjects randomized in the following analysis sets
    - FAS
    - PP analysis set
    - Safety analysis set
    - HRQoL analysis set
    - [redacted]
- Number and percentage of subjects randomized but not treated by study treatment by reason
    - Subject with only BSC after randomization
    - Other reasons collected on “Study Termination” eCRF page
- Number of randomized subjects still on treatment
- Number of randomized subjects off-treatment overall and by the main reason of discontinuation
- Number of randomized subjects who discontinued treatment but are still in follow-up
- Number of subjects who discontinued the study and primary reasons for study discontinuation
- Number and percentage of subjects who re-initiated avelumab treatment

- Number of subjects discontinued the treatment after re-initiation

Number and percentage of subjects who received new anti-cancer therapy after study treatment discontinuation. In addition, the following will be summarized:

- Number and percentage of randomized subjects overall, by region (Europe, European Economic Area (EEA) (required by EudraCT), North America, Latin America, Western Europe, Eastern Europe, Asia, Australasia, Africa and Middle East, by country within region
- Number and percentage of randomized subjects by center.
- Cross tabulation: subjects randomized (avelumab + BSC or physician's choice + BSC) versus subjects treated (avelumab + BSC or physician's choice + BSC),
  - Reasons if randomized ≠ treated

## 12.2 Protocol Deviations

*Analysis sets: FAS*

All important protocol deviations (IPDs) 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.

IPDs 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.

In addition, the IPDs that lead to the exclusion of a subject from the PP analysis set as specified in Section 10 are clinically important protocol deviations (CIPD), which are a subset of important protocol deviations will be reported.

All IPDs will be documented in CDISC datasets whether identified through site monitoring, medical review or programming. Other protocol deviations will not be included in the CDISC datasets but will be maintained in separate clinical trial master file. The complete list of IPDs and CIPDs are maintained by the medical team and will be finalized prior to database lock. IPDs to be identified by both medical review process and programming as well as all CIPDs are specified in Appendix I.

All IPDs and CIPDs identified by medical review process and/or programming will be presented in the summary tables by treatment arm and in data listings.

# 13 Demographics and Other Baseline Characteristics

## 13.1 Demographics

*Analysis sets: FAS, Safety Analysis Set*

Demographic characteristics will be summarized by treatment arm and overall using the following information from the Screening/Baseline Visit CRF pages.

- Demographic characteristics
  - Gender: Male, Female
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Unknown
  - Ethnic origin: Hispanic/Latino (Yes/No), Japanese (Yes/No)
  - Age (years): summary statistics
  - Age categories :
    - < 65 years,
    - ≥ 65 years
      - 65-<75
      - 75-<85
      - ≥ 85
  - Pooled Geographical Region:
    - North America
    - Europe
    - Asia
    - Rest of the World (Australia and/or Latin America will be included as additional pooled geographical regions if including > 10% of the overall randomized population)
  - Geographic Region:
    - North America
    - Latin America
    - Western Europe
    - Eastern Europe
    - Australasia
    - Asia
    - Africa
    - Middle East
  - EEA: Yes or No
  - Eastern Cooperative Oncology Group (ECOG) Performance Status (PS): 0 or 1
  - Physical measurements
    - Height (cm)
    - Weight (kg)

- Body Mass Index (BMI) ($kg/m^2$)

– Body Surface Area (BSA) ($m^2$)

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, treatment group, age, sex, race, ethnicity, height (cm), weight (kg), BMI ($kg/m^2$), BSA ($m^2$), and ECOG performance status.

## 13.2 Medical History

*Analysis sets: FAS, Safety Analysis Set*

Medical history reported at the time of the Screening/Baseline procedures will be coded using the most current available version of Medical Dictionary for Regulatory Activities (MedDRA) and will be summarized from the "Medical History" eCRF page. Medical history will be summarized as the numbers and percentages of 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 by treatment arm based on the safety analysis set and FAS.

## 13.3 Other Baseline Characteristics

*Analysis sets: FAS, Safety Analysis Set*

### 13.3.1 Disease Characteristics

Information on disease characteristics collected on the "Disease History" eCRF page will be summarized by treatment arm and overall. Summary statistics will be presented for the following:

- Primary site of tumor
- Sub-site of tumor
- Histopathologic / cytologic type of tumor
- Time since first occurrence of metastatic disease (months), defined as (date of randomization - date of first occurrence of metastatic disease)/30.4375
- Time since documented, locally advanced or inoperable disease diagnosis (months), defined as (date of randomization – date of documented locally advanced, inoperable or metastatic disease)/30.4375
- TNM classification at initial diagnosis
- TNM classification at study entry

Baseline characteristics with respect to ECOG PS, vital signs, physical examinations, ECGs, 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 (i.e., anti-cancer medications other than study treatment taken at any time pre-treatment) categories will be tabulated by treatment arm and overall based on the safety analysis set and FAS:

- 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 anti-cancer drug therapy will be summarized as follows by treatment arm and overall:

- Number of subjects with at least one prior anti-cancer drug therapy
- Number of any prior anti-cancer therapy regimens: missing / 1 / 2 / 3 / ≥4
- Number of prior anti-cancer therapy regimens for metastatic or locally advanced disease: missing / 1 / 2 / 3 / ≥4
- Type of prior anti-cancer therapy: Cytotoxic therapy / Endocrine therapy / Monoclonal antibody therapy / Small molecules / Immunotherapy / Other
- Intent of Therapy: Neo-Adjuvant / Adjuvant / Metastatic or Locally advanced
- Type of prior anti-cancer treatment: Platinum-based regimen / Ramucirumab-based regimen / HER-2 directed regimen / Other regimens. Detailed grouping of type of prior anti-cancer regimens will be determined before database lock after a comprehensive medical review on all such therapy in the database.
- Best response: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive Disease (PD) / Not assessable / Unknown / Not applicable. Best response will be summarized by each regimen.

The prior anti-cancer drugs will also be summarized based on the number and percentage of subjects by the drug class level and preferred term. Within each regimen, a subject will be counted only once within a given drug class and within a given preferred term, even if he/she received the same medication at different times in the same regimen. The summary will be sorted on decreasing frequency of drug class and preferred term in the 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. These listings will include the subject identification number and all relevant collected data on the corresponding eCRF pages.

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

### 13.3.3 PD-L1 Expression Status at Baseline

Three scoring methods to evaluate PD-L1 expression status will be used as follows:

- Conventional tumor cell (TC) (cell number-based percentage) scoring method
- Non-tumor cell elements area-based scoring method
- Whole tumor area-based scoring method

Additional details of these scoring algorithms are specified in the Pathology Report Form (PRF).

PD-L1 expression status (positive vs. negative) at baseline with different cutoff values will be summarized in total, by treatment arm, by scoring methods, as well as by treatment arm within region (Asia vs. non-Asia). PD-L1 assay status of total % of PD-L1 positive tumor cells will be summarized as below:

- $< 1\%$
- $\geq 1\%$ to $< 5\%$ (cut-off is 1%)
- $\geq 5\%$ to $< 25\%$, (cut-off is 5%)
- $\geq 25\%$ (cut-off is 25%)

## 14 Previous or Concomitant Medications/Procedures

*Analysis sets: FAS*

### 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 treatment, which are started before first dose date of study treatment.

Prior and concomitant medications will be summarized from the "Concomitant Medications Details" eCRF page. Premedications 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.

Summaries of prior and concomitant medications will include the number and percentage of subjects by Anatomical Therapeutic Chemical (ATC) classification level 2 and preferred term within treatment arm and overall. A subject will be counted only once within a given ATC class and within a given preferred term, 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 the "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, which were undertaken any time during the on-treatment period, will be summarized by treatment arm and overall. Number of subjects with concurrent procedures will be tabulated overall and by reason for procedure as collected in the eCRF page "Concomitant Procedures Details".

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

Premedications for study drug will also be presented in the listing.

## 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 and the earliest date of start of any new anti-cancer therapy will be used for censoring for efficacy analyses.

Number and percentage of subjects with any anti-cancer treatment after discontinuation will be tabulated overall and by type of therapy within treatment arm based on data collected from the "Anti-Cancer Treatment after Discontinuation Details" eCRF page.

Summary statistics will be created for best response across all post study treatments based on the data collected from "Anti-Cancer Treatment after Discontinuation Details" eCRF page. For subjects who received more than one anti-cancer drug therapy after treatment discontinuation, the best overall response among all anti-cancer drug therapies will be summarized.

Summary of subsequent anti-cancer treatment will include the number and percentage of subjects by ATC Classification level 2 and preferred term. The same approach as prior and concomitant medications will be applied in presenting the summary table of subsequent anti-cancer treatment.

# 15 Treatment Compliance and Exposure

*Analysis sets: Safety Analysis Set*

All dosing calculations and summaries will be based on the administration of study medications from the corresponding eCRF pages.

- Subjects randomized to the avelumab + BSC arm 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).
- Subjects randomized to the physician's choice arm (paclitaxel + BSC or irinotecan + BSC) will receive an IV infusion of one of the following therapies:
    - Paclitaxel – Days 1, 8, and 15 of a 4-week treatment cycle
    - Irinotecan – Days 1 and 15 of a 4-week treatment cycle

Analysis of exposure is based on the calculated actual dose level. The last available weight of the subject on or prior to the day of dosing will be used. Height at Screening visit will be used in deriving BSA.

- Avelumab: total dose administered/weight [mg/kg]
- Physician's choice (paclitaxel + BSC or irinotecan + BSC treatment: total dose administered/BSA [mg/m$^2$]

Any BSC therapy will not be included in the compliance and exposure calculation.

**Duration**

The duration of *avelumab* treatment (in weeks) during the study is defined as:

Treatment duration (weeks) = (date of last dose – date of first dose + 14)/7

The duration of *paclitaxel* and *irinotecan* treatment (in weeks) during the study is defined as:

Treatment duration (weeks) = {date of last dose – date of first dose + [28 – (date of last dose of last cycle – date of first dose of last cycle)]} /7

**Cumulative Dose**

The overall cumulative dose (mg/kg) of avelumab per subject across all cycles is the sum of the actual dose levels that the subject received (i.e. total dose administered (mg) / weight (kg)).

The overall cumulative dose (mg/m$^2$) of paclitaxel and irinotecan per subject across all cycles is the sum of the actual dose levels that the subject received (i.e. total dose administered (mg) / BSA (m$^2$)).

**Dose Intensity**

The dose intensity will be calculated for each subject across all cycles. The dose intensity of avelumab (mg/kg/cycle) is defined as:

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

The dose intensity of irinotecan (mg/m$^2$/cycle), and paclitaxel (mg/m$^2$/cycle) is defined as

$$\text{dose intensity} = \left(\frac{\text{Cumulative dose of irinotecan or paclitaxel}}{(\text{duration of therapy (in weeks)})/4}\right)$$

**Relative Dose Intensity**

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

Summaries of treatment exposure and compliance will include the following information:

- Duration of therapy (weeks)
- Total number of infusions received
- Cumulative dose of therapy (mg/kg or mg/m$^2$)
- Dose intensity (mg/kg/cycle) or (mg/m$^2$/cycle)
- Relative dose intensity of therapy categories: 80%-<90%, 90%-<110%, >=110%

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 an actual non-zero dose < 90% of the planned dose. The number of subjects with at least one dose reduction as well as a categorical breakdown of dose reductions (1 / 2 / 3 / >=4) will be summarized for the physician's choice (paclitaxel or irinotecan) + BSC arm. There are no dose reductions allowed for the avelumab treatment arm in the protocol.

**Dose Delays**

The number of subjects with delayed infusions will be summarized. Delays are defined as infusions given ≥3 days from the planned administration date and will be derived based on study drug administration 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 a subject receives avelumab on Day 1, then the next avelumab administration date should be on Day 15; however, if the subject receives avelumab on Day 16 or 17, this is not considered a delay. In the event that a subject has more than one delay, the maximum delay will be used.

Number and percentage of subjects with delayed study drug administration and maximum length of delay, i.e. the worst case of delay if subjects have multiple dose delays will be summarized by treatment group.

**Infusion Rate Reductions**

Infusion rate reductions will be calculated for avelumab + BSC arm only. It will be derived based on the infusion rate recorded by visit on the eCRF. Number of subjects with at least one infusion rate reduction as well as a breakdown of the number of infusion rate reductions (1 / 2 / ≥3) will be summarized.

A listing of study drug administration will be created with the information collected on the applicable treatment administration details eCRF pages.

Subjects with dose reductions, dose delays or infusion rate reductions and the corresponding reasons will be summarized in the listing.

# 16 Endpoint Evaluation

The primary endpoint of the study is Overall Survival (OS). The key secondary efficacy endpoints include Progression-free Survival (PFS) and Best Overall Response (BOR) according to RECIST 1.1 criteria and as adjudicated by the IRC.

The type I error rate for the primary endpoint (OS) and two key secondary endpoints of PFS and BOR will be controlled at 2.5% (one-sided) significance level using a gatekeeping procedure. The primary hypothesis (overall survival) will serve as a gate keeper and be tested at an overall one-sided type I error of 2.5%. Only if the primary endpoint is statistically significant, the two key secondary endpoints of PFS and BOR will be tested, and the Hochberg step-up procedure will be used to control the type I error rate at 2.5% (one-sided) for PFS and BOR.

No formal adjustment for multiplicity will be undertaken for sensitivity or subgroup analyses involving the primary endpoint, analyses of non-key secondary endpoints, or analyses involving exploratory endpoints.

## 16.1 Primary Endpoint Analyses

### 16.1.1 Primary Efficacy Analysis of Overall Survival

*Analysis sets: FAS, PP Analysis Set,* CCI [redacted]

The primary efficacy analysis of Overall Survival (OS) will be performed on the FAS. All data required for the calculation of time to event will be taken from the eCRF.

OS is defined as the time from randomization to the date of death, regardless of the actual cause of the subject's 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 as specified in Section 11 as of the data cut-off date for the analysis.

OS (in months) = (Date of death – Date of randomization +1) / 30.4375

The primary efficacy analysis will compare the OS time between the treatment arms, and will be performed using a one-sided stratified log-rank test at the alpha level of 0.025. The stratification factor is region (Asia vs. non-Asia).

The following null hypothesis will be tested:

$$H_0: \lambda_A(t) = \theta\lambda_B(t),\ \theta \geq 1, \text{ versus } H_1: \lambda_A(t) = \theta\lambda_B(t),\ \theta < 1,$$

where $\lambda(t)$ represents the hazard at time t and $\theta$ the unknown constant of proportionality of hazards in treatment groups A (avelumab + BSC) and B (physician's choice + BSC).

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

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (i.e., ties=discrete option in SAS PHREG procedure).

Kaplan-Meier (i.e. product-limit) estimates of median OS time will be presented by treatment arm together with two-sided 95% confidence intervals (CIs) calculated according to Brookmeyer and Crowley (1). Finally, the number of subjects at risk vs. failed along with Kaplan-Meier estimates of survival probability at 3, 6, 9 and 12 months will be estimated with corresponding two-sided 95% CIs derived using the log-log transformation according to Kalbfleisch and Prentice (2) (i.e., conftype=loglog default option in SAS LIFETEST procedure). The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of subjects with an event (death) and censoring reasons will be presented by treatment arm. Censoring reasons are as follows:

- Alive
- Withdrawal of consent
- Lost to follow-up

Lost to follow-up will include the following subjects:

- Lost to follow-up status is collected on the eCRF page prior to the analysis cut-off;
- Subjects with the last contact date > 13 weeks prior to the analysis cut-off date (duration of 13 weeks is based on the assessment schedule of every 12 week for survival follow-up interval + 1 week window).

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

### 16.1.2 Sensitivity Analyses of Primary Endpoint

*Analysis sets: FAS*

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

- The primary analyses described above will be repeated based on the PP analysis set if the PP analysis set includes less than 90% of subjects in the FAS.
- An unstratified analysis will be performed to compare the OS time between the treatment arms.
- If the actual number of events is 10% more than the planned number at the primary analysis, i.e. if there are > 282 events based on the prospectively determined cut-off date at the primary analysis, the primary analysis will be repeated by using the first cut-off date at which exactly 256 events that were observed in the study for the primary analysis.
- The primary analysis will be repeated by censoring those subjects in FAS who received subsequent anti-cancer therapy after discontinuing the study treatment with the date of the first dose of subsequent anti-cancer therapy minus 1 day. The final list of subsequent anti-cancer therapy will be provided upon medical review of all subsequent anti-cancer therapies.

**Methods for Evaluating the Validity of Model Assumptions**

In order to assess the usefulness of the treatment difference estimate from the Cox proportional hazards model, the validity of the proportional hazards assumption will be assessed visually by plotting log(-log(survival)) versus log(time) by treatment.

Schoenfeld residuals including a LOESS curve will be plotted to graphically investigate violations of the proportional hazards assumption. Schoenfeld residuals will be computed in SAS using the PHREG procedure via the OUTPUT statement (keyword=RESSCH). With proportional hazards the LOESS curve should be parallel to the x-axis.

If these show large departures from proportional hazards then OS will also be analyzed based on restricted mean survival time (RMST) differences (Zhang, 2013) (6).

**Restricted Mean Survival Time (RMST)**

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

- $\tau_1$ = minimum of (largest observed survival time for avelumab + BSC arm, largest observed survival time for physician's choice + BSC arm).
- $\tau_2$ = minimum of (largest survival event time for avelumab + BSC arm, largest survival event time for physician's choice + BSC arm).

In this section, 'survival' is meant to denote OS.

The treatment effect between the two treatment arms will be assessed based on the difference in RMST. The associated 95% CI for the difference in means and 1-sided p-value will be generated. RMST as a function of τ and the associated treatment effect between the two treatment arms will be plotted against time τ

**Exploratory Analyses to Investigate the Impact of Potential Prognostic Factors**

Multivariable Cox regression analysis will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact.

To include baseline variables into Cox's proportional hazards model, the subgroup levels defined in Section 10 "Subgroup Analysis Sets" are to be used. For variables with more than two categories, an indicator variable will be defined for each category except for the first category, which defines the reference always. The final parameterization will be updated and fixed at the Data Review Meeting at the latest and documented in an amendment to this SAP if different from the following definition.

A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The Cox's Proportional Hazard model is defined as: h(t) = h(0;t) exp(Xb), where h(0;t) defines the baseline hazard function and X defines the vector of explanatory variables and b the unknown vector of regression parameters. Variables are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the

subsequent backward elimination. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald test). This analysis will be performed using the stepwise selection method in SAS (Proc PHREG). Once this procedure stops, the factor 'treatment group' will be added to the last selected model in order to evaluate the effect of treatment on OS time when adjusted for the selected explanatory variables. The hazard ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% confidence intervals. No interactions will be considered. Post-baseline factors will not be considered for the model.

### 16.1.3 Subgroup Analysis of Primary Endpoint

*Analysis sets: FAS*

Subgroup analyses will be performed on the primary endpoint using all subgroup levels defined in Section 10 “Subgroup Analysis Sets”. All subgroup analyses are exploratory and will be performed as unstratified. No adjustment for multiplicity will be performed. In the case of a low number of subjects within a category (< 5% of the randomized population), the categories may be pooled.

The OS time between the two treatment arms will be compared using a two-sided unstratified log rank test per subgroup level; the unstratified HR and its corresponding 95% CI will be computed per subgroup level.

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

- Model 1: treatment + subgroup
- Model 2: treatment + subgroup + treatment * subgroup-variable

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

The HR and its corresponding 95% CI of all subgroups will also be presented in a forest plot.

### 16.1.4 Time of Follow-up for OS

*Analysis sets: FAS*

A Kaplan-Meier plot for OS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the OS censoring and event indicators.

Kaplan-Meier estimates will be presented by treatment group together with the median time of follow-up for OS. In particular, the follow-up rate at 3, 6, 9 and 12 months will be estimated with corresponding two-sided 95% CIs.

**Table 2 Censoring Rules for Duration of Overall Survival Follow-up**

| | Date of event / censoring | Censoring |
|---|---|---|
| Subjects alive or lost to follow-up | Time from randomization to last date known to be alive | No |
| Subjects who died | Time from randomization to date of death | Yes |

## 16.2 Secondary Endpoint Analyses

Based on the gatekeeping testing strategy described at the beginning of Section 16, if the one-sided p-value from the stratified log-rank test of the primary endpoint OS is less than 0.025, the key secondary endpoints (i.e., PFS and BOR) will be tested per the Hochberg step-up procedure used to control the type I error rate at 2.5% (one-sided). Per this procedure according to Westfall, et al. (3), the one-sided p-value from the stratified log-rank test of PFS and the Cochran-Mantel-Haenszel test (CMH) of ORR will be calculated. If the least significant p-value (the larger p-value is less than 0.025 then both null hypotheses corresponding to the key secondary endpoints of PFS and BOR will be rejected. Otherwise, if the least significant p-value (the larger p-value) of the two secondary endpoints is greater than 0.025, then the null hypothesis of this secondary endpoint will not be rejected at the one-sided alpha level of 0.025, and the other secondary endpoint will be tested at the one-sided p-value of 0.0125.

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 IRC assessment results.

### 16.2.1 Progression-Free Survival (PFS)

*Analysis sets: FAS, PP Analysis Set,* CCI [redacted]

PFS is defined as the time from date of randomization until date of the first documentation of PD or death by any cause (whichever occurs first). The tumor response will be determined according to RECIST 1.1 and adjudicated by an IRC. Details on determination of the first disease progression date are specified in the IRC charter.

PFS time (in months) = (Date of PD or death - Date of randomization + 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 new anti-cancer treatment prior to an event, or for subjects with an event after two or more consecutive 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 randomization unless death occurred on or before the time of the second planned tumor assessment in which case the death will be considered an event.

If a tumor assessment was performed on the same day as start of 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 the tumor assessment will be included in the derivation of the PFS time.

Last adequate tumor assessment is defined as the last tumor assessment result that is not "NE" or "NA".

The censoring and event date options to be considered for the PFS and Duration of Response (DR) analysis are presented in Table 3.

**Table 3 Outcome and Event Dates for PFS and DR Analyses**

| Scenario | Date of event / censoring | Outcome |
|---|---|---|
| No baseline assessment | Date of randomization | Censored [a] |
| PD or death<br>- After at most one missing or inadequate post-baseline tumor assessment, OR<br>≤ 12 weeks after date of randomization | Date of progression or death | Event |
| PD or death after 2 or more missing or inadequate post-baseline tumor assessments | Date of last adequate tumor assessment[b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments | Censored |
| No progression and no death | Date of last adequate tumor assessment[b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments | Censored |
| Treatment discontinuation due to "Disease progression" without documented progression | Not applicable | Information collected on treatment discontinuation page is ignored since outcome 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.

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

The analysis of PFS will be analogous to that for OS time as described in section 16.1.1.

Frequency (number and percentage) of subjects with each event type (PD or death) and censoring reasons will be presented by treatment arm. Censoring reasons are as follows:

- Ongoing in the study without an event
- No baseline assessment
- No adequate post-baseline assessment
- Start of new anti-cancer therapy prior to PD

- Event after 2 or more missing or inadequate post-baseline assessments
- Withdrawal of consent
- Lost to follow-up

Lost to follow-up will include the following subjects:

- Lost to follow-up status is collected on the eCRF page prior to the analysis cut-off;
- Subjects with the last contact date > 13 weeks prior to the analysis cut-off date (duration of 13 weeks is based on the assessment schedule of every 12 week for survival follow-up interval + 1 week window).

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

**Sensitivity Analysis**

The following sensitivity analyses of PFS will also be performed.

1. The primary analyses will be repeated on the PP analysis set if the PP analysis set includes less than 90% of subjects in the FAS;
2. An unstratified analysis will be performed to compare PFS time between the treatment arms;
3. PFS based on IRC assessment and counting all PD and deaths as events without considering the censoring rule for PD or death as described in Table 3;
4. PFS based on IRC assessment and initiation of subsequent anticancer therapies will not be used as a censoring reason for PFS.
5. PFS based on investigator assessment will be analyzed using the same censoring rules described in Table 3. For this sensitivity analysis, the date of tumor assessment will be derived from the eCRF. The earliest tumor assessment date corresponding to the respective visit that is collected on this page will be used for date of disease progression;
6. A summary of concordance of disease progression status between IRC and investigator assessment will be provided including status of "No Event", "Progressive disease" and "Death as event".

### 16.2.2 Best Overall Response (BOR)

*Analysis sets: FAS, PP Analysis Set,*[redacted]

The confirmed BOR is defined as the best confirmed response obtained among all tumor assessment visits after the date of randomization until documented disease progression, taking requirements for confirmation into account as detailed below. The tumor response at each assessment visit will be determined according to RECIST 1.1 and as adjudicated by an IRC. Details of determination of tumor response date are provided in the Imaging Review Charter.

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 start of 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 the tumor assessment will be included in the assessment of BOR.

The following requirements must be met for confirmation of response:

- PR or CR needs to be confirmed at a subsequent tumor assessment, preferably at the regularly scheduled 6-week assessment interval (after 12 months of treatment, 12-week assessment interval will be used), but no sooner than 5 weeks after the initial documentation of CR or PR.
- PR or CR can be confirmed at an assessment later than the next assessment after the initial documentation of PR or CR, respectively.
- The minimum duration for a BOR of stable disease (SD) is defined as at least 6 weeks after randomization.
- PD = progression ≤ 12 weeks after date of randomization (and not qualifying for CR, PR or SD), i.e. tumor assessment of PD that is >12 weeks after date of randomization and there is no tumor assessment in between will have a BOR of NE.

Table 4 summarizes the derivation rules described by Eisenhauer, et al. for the BOR when confirmation from subsequent assessment is needed (4). For subjects who have non-target lesions only at baseline, the time-point tumor assessment of "Non-CR/non-PD" will be evaluated with the same criteria as SD (minimum criteria for SD duration) in deriving the overall BOR.

**Table 4 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 |

| Overall response 1st time point | Overall response subsequent time point | Best overall response (BOR) |
|---|---|---|

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 patient 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 objective response rate (ORR) is defined as the proportion of subjects having reached a confirmed BOR of CR or PR according to RECIST v1.1 and as adjudicated by the IRC. Subjects with BOR of non-CR/non- PD are not considered as having achieved objective response. Therefore, these subjects will only be counted in the denominator of the rate, but not in the numerator.

The confirmed ORR by treatment group 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).

The association of treatment effect and objective response (OR) will be tested by the General Association Statistic of the Cochran-Mantel-Haenszel test (CMH) according to Cochran (5) with the stratification factor of region (Asia vs. non-Asia). The null hypothesis of no association in any of the strata is tested against the alternative, which specifies that there is an association between treatment effect and ORR in at least one stratum. The one-sided CMH test will be performed with the alpha level of 0.025.

In case the assumptions of the CMH test are violated due to a small number of subjects in some strata (i.e., number of subjects ≤ 10 in at least one of the stratum), a Fisher's exact test will be used.

The stratified (i.e. adjusted) odds ratio in terms of ORR will also be estimated with a Mantel-Haenszel estimator along with its 95% CI to compare the treatment effect. The odds ratio is defined as the odds of OR with avelumab + BSC divided by the odds of OR with physician's choice + BSC. The Breslow-Day test will be used to test the null hypothesis that odds ratios in all regions are equal against the alternative hypothesis that the odds ratio in at least one region is different.

In case the null hypothesis of homogeneity of odds ratios across regions is not rejected at the two-sided alpha level of 5%, the common odds ratio will be determined as the Mantel-Haenszel estimate (by the SAS FREQ procedure using the CMH option); if the null hypothesis of homogeneity of odds ratios across all regions is rejected, the odds ratio per region will be calculated with the corresponding exact CI.

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. Subjects with BOR of NE will be summarized by reason for having NE status. The following reasons will be used:

- No baseline assessment (independent review committee identifies neither any target nor any non-target lesions)
- No post-baseline assessments due to death within 6 weeks after randomization
- No post-baseline assessments due to other reasons
- All post-baseline assessments have overall response NE
- New anticancer therapy started before first post-baseline assessment
- SD of insufficient duration (<6 weeks after date of randomization without further evaluable tumor assessment)
- PD too late (i.e. tumor assessment of PD was >12 weeks after date of randomization and there was no 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'.

**Sensitivity Analysis**

The following sensitivity analyses of ORR will be performed:

- The analysis of ORR will be repeated on the PP analysis set if this analysis set includes less than 90% of subjects in the FAS;
- ORR according to investigator assessment will be analyzed in the same manner as ORR according to IRC assessment. The confirmed BOR according to Investigator assessment will be derived in the same way as the confirmed BOR according to the IRC.
- A summary of concordance of BOR between IRC and investigator assessment will also be provided. The confirmed BOR according to investigator assessment will be derived in the same way as the confirmed BOR according to the IRC.
- A multivariable logistic regression analysis will be performed to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. The subgroup variables defined in Section 10 "Subgroup Analysis Sets" will be included in the model. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald test). Once this procedure stops, the factor 'treatment group' will be added to the last selected model in order to evaluate the effect of treatment on BOR when adjusted for the selected explanatory variables. The odd ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% confidence intervals. No interactions will be considered. Post-baseline factors will not be considered for the model.

## 16.2.3 Time of Follow-up for PFS

*Analysis sets: FAS*

A Kaplan-Meier plot for PFS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the PFS censoring and event indicators.

**Table 5 Reverse Censoring Rules for Duration of PFS Follow-up**

| | Date of event / censoring | Censoring |
|---|---|---|
| Subjects alive or lost to follow-up | Time from randomization to last date known to be alive | No |
| Subjects who died or had PD | Time from randomization to date of first PD or death | Yes |

Kaplan-Meier estimates will be presented by treatment group together with the median time of follow-up for PFS. In particular, the follow-up rate at 3, 6, 9 and 12 months will be estimated with corresponding two-sided 95% CIs.

## 16.2.4 Subgroup Analysis of Secondary Efficacy Endpoints

*Analysis sets: FAS*

Analysis on subgroups as defined in Section 10 will be performed for the secondary efficacy endpoints, PFS and BOR.

All the subgroup analyses will be exploratory; no adjustment for multiplicity will be performed. In the case of a low number of subjects within a category (< 19 subjects, which is about 5% of the randomized population), the categories may be pooled when meaningful.

All the subgroup analyses will be performed as unstratified. The unadjusted 95% CI will be calculated for all subgroup analyses.

The subgroup analysis of PFS will be analogous to that of subgroup analysis of OS time as described in section 16.1.3.

For the subgroup analysis of BOR, the association of treatment effect and BOR will be tested using the two-sided CMH test per subgroup level. The ORR along with the two-sided exact Clopper-Pearson 95% CIs will be calculated for each subgroup.

In addition, to assess the heterogeneity of treatment effect across the subgroup levels for the secondary endpoint of BOR, a logistic regression model will be fitted with BOR as the dependent variable (=1 for subjects with a confirmed BOR of PR or CR; =0 otherwise); subgroup, treatment, and with and without the treatment-by-subgroup interaction as explanatory variables. A p-value for the interaction term (Wald Chi-Square test) will be provided together with the odds ratio and corresponding 95% CI of the interaction model parameter.

## 16.2.5 Health-Related Quality of Life

*Analysis sets: HRQoL Analysis Set*

Health-related quality of life (HRQoL) will be assessed by the EuroQOL 5-dimensions questionnaire (EQ-5D), the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 supplemented by the gastric cancer module QLQ-STO22 questionnaire. The analyses specified below will be performed for HRQoL analysis set. Further exploratory and sensitivity analyses on HRQoL assessment will be detailed in an addendum to this analysis plan.
The EQ-5D questionnaire is a generic measure of health status that provides a descriptive profile and a simple index value. It includes 5 items assessing mobility, self-care, usual activities, pain / discomfort, anxiety / depression on a 5-level response scale (EQ-5D-5L). It also includes a visual analogue scale (VAS) ranging from 0 to 100 for self-rated health status. A higher score indicates better health status. Utilities can be derived from the EQ-5D, ranging from 0.0 (worst health state) to 1.0 (best health state).

The EQ-5D-5L scoring system will also be converted into a single index value. The index value is country specific and is a major feature of the EQ-5D instrument, facilitating the calculation of quality-adjusted life years (QALYs) that are used to inform economic evaluations of health care interventions. The Japan country specific value set will be used in deriving the index value for Asian countries which ranges from -0.111 (worst health state) to 1.000 (best health state); the UK country specific value set will be used in deriving the index value for non-Asian countries which ranges from -0.594 (worst health state) to 1.000 (best health state).

The QLQ-C30 self-assessment questionnaire incorporates five functional scales (Physical, Role, Cognitive, Emotional, and Social), a global health status / QoL scale, and a number of single items assessing additional symptoms commonly reported by cancer subjects (e.g. dyspnea, loss of appetite, insomnia, constipation, and diarrhea), and financial impact of the disease.

All of the scales and single-item measures range in score from 0 to 100. A high scale score represents a higher response level. Therefore, a high score for a functional scale represents a high/healthy level of functioning, a high score for the global health status / QoL represents a high QoL, but a high score for a symptom scale/item represents a high level of symptomatology/problems. The scoring procedure for each of the scales is the same and consists of computing the raw score (RS) and then computing the actual scale score (S) by making a linear transformation to standardize the score to values from 0 to 100 as shown below.

$$\text{Raw Score} = RS = (I_1+I_2+\ldots+I_n) / n$$

For **Functional scales**:

$$\text{Score} = 100 \times [1 - (RS - 1) / \text{Range}]$$

For **Symptom scales / items** and **Global health status / QoL**:

$$\text{Score} = 100 \times [(RS - 1) / \text{Range})]$$

Where $I_1$, $I_2$, … $I_n$ are the individual items and Range is the difference between the maximum possible value of RS and the minimum possible value. The range of RS equals the range of the item values. Most items are scored 1 to 4, giving Range = 3. The exceptions are the items contributing to the Global Health Status / QoL, which are 7-point questions with Range = 6.

**Table 6 QLQ-C30 Global Health Status and Functional Scales**

| Scale | Items |
|---|---|
| **Global Health Status / QoL** | |
| Global Health Status / QoL | 29, 30 |
| **Functional Scales** | |
| Physical Functioning | 1, 2, 3, 4, 5 |
| Role Functioning | 6, 7 |
| Emotional Functioning | 21, 22, 23, 24 |
| Cognitive Functioning | 20, 25 |
| Social Functioning | 26, 27 |
| **Symptom Scales / Items** | |
| Fatigue | 10, 12, 18 |
| Nausea and vomiting | 14, 15 |
| Pain | 9, 19 |
| Dyspnoea | 8 |
| Insomnia | 11 |
| Appetite loss | 13 |
| Constipation | 16 |
| Diarrhoea | 17 |
| Financial difficulties | 28 |

The QLQ-STO22 self-assessment questionnaire is a gastric cancer module consisting of 22 items concerning disease and treatment-related symptoms and side effects, dysphagia, nutritional aspects and items about the emotional problems of gastric cancer. The module is composed of 5 multi-item symptom scales (Dysphagia, Pain, Reflux, Eating Restrictions, and Anxiety) and four single items (Dry Mouth, Tasting, Body Image, and Hair Loss. A high score represents more problems.

**Table 7 QLQ-STO22 Multi-Item Symptom Scales**

| Scale | Item(s) |
|---|---|
| Dysphagia | 31, 32, 33 |
| Pain | 34, 35, 36, 37 |
| Reflux | 38, 39, 40 |
| Eating Restrictions | 41, 42, 43, 46 |
| Anxiety | 47, 48, 50 |
| Dry Mouth | 44 |
| Tasting | 45 |
| Body Image | 49 |

| | |
|---|---|
| Hair Loss | 51, 52 |

The analysis will be performed on the following scores:

1. Each of the EQ-5D scores
2. QLQ-C30 Global health status/QoL scale, each of the five QLQ-C30 functional scales, and each of the nine QLQ-C30 symptoms/items
3. Each of the nine QLQ-STO22 gastric cancer module symptom scales

The following analyses will be performed for each of these score/scale systems:

- HRQoL questionnaires compliance and full completion rate:

  HRQoL compliance rate across all instruments will be summarized by considering the EQ-5D-5L, the QLQ-C30, and the QLQ-STO22 for each scheduled visit using the following definition:

$$\%Compliance = 100 \times \frac{\text{number of subjects with at least one HRQoL questionnaire available}}{\text{number of subjects for whom a HRQoL questionnaire is expected}}$$

$$\%Full\ completion\ rate = 100 \times \frac{\text{number of subjects with all items in HRQoL questionniare available}}{\text{number of subjects for whom HRQoL questionnaire is expected}}$$

  The compliance and full completion rate for each questionnaire will be displayed using a line plot with time on the x-axis and completion rate on the y-axis. Separate lines will be presented for each treatment arm.

- HRQoL questionnaires descriptive statistics

  Observed and change from baseline values will be summarized descriptively at planned visits during the treatment phase, the Discontinuation visit, End of Treatment visit and 30-day Safety Follow-up visit by treatment arm. The percent of subjects at the worst possible score (i.e., 5 on each of the EQ-5D scores, 0 on VAS, 0 on index score, 100 on symptom scales and 0 on QLQ-C30 functional scales and quality of life) and at the best possible score (i.e., 0 on each of the EQ-53 scores, 100 on VAS, 1 on index score, 0 on symptom scales and 100 on QLQ-C30 functional and quality of life) will be reported. In addition, the best (largest positive change), worst (largest negative change), and last observed post-baseline change from baseline values will be summarized.

  Change from baseline will also be displayed using a line plot with time point (scheduled visit) on the x-axis, mean change score on the y-axis, standard errors around each mean score, and separate lines for each treatment arm.

  If there are multiple complete assessments for any scheduled visit during the treatment phase, the assessment that is closest to the planned visit per protocol will be used in the analysis. If there are multiple complete assessments at Discontinuation visit and End of Treatment visit, the first assessment will be used in the analysis.

- Longitudinal analysis of change from baseline

  A mixed-effects model repeated measures (MMRM) analysis will evaluate longitudinal change from baseline on the QLQ-C30 physical function scale and the QLQ-STO22 scales. Covariates will include the baseline score for the domain score being evaluated and randomization stratification factors, and the questionnaire entry time point including both scheduled and unscheduled visits will be analyzed as a continuous variable. The overall significance of the difference between the MMRM trajectories for the treatment arms will be tested at a significance level of 0.05 (two-sided). An unstructured covariance structure will be used, if possible, though a simpler covariance structure will be used if the model does not converge.

- Subject level change from baseline

  For the HRQoL domain scores analyzed using MMRM, an empirical cumulative distribution function (eCDF) curves, separated by treatment arm, will display the proportion of subjects who experienced specific changes from baseline to end of treatment. For each eCDF, the range of change scores will be displayed on the x-axis, and the cumulative percentage of patients achieving that change score or better will be displayed on the y-axis. The eCDF may be interpreted by choosing a change score magnitude that may be considered clinically meaningful on the x-axis and comparing the rate of subjects who achieved at least that amount of change in the treatment arms on the y-axis.

**Missing Values and Imputation Rules**

If 50% or more of the items for a given scale are non-missing, then the Raw Score should be computed as the average of the non-missing items; if more than 50% of the items for a given scale are missing the scale score will be set to missing.

Missing data will be retained as observed, and no imputation of HRQoL scores will be conducted.

## 16.3 Other Endpoint Analyses

*Analysis sets: FAS*

### 16.3.1 Duration of Response (DR)

DR is defined, for subjects with an OR, as the time from first documentation of OR (CR or PR according to RECIST 1.1 and adjudicated by an IRC) to the date of first documentation of objective progression of disease or death due to any cause. For subjects with an OR but neither documented disease progression nor death as of the cut-off date for the analysis, duration of response will be censored at the date of the last adequate tumor assessment. The censoring rules for DR are the same as described for PFS in Section 16.1.2.

DR (months) = (date of PD or death – date of 1$^{st}$ OR + 1)/30.4375

DR will be displayed graphically and analyzed using Kaplan-Meier methodology analogous to that used for OS as described in Section 16.1.1. In particular, the DR rate at 3, 6 and 12 months will be estimated with corresponding two-sided 95% CIs.

In addition, DR according to investigator assessment will be analyzed in the same manner as DR according to IRC assessment.

If the number of subjects with OR is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided.

### 16.3.2 Time to Response

Time to response (TTR) is defined, for subjects with an OR according to RECIST v1.1, as the time from randomization to the first documentation of OR which is subsequently confirmed.

TTR (weeks) = (First date of CR or PR – date of randomization + 1)/7

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

### 16.3.3 Tumor Shrinkage

Tumor shrinkage will be summarized as the percent change from baseline in target lesions (sum of longest diameter for non-nodal lesions and short axis for nodal lesions) per time point. See Section 11 for the definition of baseline. Tumor response will be based on the IRC assessment. It will be derived as:

[(Sum of target lesions at week *XX* – sum of target lesions at baseline)/sum of target lesions at baseline] x 100

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

Minimum of [(sum of target lesions at week *XX* – sum of target lesions at baseline)/sum of target lesions at baseline] x 100

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

## 16.3.4 Disease Control Rate (DCR)

Disease Control Rate (DCR) is the proportion of subjects with confirmed BOR of CR, PR, SD and Non-CR/Non-PD (applicable only to subjects with non-measurable disease at baseline) according to RECIST v1.1 and adjudicated by the IRC obtained among all tumor assessment visits after the date of randomization until documented disease progression in the analysis set of interest.

DCR by treatment group will be calculated along with the two-sided 95% CI using the Clopper-Person method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option).

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

| | |
|---|---|
| [redacted] | [redacted] |
| [redacted] | [redacted] |
| [redacted] | [redacted] |

[redacted]

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

[redacted]

## 17 Safety Evaluation

*Analysis sets: Safety Analysis Set*

The subsections in this section include specifications for summarizing safety endpoints that are common across clinical trials such as adverse events, laboratory tests and vital signs. All safety related summary tables will be based on the analysis sets as indicated above. The safety endpoints will be tabulated using descriptive statistics.

## 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 unless otherwise specified. The adverse event (AE) listings will include all AEs (whether treatment-emergent or not). A separate listing for AEs that 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. Details are included in Table 14 in Appendix IV).
- **Infusion Related Reactions (IRR)**: IRRs are identified based on a list of MedDRA PTs. The detailed criteria of the timing relationship to infusion are specified in Table 15 in Appendix IV.

Unless otherwise stated adverse events will be displayed in terms of frequency tables: PT and primary SOC by decreasing frequency based on the avelumab + BSC treatment arm.

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

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 grade per the latest National Cancer Institute – Common Terminology Criteria for Adverse Events (NCI-CTCAE version 4.03) per subject, using the latest version of MedDRA PT as event category and MedDRA primary SOC as body system category.

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

The following tables will be created:

- The overall summary of AEs table will include the following categories:
  - 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% in any treatment arm by SOC and PT

In addition, a listing of AEs with onset or worsening date after the on-treatment period 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:

- TEAEs leading to treatment discontinuation by SOC and PT
- Related TEAEs leading to treatment discontinuation by SOC and PT

A data listing of TEAEs leading to 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 treatment, deaths within 60 days after the date of first dose of study treatment, and primary reason for death will be tabulated based on information from the "Report of Subject Death" and "Survival Follow-Up" CRFs.

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

In addition, date and primary reason for death will be provided in a data listing together with selected dosing information (study treatment received, date of first / last dose administration, dose and number of infusions received of study treatment) and will include the following information:

- AEs with fatal outcome (list preferred terms of AEs with outcome=Fatal, as well as AEs of Grade 5)
- Flag for death within 30 days of last dose of study treatment
- Flag for death within 60 days of first dose study treatment

### 17.2.2 Serious Adverse Events

The following overall frequency tables will be prepared for treatment-emergent SAEs:

- SAEs by SOC and PT
- Related SAEs by SOC and PT

A data listing of SAEs will also be provided with the relevant information with a flag for treatment-emergent SAEs.

### 17.2.3 Other Significant Adverse Events

The following overall frequency tables will be prepared for treatment emergent irAEs, by treatment group. The cluster is a compilation of PTs that are categorized by immune-related event of special interest as specified in Appendix IV.

- The overall summary of irAEs table will include the following categories:
  - All irAEs
  - Related irAEs
  - Serious irAEs
  - Related serious irAEs
  - irAEs, Grade ≥ 3
  - Related irAEs, Grade ≥ 3
  - irAEs leading to permanent treatment discontinuation
  - Related irAEs leading to permanent treatment discontinuation
  - irAEs leading to death
  - Related irAEs leading to death
- irAEs leading to death, by Cluster and PT
- Related irAEs leading to death, by Cluster and PT
- irAEs by Cluster and PT
- irAEs, grade ≥ 3, by Cluster and PT
- Related irAEs by Cluster and PT
- Related irAEs, grade ≥ 3, by Cluster and PT
- irAEs leading to permanent treatment discontinuation by Cluster and PT
- Related irAEs leading to permanent treatment discontinuation by Cluster and PT
- irAEs by Cluster and PT and worst grade
- Related irAEs by Cluster and PT and worst grade

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

The following overall frequency tables will be prepared for treatment emergent IRRs, by treatment group:

- The overall summary of IRR table will include the following categories:
  - All IRRs
  - Related IRRs
  - Serious IRRs
  - Related serious IRRs
  - IRRs, Grade ≥ 3
  - Related IRRs, Grade ≥ 3
  - IRRs leading to permanent treatment discontinuation
  - Related IRRs leading to permanent treatment discontinuation
  - IRRs leading to death
  - Related IRRs leading to death
- 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
- IRRs leading to death, by PT
- Related IRRs leading to death, 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]

[redacted]

## 17.3 Clinical Laboratory Evaluation

### 17.3.1 Hematology and Chemistry Parameters

Hematology and chemistry 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).

The worst grade during the on-treatment period will be summarized per treatment group considering only subjects with post baseline laboratory samples: Laboratory tests by NCI-CTC grade (0, 1, 2, 3, 4).

Quantitative data will be summarized using descriptive statistics (mean, StDev, 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). Changes from baseline at each nominal visit over time will also be presented in boxplots. End of Treatment visit will be summarized separately. Qualitative data based on reference ranges will be described according to the categories (i.e. Low, Normal, High).

If both central and local labs are collected for a subject, these summary statistics by visit will be based only on the central lab collected data, while summaries of worst on-treatment abnormalities will be based on both local and central lab data.

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

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

Corrected Calcium (mg/dL) = measured total Calcium (mmol/L) + 0.02 (40 - serum 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 by treatment group:

- 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, with different symbols for different treatment groups, by graphically displaying

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

In addition, a listing of all TBILI, ALT, AST and ALP values for 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 (number of subjects and percentages) during the on-treatment period.

- 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, 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 (ALP) (alkaline phosphatase increased), Alanine Aminotransferase (ALT) (ALT increased), Amylase (serum amylase increased), Aspartate Aminotransferase (AST) (AST increased), Total Bilirubin (blood bilirubin increased, Cholesterol (cholesterol high), Creatinine (creatinine increased), Creatine Kinase (CPK increased), Potassium (hypokalemia/ hyperkalemia), Sodium

(hyponatremia/ hypernatremia), Magnesium (hypomagnesemia/hypermagnesemia), Calcium (hypocalcemia/ hypercalcemia), Glucose (hypoglycemia/hyperglycemia), Gamma Glutamyl Transferase (GGT) (GGT increased), Lipase (lipase increased), Phosphates (hypophosphatemia), Triglycerides (hypertriglyceridemia).

**Parameters with NCI-CTC grades not available:**

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

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

In this study, these apply to the following laboratory tests:

- Hematology:

  Basophils (absolute and percent), Eosinophils (absolute and percent), Monocytes (absolute and percent), Neutrophils (absolute and percent), 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.

### 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.

- Hemostaseology: activated partial thromboplastin time (aPTT), prothrombin time (PT), and International Normalized Ratio (INR).
- Urinalysis: all urinalysis parameters
- Other parameters: hormone and immunology parameters
- Pregnancy test

Listings of laboratory results will be provided for all laboratory parameters. These listings will be sorted by parameters and assessment dates or visits for each subject; laboratory values that are outside the normal range will be flagged along and corresponding normal ranges will be provided in addition to NCI-CTCAE grade.

In addition, listings of abnormal values will be provided for chemistry, hematology, hemostaseology, and urinalysis parameters. If there is at least one abnormal result for any parameter, all the data for that parameter will be included in 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, StDev, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each planned visit over time. Discontinuation/End of Treatment visit will be summarized separately. The changes computed will be the differences from baseline.

The maximum on-treatment change from baseline will be calculated and categorized by vital sign measurement (see Table 8). Missing values will define a separate category. A summary of maximum shift from baseline by category will be provided by treatment arm.

**Table 8 Categories of Change from Baseline for Vital Sign Parameters**

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

bpm = beats per minute; DBP=diastolic blood pressure; SBP=systolic blood pressure.

## 17.5 Other Safety or Tolerability Evaluations

### 17.5.1 ECG

The 12-lead Electrocardiogram (ECG) assessment will be performed during screening (baseline) and at the Discontinuation / End-of-Treatment visit. For each of the ECG parameters, descriptive statistics at baseline and at the Discontinuation / End-of-Treatment visit and changes from baseline will be summarized by treatment group.

The incidence and percentage of subjects with the worst potentially clinically significant abnormalities (PCSA) for ECG parameters will be summarized during the on-treatment period. Each subject will be counted only once within each category. As ECG assessments are only performed during screening and at the Discontinuation/End-of-Treatment visit, the denominator to calculate percentages for each PCSA category is the number of subjects with Discontinuation/End of Treatment visit. The PCSA criteria are provided in Table 9.

**Table 9 Potentially Clinically Significant Abnormalities criteria for ECG**

| Test | Potentially Clinically Significant Abnormalities (PCSA) Criteria |
|---|---|
| Heart Rate | ≤ 50 bpm and decrease from baseline ≥ 20 bpm<br>≥ 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 | >450 ms<br>>480 ms<br>>500 ms |
| QTcF and QTcB change from baseline | Increase from baseline > 30 ms<br>Increase from baseline > 60 ms |

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/heart rate. Baseline QTcF and QTcB will be derived from the visit that other ECG parameters are flagged as baseline.

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

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

## 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.

CCI

CCI

CCI

[redacted]

# 18 References

1. Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982;38:29-41.
2. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons 1980.
3. Westfall P, Tobias R, Rom D, et al. Multiple Comparisons and Multiple Tests Using SAS, Cary, NC: SAS Institute Inc. 1999: 32-33.
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. Cochran WG. Some Methods for Strengthening the Common Chi-square Tests. Biometrics 1954;10:417-51.
6. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 Bio-Algorithms and Med-Systems 2013; 9(4): 183–9.
7. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. DOI 10.1200/jco.1998.16.1.139 J Clin Oncol. 1998 Jan;16(1):139-44.
8. AS Pickard, MP Neary, D Cella : Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer Health Qual Life Outcomes 5: 70,2007
9. Maringwa JT, Quinten C, King M, Ringash J, Osoba D, Coens C, et al. Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials.Support Care Cancer2011;19:1753-60

# 19 Appendices

## Appendix I Important and Clinically Important Protocol Deviations by Programming Check and Medical Review

Inclusion and exclusion criteria are referenced according to Protocol version 7.0 dated 10 May 2017

| Description of Protocol Deviation | Deviation Code | Clinically Important PD? | Proposed check / comment |
|---|---|---|---|
| **Inclusion/Exclusion Criteria** | | | |
| Subject did not meet inclusion criterion #4; e.g. histologically confirmed adenocarcinoma of the stomach or the gastroesophageal junction | INCEXC02 | Yes | Programming check and Medical review required |
| Subject did not meet inclusion criterion #5; e.g. 5. Documented objective radiographic or clinical disease progression that may be confirmed by pathologic criteria. | INCEXC03 | | Programming check and Medical review required |
| Subject did not meet inclusion criterion #6a; e.g. Subject has received a different regimen from the list of the approved protocol, as the 1st line of treatment | INCEXC04 | Yes | Programming check and Medical review required |
| Subject did not meet inclusion criterion #6b; e.g. Subject has received a different regimen from the list of the approved protocol, as the 2nd line of treatment | INCEXC06 | Yes | Programming check and Medical review required |
| Subject did not meet inclusion criterion #8; e.g. Subject did not have adequate hematological function defined as per protocol | INCEXC08 | | Programming check and Medical review required |
| Subject did not meet inclusion criterion #9; e.g. Adequate hepatic function defined as per protocol | INCEXC09 | | Programming check and Medical review required |
| Subject did not meet inclusion criterion #10; e.g. Adequate renal function defined as per protocol | INCEXC10 | | Programming check and Medical review required |
| Subject did not meet inclusion criterion #12; e.g. Effective contraception for both male and female subjects if the risk of conception exists | INCEXC11 | | Programming check and Medical review required |
| Subject met exclusion criterion #1; e.g. Subject has received prior therapy with any antibody/drug targeting T cell co-regulatory proteins | INCEXC12 | Yes | Programming check and Medical review required |
| Subject met exclusion criterion #2; e.g. Subject receives concurrent anticancer treatment | INCEXC13 | Yes | Programming check and Medical review required |
| Subject met exclusion criterion #3; e.g. Subject had major surgery for any reason, except diagnostic biopsy, within 4 weeks of the trial treatment and/or if the subject has not fully recovered from the surgery within 4 weeks of the study treatment | INCEXC14 | | Programming check and Medical review required |

| Description of Protocol Deviation | Deviation Code | Clinically Important PD? | Proposed check / comment |
|---|---|---|---|
| **Inclusion/Exclusion Criteria** | | | |
| Subject met exclusion criterion #4; e.g. Subject has received immunosuppressive agents and were not tapered off these drugs before initiation of the study treatment | INCEXC15 | Yes | Programming check and Medical review required |
| Subject met exclusion criterion #6; e.g. Subject has previous malignant disease (other than gastric cancer) within the last 5 years with the exception of basal or squamous cell carcinoma of the skin or carcinoma in situ (bladder, cervical, colorectal, breast) | INCEXC16 | Yes | Programming check and Medical review required |
| Subject met exclusion criterion #9; e.g. Subject has active immune disease | INCEXC17 | Yes | Programming check and Medical review required |
| Subject met exclusion criterion #14; e.g. Pregnancy or lactation | INCEXC18 | | Programming check and Medical review required |
| Any other inclusion criteria response "No" and/or any "Yes" response to any other exclusion criteria and subject was subsequently randomized | INCEXC19 | | Programming check and Medical review required |
| ECOG PS not performed or >1 (at screening) | INCEXC21 | | Programming check and Medical review required |
| **Informed Consent/ Subject information** | | | |
| Subject did not sign ICF and was enrolled | INFCON01 | | Programming check and Medical review required |
| ICF is signed after study procedure was done (exceptions: assessments such as CT/MRI that were performed as part of normal site procedures shall not raise a deviation. However, they must be performed within the 28 day screening window) | INFCON02 | | Programming check and Medical review required |
| Subject did not sign ICF for Pharmacogenetics even though is participating in the sub-study | INFCON04 | | Programming check and Medical review required |
| **Investigational Product** | | | |
| Subjects randomized in one treatment but received the incorrect treatment | INVPRO05 | Yes | Programming check and Medical review required |
| Avelumab premedication was not administered prior to IP dose. | INVPRO10 | | Programming check and Medical review required |
| **Randomization** | | | |
| Patient randomized more than once | RANDOM02 | | Programming check and medical review required |

# Appendix II EORTC QLQ-C30 Version 3.0

**Table 1: Scoring the QLQ-C30 version 3.0**

| | Scale | Number of items | Item range* | Version 3.0 Item numbers | Function scales |
|---|---|---|---|---|---|
| **Global health status / QoL** | | | | | |
| Global health status/QoL (revised)† | QL2 | 2 | 6 | 29, 30 | |
| **Functional scales** | | | | | |
| Physical functioning (revised)† | PF2 | 5 | 3 | 1 to 5 | F |
| Role functioning (revised)† | RF2 | 2 | 3 | 6, 7 | F |
| Emotional functioning | EF | 4 | 3 | 21 to 24 | F |
| Cognitive functioning | CF | 2 | 3 | 20, 25 | F |
| Social functioning | SF | 2 | 3 | 26, 27 | F |
| **Symptom scales / items** | | | | | |
| Fatigue | FA | 3 | 3 | 10, 12, 18 | |
| Nausea and vomiting | NV | 2 | 3 | 14, 15 | |
| Pain | PA | 2 | 3 | 9, 19 | |
| Dyspnoea | DY | 1 | 3 | 8 | |
| Insomnia | SL | 1 | 3 | 11 | |
| Appetite loss | AP | 1 | 3 | 13 | |
| Constipation | CO | 1 | 3 | 16 | |
| Diarrhoea | DI | 1 | 3 | 17 | |
| Financial difficulties | FI | 1 | 3 | 28 | |

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

Then for **Functional scales**:

$$Score = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$$

and for **Symptom scales / items** and **Global health status / QoL:**

$$Score = \{(RS-1)/range\} \times 100$$

**Examples:**

Emotional functioning

$$RawScore = (Q_{21} + Q_{22} + Q_{23} + Q_{24})/4$$
$$\text{EF } Score = \{1 - (RawScore - 1)/3\} \times 100$$

Fatigue

$$RawScore = (Q_{10} + Q_{12} + Q_{18})/3$$
$$\text{FA } Score = \{(RawScore - 1)/3\} \times 100$$

# Appendix III EORTC QLQ-C30-STO22: Gastric Cancer Module

## Gastric cancer module: QLQ-STO22

### Scope

The gastric cancer module is meant for use among patients with gastric cancer varying in disease stage and treatment modality (i.e. surgery, chemotherapy, radiotherapy, etc.). It should always be complemented by the QLQ-C30.

### Scoring

| | Scale name | Number of items | Item range | QLQ-STO22 item numbers |
|---|---|---|---|---|
| **Functional scales** | | | | |
| Body image | STOBI | 1 | 3 | 49 |
| **Symptom scales** | | | | |
| Dysphagia | STODYS | 3 | 3 | 31 – 33 |
| Pain | STOPAIN | 4 | 3 | 34 – 37 |
| Reflux symptoms | STORFX | 3 | 3 | 38 – 40 |
| Eating restrictions | STOEAT | 4 | 3 | 41 to 43,46 |
| Anxiety | STOANX | 3 | 3 | 47,48,50 |
| Dry mouth | STODM | 1 | 3 | 44 |
| Taste | STOTA | 1 | 3 | 45 |
| Body image | STOBI | 1 | 3 | 49 |
| Hair loss | STOHL | 2/1 | 3 | 51,52* |

### Remarks

- Item 52 is an optional item and depends on the answer to item 51. Item 52 should only be answered and assessed if 'yes' has been answered to item 51.

# Appendix IV Description of the Case Review for Assessment of Immune-Related AEs and IRR

In order to thoroughly and consistently analyze potential immune-mediated adverse events (AEs), a two-level approach is proposed including:

1. A MedDRA Preferred Term (PT) query is proposed 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, the MedDRA PT queries will be used to search for AEs of interest in the clinical database. The proposed event categories such as:

Immune-mediated rash, Immune-mediated colitis, Immune-mediated pneumonitis, Immune-mediated hepatitis, Immune-mediated nephritis and renal dysfunction, Immune-mediated endocrinopathies (Thyroid disorders: Hypothyroidism, Hyperthyroidism, and Thyroiditis), Immune-mediated endocrinopathies (Adrenal insufficiency, Immune-mediated endocrinopathies (Type 1 Diabetes Mellitus), Immune-mediated endocrinopathies (Pituitary dysfunction), Immune-mediated endocrinopathies (Hypogonadism), Other immune-mediated adverse events. Further details e.g. MedDRA PT queries are regularly updated based on the current MedDRA version.

In order to standardize the MedDRA PT queries as much as possible, High Level Terms (HLT) and Standardized MedDRA Queries (SMQ) were used whenever a choice, that was considered reflective of the events of interest, was available.

**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 14 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.<br>*For endocrinopathies only*: AE required hormone replacement* and /or (corticosteroid or other immunosuppressive therapy) |

| | |
|---|---|
| **Etiology** | No other clear etiology or<br>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 endocrinopathy disorders only as follows:<br>• Thyroid disorders (HLT): Thyroid therapy (ATC codes (H03A, H03B))<br>• Diabetes mellitus (including hyperglycaemia): 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 17.

### Table 15 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):**<br>• Infusion related reaction<br>• Drug hypersensitivity<br>• Anaphylactic reaction<br>• Hypersensitivity<br>• Type 1 hypersensitivity<br><br>**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**<br>• Pyrexia<br>• Chills<br>• Flushing<br>• Hypotension<br>• Dyspnoea<br>• Wheezing<br>• Back pain<br>• Abdominal pain<br>• Urticaria |