Appendix 16.1.9 Documentation of Statistical Methods

- EMR100070-007 Statistical Analysis Plan; Version 4.0, 23 July 2019

CCI

Statistical Analysis Plan

Clinical Trial Protocol Identification No.	EMR 100070-007
Title:	A Phase III open-label, multicenter trial of maintenance therapy with avelumab (MSB0010718C) versus continuation of first-line chemotherapy in patients with unresectable, locally advanced or metastatic, adenocarcinoma of the stomach, or of the gastro-esophageal junction
Trial Phase	Phase III
Investigational Medicinal Product(s)	MSB0010718C Avelumab
Clinical Trial Protocol Version	17 July 2019/Version 7.0
Statistical Analysis Plan Author	PPD
Statistical Analysis Plan Date and Version	23 July 2019/ Version 4.0
Statistical Analysis Plan Reviewers	PPD PPD PPD PPD PPD PPD PPD

1 Table of Contents

[redacted]

2 List of Abbreviations and Definition of Terms

CCI	[redacted]
ADaM	Analysis Data Model
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BOR	Best Overall Response
BSA	Body Surface Area
CCI	[redacted]
CI	Confidence Interval
CIPD	Clinically Important Protocol Deviation
CMH	Cochran-Mantel-Haenszel
CR	Complete Response
CRF	Case Report Form
CSR	Clinical Study Report
CT	Computed Tomography
CTP	Clinical Trial Protocol
CTMS	Clinical Trial Management System
DCR	Disease Control Rate
CCI	[redacted]
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic Case Report Form
EEA	European Economic Area
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
EORTC	European Organization for Research and Treatment of Cancer
EOT	End of Treatment
EQ-5D-5L	European Quality of Life (EuroQoL) - 5 Dimensions - 5 Levels
EU	European Union

[redacted]

FAS	Full Analysis Set
GCP	Good Clinical Practice
GEJ	Gastro-esophageal Junction
HR	Hazard Ratio
HRQoL	Health Related Quality of Life
ICF	Informed Consent Form
ICH	International Conference on Harmonization
IDMC	Independent Data Monitoring Committee
IPD	Important Protocol Deviation
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
OR	Objective Response
ORR	Objective Response Rate
OS	Overall Survival
PCSA	Potentially Clinically Significant Abnormalities
PD	Progressive Disease
CCI	[redacted]
PFS	Progression-Free Survival
CCI	[redacted]
PP	Per-Protocol
PR	Partial Response

[redacted]

PT	Preferred Term
Q1	First Quartile
Q3	Third Quartile
QoL	Quality of Life
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SAF	Safety
SAP	Statistical Analysis Plan
SD	Stable Disease
SDTM	Study Data Tabulation Model
SOC	System Organ Class
StDev	Standard Deviation
TEAE	Treatment Emergent Adverse Event
VAS	Visual Analogue Scale

3 Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	PPD	PPD	NA. The first version.
2.0	PPD	PPD	Details of changes are specified in Section 4.1 "Changes to Previous Version"
3.0	PPD	PPD	Details of changes are specified in Section 4.1 "Changes to Previous Version"
4.0	PPD	PPD	Details of changes are specified in Section 4.1 "Changes to Previous Version"

4 Purpose of the Statistical Analysis Plan

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specifications for the primary analysis of data collected for Clinical Trial Protocol (CTP) EMR100070-007. 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 regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective SAP will be clearly identified in the CSR.

The SAP is based upon Section 8 (Statistics) of the CTP dated 21 June 2018/ version 6.0 and is prepared in compliance with ICH E9.

4.1 Changes to Previous Version

Version 4.0

1. *Statistical Analysis Plan Reviewers* updated for current list of applicable study team members.

CCI [redacted]

3. Sections 5 *Summary of Clinical Trial Features*, Section 6 *Sample Size/Randomization*, and Section 7 *Overview of Planned Analyses* updated to remove the requirement of 112 PD-L1+ OS events required to proceed with the Primary Analysis.

4. Section 6 *Sample Size/Randomization* updated to clarify that target of 112 PD-L1+ OS events will not be reached during the study; use of 112 OS events in this population in efficacy boundary calculations is for illustrative purposes only.

[redacted]

5. Section 9 *Analysis Sets* updated to add new, additional subgroups

6. Section 10 *General Specifications for Statistical Analyses* updated to specify that only central labs are used as baseline, as per the project-level harmonized SAP. Updated to specify imputation of laboratory values, in the case the result is above or below the limit of quantitation; updated to specify that only on-treatment records are summarized in safety by-visit summaries; updated to consider BSC-Only "treatment" as having ended if subject starts subsequent anti-cancer therapy; Maintenance Phase baseline definition for HRQoL and ECOG assessments updated to clarify that assessments prior to the pre-randomization window (28 or 14 days, respectively) will not be available for selection as baseline.

CCI [redacted]

8. Section 14 *Treatment Compliance and Exposure* updated to remove consideration of reasons when calculating length of dose interruptions, as the data as collected does not allow for this to be done.

9. Section 15.1.2 *Sensitivity Analyses and Model Checks* updated to remove scenario wherein OS events in PD-L1+ subjects exceeds planned number, as it is impossible to achieve the planned 112 PD-L1+ events

CCI [redacted]

11. Appendix I *Clinically Important Protocol Deviations* updated to include additional important deviations defined by the study team.

Version 3.0

1. The following changes are made in version 3.0 of the SAP: Section 2, previously signature page, was deleted as SAP will be signed electronically. Subsequent sections re-numbered.

CCI [redacted]

3. Section 9 *Analysis Sets* and section 16.2.2 *Best Overall Response and Objective Response* updated to include analysis by HRQoL analysis set.
4. Section 10 *General Specifications for Statistical Analyses* updated to add specifications for determination of Maintenance Phase baseline for ECOG and HRQoL analyses;
   Date of last contact updated to include "Date last known to be alive" on eCRF page, even if status is other than "Alive" at time of entry;
   sub-section "Definition of study treatment, in the context of subjects randomized to receive BSC-only" updated to correct use of "End of Assessment" eCRF page rather than "End of Treatment" eCRF page, which is not used for this population;

imputation of concomitant medication dates updated to use induction phase medication start date as the reference, not randomized medication start date;
imputation of subsequent anti-cancer therapy dates to align with new program-level SAP template;
Induction Phase On-Treatment period updated to continue to start of Maintenance Phase treatment start, rather than randomization date.

5. Section 11.1 *Disposition of Subjects and Discontinuations* updated to clarify that subjects who complete Induction Phase chemotherapy (randomized to avelumab or BSC-Only treatment groups) will not have discontinuation reason presented on the table.
6. Section 13.2 *Subsequent Anti-Cancer Therapies/Procedures* updated to provide one listing of subsequent anti-cancer treatments with all subjects, rather than two separate listings with subsets of subjects in each.
7. Section 14 *Treatment Compliance and Exposure* updated to clarify that dose interruptions are for IV medications; correct treatment duration calculation for Oxaliplatin + Capecitabine group to use a 3-week cycle; define duration of regimen; remove the summary of Infusion Rate Reduction ≥ 50%, as this value cannot be calculated with eCRF data.
8. Section 15.1.1 *Primary Efficacy Analysis* of Overall Survival updated to include month 21 in time list presented in outputs.
9. Section 15.1.2 *Sensitivity Analyses and Model Checks* updated to include sensitivity analysis of OS which censors subjects at start date of subsequent anti-cancer treatment.
10. Section 15.2.1 *Progression-Free Survival* updated to include months 18 through 24 in time list presented in outputs, consistent with OS display.
11. Section 15.2.2 *Best Overall Response and Objective Response* updated to present "NED" as the overall timepoint response for subjects who have CR at Re-Baseline and no visible disease on the subsequent scan. NED is not a possible "Best Overall Response" result.
12. Section 15.2.3 *Subject Reported Outcomes /Health-Related Quality of Life* was updated to specify that "baseline" is the Maintenance Phase baseline for all analyses;
Abbreviations of QoL were updated to HRQoL throughout the document, except where contained in the
CCI [redacted]
14. Section 16.3.2 *Other Laboratory Parameters* updated to remove additional listings for abnormal hemostaseology and urinalysis parameters. Full listings for these parameters will be provided and will indicate abnormal values.
15. Appendix IV *Description of the Case Review for Assessment of Immune-Related AEs and IRRs* updated for new irAE selection criteria.

Version 2.0

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

1. Updates to spelling, grammar, and word choice were updated and clarifications of existing ideas were added throughout the document.
2. SAP Author and Reviewers updated.

[redacted]

3. Section 2 *List of Abbreviations and Definitions* was updated to include additional abbreviations used in this document and remove others that are not used.
4. Section 5 *Summary of Clinical Trial Features* was updated as per Protocol Version 6.0, including updating primary endpoint to OS in FAS and PD-L1+ subjects; dropping PFS as primary endpoint; adding PFS as key secondary endpoint; CCI [redacted]
5. Section 6 *Sample Size/Randomization* was updated to drop considerations for PFS, add OS in PD-L1+ subjects, including details of simulation used for power calculation.
6. Section 6 *Sample Size/Randomization and Section* 8 *Overview of Planned Analyses* were updated to specify that data cut-off date for the interim and the primary analysis will be based on a prospectively-determined date; updated O'Brien-Fleming boundaries to account for the actual number of events at the interim and the final analysis, in the case observed events deviates from the planned number of events
7. Section 7.1 *Sequence of Analysis* was updated to specify that the primary analysis will occur after a minimum 18 months after the last subject is randomized and 112 OS events in the PD-L1+ population.
8. Section 7.2 *Interim Analysis* was updated to clarify role of IDMC members
9. Section 7.3 *Primary Analysis* includes explanation for removal of IRC assessment
10. Section 9 *Analysis Sets* was updated to include PD-L1+ Analysis Set, for Primary Endpoint; Per-Protocol Analysis Set updated for additional exclusion criteria, remove requirement for post-baseline tumor assessment; Safety-Maintenance Analysis Set updated to clarify handling of subjects randomized to control, with BSC-Only as treatment; addition of HRQoL Analysis Set CCI [redacted] Table 1 updated for analyses; Subgroups updated to add second age subgroup, remove Intent of Therapy subgroup, add Re-Baseline response subgroup.
11. Section 10 *General Specifications for Statistical Analyses* was updated to state that Maintenance Phase begins at Randomization; explanation of Re-Baseline tumor scan included; baseline definition updated for subjects randomized to control arm, with BSC-Only as treatment; Study Day definition broken into Induction Phase Study Day and Maintenance Phase Study Day; explanation of study treatment in the context of BSC-only subjects added; added calculation for Age relative to Maintenance Phase; added explanation of summary statistics over time; date imputation for subsequent anti-cancer therapies added.
12. Section 11.1 *Disposition of Subjects and Discontinuations* was updated for separate summaries based on phase; for region.
13. Section 11.1 *Protocol Deviations* updated for reporting of subjects randomized more than once.
14. Section 12.1 *Demographics* was updated to add EEA; display of regions; clarify Maintenance Phase version of the summary.

15. Section 12.3.1 *Disease Characteristics* was updated to include re-baseline tumor response.
16. Section 12.3.3 *Induction Phase ORR* was eliminated, as this information is now included in Disease Characteristics summary; Section 13.3.3 *PD-L1 Expression Status* was added to clarify summary of PD-L1 status.
17. Section 13.1 *Prior and Concomitant Medications/Procedures* was updated to include definition of premedications, concomitant procedures; summary table for pre-medications and concomitant procedures.
18. Section 13.2 *Subsequent Anti-Cancer Therapies/Procedures* was updated with more details;
19. Section 14 *Treatment Compliance and Exposure* was updated to add clarification for the definition of relative dose intensity and dose reductions; addition of definition of Dose Interruptions;
20. Section 15 *Endpoint Evaluation* was updated for new primary/secondary endpoints; inclusion of explanation of alpha recycling.
21. Section 15.1.2 *Progression Free Survival* removed from section 16.1, relocated to 16.2.1
22. Section 15.1.2 (previously 16.1.3) *Sensitivity Analyses* updated for numbering (16.1.2); inclusion of additional analyses; details of RMST analysis was added.
23. Section 15.1.4 (previously 16.1.5) *Subgroup Analysis of Primary Endpoint* was updated to include the overall HR and 95% CI in the forest plot; the text was corrected to specify the point estimate of the interaction model parameter would be provided; models were updated to remove "factors".
24. Section 15.2.2 *Best Overall Response and Objective Response* updated to clarify handling of subjects with CR at Re-Baseline; reasoning for confirmation of response after non-consecutive assessment; inclusion of interaction analysis and multivariate analysis;
25. Section 15.2.3 *Subject-reported Outcomes/Quality of Life* was updated to include more details on the HRQoL scoring system; completion and compliance summary, mixed-effects model repeated measures analysis and more details on summary of change from baseline;

CCI
CCI

28. Section 16 *Safety Evaluation* was updated to include Table 8 detailing analyses and analysis sets.
29. Section 16.1 *Adverse Events* was updated to including details of pooling the same AE with different toxicity grade, outcome or seriousness recorded as different entries on the eCRF; drop summary of related irAE; was updated for the definition of immune-related AE (irAE) and infusion-related reactions (IRRs).
30. Section 16.2.3 *Other Significant Adverse Events* was updated to delete related irAEs from the irAE analysis; replace SOC with Cluster; include overall summary table of IRRs

CCI

32. Section 16.3.1 *Hematology and Chemistry Parameters* was updated to specify the denominator for worst-grade during the on-treatment period analysis;
33. Section 16.5.1. *ECG* was updated to specify the denominator of ECG PCSA summary table;

5 Summary of Clinical Trial Features

Trial objectives	**Primary objectives:** The primary objectives are to demonstrate superiority of maintenance therapy with avelumab versus continuation of first-line chemotherapy with regard to Overall Survival (OS) in all randomized subjects or in PD-L1+ subjects who have not progressed on first-line chemotherapy. **Secondary objectives:** • To demonstrate superiority of maintenance therapy with avelumab versus continuation of first-line chemotherapy with regard to Progression-Free Survival (PFS) as per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1) and per Investigator assessment. • To demonstrate superiority of maintenance therapy with avelumab versus continuation of first-line chemotherapy with regard to objective response rate (ORR) as per RECIST v1.1 according to investigator assessment. • To compare the subject-reported outcomes / quality of life (HRQoL) of subjects when treated with avelumab versus continuation of first-line chemotherapy as assessed by the European QoL 5-dimensions 5-levels questionnaire (EQ-5D-5L), and the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 and module QLQ-STO22 • To determine the safety and tolerability of avelumab CCI

	[redacted]
Trial design and plan	This is a multicenter, international, randomized, open-label Phase III trial of avelumab in subjects with advanced (unresectable, locally advanced or metastatic) adenocarcinoma of the stomach, or of the gastroesophageal junction (GEJ) who are treatment naïve and have not yet received chemotherapy for the treatment of metastatic or locally advanced disease. This study includes a screening period, Induction Phase, Maintenance Phase, and Follow-up Phase. The actual number of subjects enrolled in the Induction Phase will be driven by observed induction failure rates and will be such to allow for approximately 466 subjects to be randomized in the Maintenance Phase. Subjects enrolled in the Induction Phase will receive induction chemotherapy comprised of oxaliplatin and either 5-fluorouracil (5-FU) or capecitabine (Induction Phase) for 12 weeks. Following the Induction Phase, subjects who experience a complete response (CR), partial response (PR), or stable disease (SD), as verified by an independent radiologist, will enter the Maintenance Phase and be randomly assigned to receive either avelumab, or continuation of the same chemotherapy regimen from the Induction Phase (Maintenance Phase).

	The dose and schedule of the chemotherapy during the Induction Phase are as follows or in accordance to the label instructions and per local guidelines: • Oxaliplatin at 85 mg/m^2 intravenous (IV) on Day 1 with leucovorin 200 mg/m^2 IV on Day 1 (or equivalent levoleucovorin dose) followed by 5-FU at 2600 mg/m^2 IV continuous infusion over 24 hours on Day 1, given every 2 weeks (for up to 12 weeks) OR • Oxaliplatin at 85 mg/m^2 IV on Day 1 with leucovorin 400 mg/m^2 IV on Day 1 (or equivalent levoleucovorin dose) followed by 5-FU at 400 mg/m^2 IV push on Day 1 and 2400 mg/m^2 IV continuous infusion over 46-48 hours (Days 1 and 2) given every 2 weeks (for up to 12 weeks) OR • Oxaliplatin at 130 mg/m^2 IV on Day 1 with capecitabine at 1000 mg/m^2, twice daily for 2 weeks followed by a 1-week rest period given every 3 weeks (for up to 12 weeks). Local label and guidelines should be followed for specific starting doses due to renal or hepatic impairments and for subsequent dose modifications due to different toxicities. Upon completion of chemotherapy in the Induction Phase, subjects without disease progression (subjects with SD, PR, or CR per independent radiologist) will be eligible for randomization to the Maintenance Phase where they will receive either avelumab, or continue the same regimen of chemotherapy from the Induction Phase. Treatment during the Maintenance Phase are as follows: 1. For subjects randomly assigned to avelumab: avelumab will be given at a dose of 10 mg/kg as a 1 hour IV infusion once every 2 weeks until disease progression 2. For subjects randomly assigned to chemotherapy: the same regimen of oxaliplatin-fluoropyrimidine doublet as in the Induction Phase will be continued until disease progression. • Subjects who are not deemed eligible to receive further chemotherapy will receive best supportive care (BSC) alone with no active therapy. Prior to randomization,

	Investigators must specify BSC treatment for these subjects. • For subjects receiving chemotherapy (oxaliplatin + 5-FU or oxaliplatin + capecitabine), dose modifications after the starting dose are allowed if the continuation of the oxaliplatin-fluoropyrimidine doublet, or a component thereof, is prohibited by toxicity. For subjects intolerant to further oxaliplatin, single-agent capecitabine or 5-FU + leucovorin will be an option for dose modification. 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. Clinical decision making and study endpoint evaluation will be based on Investigator assessment of the scans using RECIST v1.1. Study treatment will continue until: • Disease progression • Significant clinical deterioration (clinical progression) by Investigator's opinion • Unacceptable toxicity by Investigator's opinion, or • Any criterion for withdrawal from the trial or trial treatment is fulfilled. Subjects receiving avelumab may continue treatment past the initial determination of disease progression per RECIST version 1.1 as long the following criteria are met: • Investigator-assessed clinical benefit, without any rapid disease progression • Tolerance of trial treatment • Stable Eastern Cooperative Oncology Group performance status (ECOG PS [PS=0 or 1]) • Treatment beyond progression will not delay an imminent intervention to prevent serious complications of disease progression (for example, central nervous system metastases). The decision to continue avelumab treatment beyond progression should be discussed with the Medical Monitor and documented in the study records.

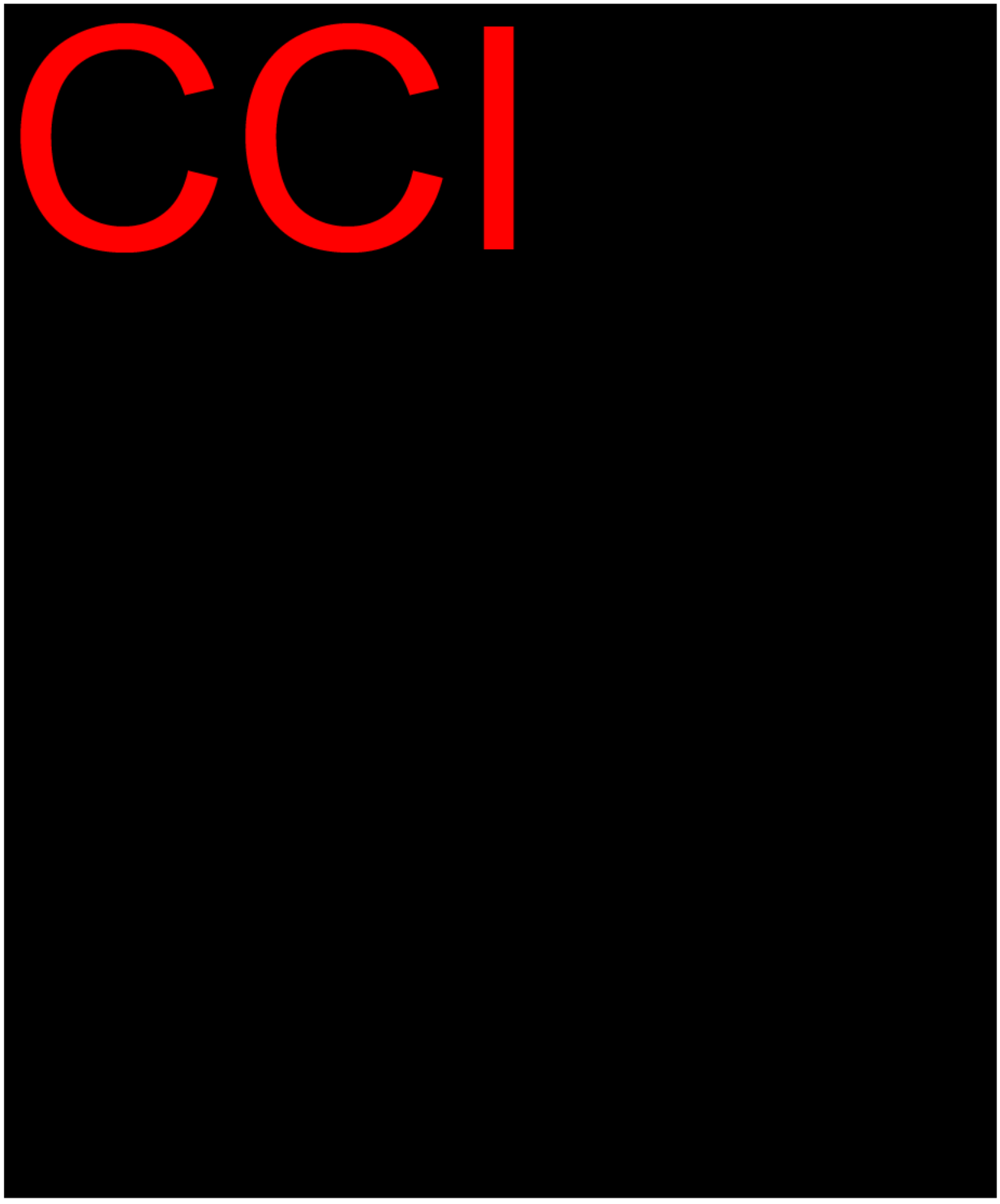
	For subjects continuing avelumab after initial progressive disease (PD), a radiographic assessment should be performed within 6 weeks of original PD to determine whether there has been a decrease in the tumor size, or continued PD. The assessment of clinical benefit should be balanced by clinical judgment as to whether the subject is clinically deteriorating and unlikely to receive any benefit from continued treatment with avelumab. For subjects receiving avelumab, if discontinuation occurs due to progression and a definitive diagnosis/radiographic confirmation is not made at the time of discontinuation, a second imaging scan may be allowed for confirmation of progression. If progression at the second imaging scan is not confirmed and the subject wishes to restart, the subject will be allowed to continue receiving avelumab as long as they meet the criteria for continuation of treatment beyond progression. If the Investigator feels that the subject continues to achieve clinical benefit by continuing treatment, the subject should remain on the trial and continue to receive monitoring according to the Schedule of Assessments. Any additional continuation of avelumab plus BSC beyond further progression must be discussed and agreed upon with the Medical Monitor. Further disease progression is defined as an additional increase in tumor burden of 20% and ≥ 5 mm absolute increase in tumor burden from time of initial PD. This includes an increase in the sum of all target lesions and/or the development of new measurable lesions. Subjects receiving avelumab who have experienced a CR should be treated for a minimum of 12 months or until disease progression or unacceptable toxicity, after confirmation of response. 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 after agreement with 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 follow the Schedule of Assessments until disease progression. Subjects in the maintenance chemotherapy arm will receive trial treatment until PD per RECIST v1.1, significant clinical deterioration (clinical progression), unacceptable toxicity, withdrawal of consent, or if any criterion for withdrawal from the trial or trial treatment is fulfilled.

	On-study subject management and study endpoint evaluation will be based on Investigator assessments. Subjects will attend clinic visits at regular intervals to receive trial treatment and for efficacy and safety assessments. After completion of the Maintenance Phase, subjects will enter the Follow-up Phase.
Planned number of patients	Approximately 466 subjects are planned to be randomly assigned into the Maintenance Phase. The actual number of subjects enrolled in the Induction Phase will be driven by observed induction failure rates and will be such to allow for approximately 466 patients to be randomized in the Maintenance Phase. At the time of this protocol amendment, subject enrollment in the Maintenance Phase is complete, with 499 subjects randomized into the Maintenance Phase.
Primary endpoint	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.
Secondary/ CCI endpoints	The key secondary endpoints are: • PFS defined as the time (in months) from randomization to the date of the first documentation of disease progression (per RECIST v1.1 and as evaluated by Investigator) or death due to any cause (whichever occurs first). • Best Overall Response (BOR) in Maintenance Phase according to RECIST v1.1 and per Investigator assessment. Other secondary endpoints include subject-reported outcomes / QoL (assessed by the EQ-5D-5L, EORTC QLQ-C30, and EORTC module QLQ-STO22 questionnaires). CCI

	CCI
Safety endpoints	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 examinations, clinical laboratory assessments, concomitant medications, vital signs, electrocardiogram (ECG) parameters, and ECOG PS.

6 CCI /Randomization

CCI



[redacted]

CCI

7 Overview of Planned Analyses

This SAP covers the analyses for efficacy and safety based on the data cut-off dates for the interim analysis, the primary analysis and the follow-up CCI analysis. Statistical analyses will be performed using cleaned eCRF data as well as external data, including quality of life (HRQoL) data, CCI and other CCI All data will be included up to a prospectively-determined cut-off date. The data cut-off date for the interim and the primary analyses will be prospectively determined based on monthly event projection provided by the unblinded team as detailed in Section 6. Since the observed number of OS events in all randomized subjects at the interim analysis may not be exactly equal to the planned 267 events, the efficacy boundary will be updated based on the actual number of observed events using the pre-specified alpha-spending function. The observed Z-test statistic at the interim analysis will be compared with the updated efficacy boundary. For the primary analysis of OS in all randomized subjects, if the number of events deviates from the target of 356 events, the primary analysis criteria will be determined taking into account the actual alpha spent at the interim analysis and the actual association between the two test z-scales, so that the overall one-sided significance level for OS in all randomized subjects is controlled at 0.02.

Since the formal efficacy boundaries will be used at the interim analysis for the statistical testing of OS in all randomized subjects, a statistically significant finding at the interim will be intended to claim superiority.

A separate SAP covers the interim analyses for periodic safety review and interim efficacy review by the Independent Data Monitoring Committee (IDMC).

Separate or supplemental analysis plans might be written to cover:

CCI

- HRQoL analysis

7.1 Sequence of Analysis

The following analyses will be performed during this trial:

- An interim analysis of OS in all-randomized subjects will be performed after approximately 267 post-randomization events (deaths) in all randomized subjects have been observed.
- The primary analysis of OS will be performed after approximately 356 post-randomization OS events (deaths) in all randomized subjects have been observed or 18 months after randomization of the last subject, whichever comes later.
- There will be ongoing interim analyses for periodic safety review by the IDMC. Details will be provided in the IDMC charter and the IDMC SAP.

7.2 Interim Analysis

An IDMC will be formed and will be responsible for periodic safety evaluations of the trial as well as the evaluation of the interim efficacy analysis. The IDMC consists of a group of three experts (2 clinicians and a biostatistician), who are neither participants in the trial nor employees of the sponsor of this trial, and the independent statistical provider, who is not a voting member of the IDMC. Details of the study's IDMC mission, composition, operations, and safeguarding of blinded data are provided in the IDMC charter.

For the periodic safety evaluations, 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.

One interim efficacy analysis is planned for OS in all randomized subjects. The programming of the interim safety and efficacy analyses will be produced by the blinded statistical team, using a dummy randomization file in which the treatment assignments are scrambled. An independent statistical provider will perform the unblinded interim safety and efficacy analyses to support the IDMC. Results from the interim analyses will be transmitted from this independent statistical provider to the IDMC members. Neither members of the blinded statistical team nor employees of the study Sponsor will receive copies of the unblinded datasets and/or results.

The interim analysis is planned to detect an overwhelming efficacy of OS in all randomized subjects and will take place when approximately 75% of the events, i.e. 267 OS events in all randomized subjects, have been observed. No interim analysis is planned for in PD-L1+ subjects. A Lan-DeMets alpha spending function with O'Brien-Fleming type boundaries will be used to control the local type I error rate of 2% (one-sided) for OS in all randomized subjects. If the planned number of events are observed at each analysis, one-sided boundaries of 0.0072 and 0.0178, respectively will be used for evaluating statistical significance. The O'Brien-Fleming efficacy boundaries based on a Lan-DeMets spending function will be updated prior to the time of analysis of OS based on the actual number of events and the information fraction for type I error control. Information fraction at the interim analysis will be calculated assuming the planned 356 events will be observed at the time of

the primary analysis. At the primary analysis, the boundaries will be updated based on the true number of events observed and information fraction used at each analysis. Since formal efficacy boundaries will be used at the interim analysis for the statistical testing of OS in all randomized subjects, a statistically significant finding at the interim will be intended to claim superiority. If OS in all randomized subjects is statistically significant at the interim analysis, only descriptive analysis of OS in PD-L1+ subjects will be performed at that time. No formal analysis of key secondary endpoints will be conducted at the interim analysis. If the trial proceeds to the final analysis and the null hypothesis of OS in PD-L1+ subjects is rejected at alpha level 0.005, then the boundary for the final hypothesis of OS in randomized subjects will be based on a local significance level of 0.025. Assuming the planned number of events are observed at each analysis, the 1-sided boundaries would then be 0.0233 for OS in all randomized subjects at primary analysis. This has been recalculated by keeping the alpha level of 0.0072 at interim analysis unchanged. Additional details on the multiple endpoint testing strategy are found in Section 15.

7.3 Primary Analysis

The primary analysis will use data up to a prospectively-determined cut-off date, upon which approximately 356 Maintenance Phase OS events (deaths) in all randomized subjects are expected to have been observed, or which falls 18 months after randomization of the last subject, whichever comes later.

The type I error and the efficacy boundary for the primary analysis will utilize an adjusted Lan-DeMets alpha spending function for O'Brien-Fleming boundaries, taking into account the actual number of OS events in all randomized subjects observed at the primary analysis cut-off date. The updated spending function will be based on the true/revised alpha spent at the interim analysis, updated to reflect the actual information fraction included at the interim analysis, considering the actual number of OS events in all randomized subjects observed at the primary analysis.

Protocol version 6 changed the evaluation of study endpoints from the Independent Review Committee (IRC) to the Investigator assessment. Because PFS is no longer a primary endpoint and is now a key secondary endpoint, there is no longer a need to have an IRC evaluate response. CCI [redacted] However, all induction phase scans were reviewed by a single independent radiologist to determine each subject's eligibility for randomization into the maintenance phase.

All efficacy and safety analyses outlined in this SAP will be performed at the time of the primary analysis.

A partial database lock will be performed for both the interim and the primary analyses.

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

8 Changes to the Planned Analyses in the Clinical Trial Protocol

The statistical methods specified in this SAP are in accordance with protocol version 7.0 (dated 17 July 2019), with the following exception:

- Subjects randomized to the chemotherapy arm, who are designated to receive BSC-only, will have the date of their Week 1 Day 1 assessment considered as "treatment start date". Consequently, this date will be used for determining baseline for Safety endpoints. This differs from the protocol, which specifies that the date of randomization will be used for both. Use of the Week 1 Day 1 assessment date more closely mirrors how subjects randomized to active treatment are handled.

9 Analysis Sets

The analysis for the Induction Phase and Maintenance Phase will be handled separately, and no data pooling will be applied across these treatment phases for the study.

Screening Analysis Set

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

Full Analysis Set

The Full analysis set (FAS) will include all randomized subjects. Subjects will be classified according to the treatment assigned at randomization as per the intent-to-treat principle. The FAS will be the primary analysis set for OS, PFS, and BOR efficacy endpoints.

PD-L1+ Analysis Set

The PD-L1+ Analysis Set is a subset of the FAS consisting of all PD-L1+ subjects determined by the PD-L1 IHC 73-10 pharmDx companion diagnostic assay using the tumor cell scoring method (CDx) at a 1% diagnostic cut-off. Analyses performed on the PD-L1+ Analysis Set will be performed according to subjects' treatment arms as randomized. One of two primary hypotheses for OS subjects will be tested in the PD-L1+ Analysis Set.

Per-Protocol Analysis Set

The Per-Protocol (PP) analysis set is a subset of the FAS and will include patients who do not meet any of the following criteria that could impact the key objectives of the study. The analysis of the primary and key secondary endpoints will be repeated for the PP analysis set if it contains <90% of the subjects in the FAS. Patients who meet any of the following criteria will be excluded from the PP analysis set

- Subjects randomized to chemotherapy or avelumab but received the incorrect active treatment (subjects who receive no active treatment are not excluded).

- Subject had an overall tumor response of PD per independent radiologist at completion of Induction Phase and was subsequently randomized.
- Subject did not have histologically confirmed unresectable locally advanced or metastatic adenocarcinoma of the stomach or GEJ as per inclusion criterion 5.
- Subject had 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 (CTLA-4) as per exclusion criterion 1.
- Subject received concurrent anticancer treatment as per exclusion criterion 2.
- Subject received prior chemotherapy for unresectable locally advanced or metastatic adenocarcinoma of the stomach or GEJ as per exclusion criterion 3.
- Subject has brain metastases not meeting the exceptions laid out in the study protocol, as per exclusion criterion 7.

Safety for Induction Phase (Safety-Induction) Analysis Set

The Safety-Induction analysis set will include all patients who were enrolled into the study (i.e., signed informed consent) and received at least one dose of Induction Phase chemotherapy.

Safety for Maintenance Phase (Safety-Maintenance) Analysis Set

The Safety-Maintenance analysis set will include all subjects randomized to receive active study medication (avelumab or chemotherapy) who received at least one dose of active study medication during the Maintenance Phase and subjects randomized to the control arm who are deemed not eligible to further chemotherapy and are planned per Interactive Web Response System (IWRS) to receive best supportive care (BSC) only, with no additional active treatment.

Subjects will be classified according to the treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period, in which case subjects will be classified according to the first study treatment received. In the case a subject randomized to receive active study medication receives no active study medication, that subject will be excluded from the Safety-Maintenance Analysis Set.

HRQoL Analysis Set

The Health-Related Quality of Life (HRQoL) analysis set is a subset of the FAS and includes all FAS subjects who meet both of the following criteria:

- Maintenance Phase baseline HRQoL assessment completed
- At least one HRQoL assessment completed after start of Maintenance Phase study treatment (or after Maintenance Phase Week 1 Day 1 visit for subjects who did not receive active study treatment).

Re-baseline HRQoL assessment occurs following completion of Induction Phase chemotherapy treatment at the Re-baseline visit.

Maintenance phase HRQoL assessment is defined as HRQoL questionnaires completed after the first dose of Maintenance Phase study treatment for subjects who received avelumab or continuation of Induction phase chemotherapy treatment, and questionnaires completed after date of randomization for subjects on BSC only.

CCI

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

Table 1 Statistical Analyses by Analysis Set

Analysis	Analysis Set						
	FAS	PD-L1+	PP	Safety-Induction	Safety-Maintenance	HRQoL	CCI
Demographic Characteristics	✓	✓		✓	✓	✓	
Other Baseline Characteristics (Medical History, Disease History, Prior Anti-Cancer Therapies, Re-baseline RECIST response)	✓	✓		✓	✓	✓	
Prior and Concomitant Therapies/Procedures	✓	✓		✓	✓		
Compliance and Exposure				✓	✓		
Efficacy: Primary Endpoint (OS)	✓	✓	✓				
Efficacy: Key Secondary Endpoints: (PFS, BOR)	✓	✓	✓			✓ (BOR only)	
Efficacy: Other Secondary Endpoint: HRQoL		✓				✓	
CCI							
Safety Endpoints (See Section 16 for additional details)				✓	✓		
PK							✓

[redacted]

Subgroup Analysis Set

Subgroup analyses will be performed on primary and key secondary efficacy endpoints based on the subgroups as defined below.

In order to include baseline variables in the Cox proportional hazards model for OS and PFS multivariate analysis, and the logistic model for BOR multivariate analysis, indicator variables will be defined for each subgroup category except for the reference category, against which all other categories are compared. In case a low number of subjects (i.e. < 5% of the randomized population) is observed in a given subgroup category, categories may be pooled. Subgroup analyses will not be performed on or include any subjects with missing values for a given subgroup.

In addition, primary and key secondary endpoint tables will be repeated for the group of subjects who are treated in Japan. Japan-specific outputs will follow the format of the overall primary/secondary end-point tables and will display one-sided p-values. These tables will omit the geographically-stratified analysis.

The parameterization used for the primary analysis will be updated and finalized at the Data Review Meeting, at the latest, and documented in an amendment to this SAP, should it differ from the following definitions.

The subgroups are as follows:

- PD-L1 Assay Status – Tumor cells
  - < 1% (Reference)
  - ≥1%
- Age Group 1
  - Age at randomization < 65 years (Reference)
  - Age at randomization ≥ 65 years
- Age Group 2
  - Age at randomization < 75 years (Reference)
  - Age at randomization ≥ 75 years
- Gender
  - Male (Reference)
  - Female
- Race
  - Caucasian / White (Reference)
  - Asian

  - Black/ African American
  - Other
- Pooled Geographical Region:
  - Asia
  - Non-Asia (Reference)
- ECOG PS at Maintenance Phase baseline
  - ECOG PS 0
  - ECOG PS 1 (Reference)
- Site of primary tumor
  - Stomach
  - Gastro-esophageal Junction (GEJ) (Reference)
- Regimen of induction chemotherapy
  - Oxaliplatin + 5-FU (Reference)
  - Oxaliplatin + Capecitabine
- Re-Baseline Response per Independent Radiologist
  - CR, PR (Reference)
  - SD, non-CR/non-PD
- Microsatellite Instability Biomarker Status
  - Stable (Reference)
  - High
- Presence of Liver Metastasis
  - Diagnosis of metastatic liver
  - No diagnosis of metastatic liver (reference)
- Presence of Peritoneal Carcinomatosis
  - Diagnosis of peritoneal carcinomatosis
  - No diagnosis of peritoneal carcinomatosis (reference)
- Prior Gastrectomy
  - Prior gastrectomy
  - No prior gastrectomy (reference)

- Number of Metastatic Sites
  - 0 (reference)
  - 1
  - 2
  - ≥ 3

10 General Specifications for Statistical Analyses

For the Induction Phase, all the analyses performed will be presented by the two chemotherapy arms and overall, unless otherwise specified. For the Maintenance Phase, demographic and baseline characteristics, as well as all the safety and efficacy endpoint analyses will be summarized by avelumab arm and the combined chemotherapy/BSC arm.

All statistical analyses will be performed using SAS® Version 9.4 or higher.

Data handling after cut-off date:

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

Pooling of centers:

In order to provide overall estimates of treatment effects, data will be pooled across trial centers. The "center" factor will not be considered in statistical models or for subgroup analyses due to the high number of participating centers and the small number of subjects expected to be 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 15.1 and 15.2. Section 7 describes the procedures for controlling the overall type I error rate in the study. The statistical tests to compare treatment arms on other secondary, CCI , and safety analyses CCI

Confidence intervals will be two-sided with a confidence probability of 95%, unless otherwise specified.

Presentation of continuous and qualitative variables:

Continuous variables will be summarized using descriptive statistics i.e., number of non-missing values and number of missing values, [i.e., n (missing)], mean, median, standard deviation (StDev),

minimum, maximum and first and third quartile (Q1 and Q3). As per general reporting conventions, mean, median, Q1, and Q2 will be displayed with one more decimal place than the raw data; StDev will be displayed using two more decimal places than the raw data. Percentages will be reported to one decimal place. Rounding will be performed to closest integer/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 frequency counts and percentages. Unless otherwise stated, the calculation of proportions/percentages for a given qualitative variable will include subjects with missing values in the denominator. When applicable, counts of missing observations for a qualitative variable will be presented on summary tables.

In case the analysis refers only to certain visits, percentages will be based on the number of subjects who had some study procedure performed at that visit, unless otherwise specified.

Study Treatment Phases:

The Induction Phase refers to the open-label chemotherapy treatment phase during which subjects receive chemotherapy comprised of oxaliplatin and either 5-fluorouracil (5-FU) or capecitabine. The induction treatment phase is defined as the period from the date of the first administration of induction chemotherapy through the date of the Induction Phase End-of-Treatment visit or the day before randomization.

The Maintenance Phase refers to the study treatment period in which randomized subjects receive their study treatment: either avelumab or the treatment prespecified by the physician (either continuation of the Induction Phase chemotherapy regimen, or BSC). The maintenance treatment phase is defined as the time from randomization to the Maintenance Phase End-of-Treatment visit.

Table 2 **Definition of Reference Date by Treatment Phase**

Treatment Phase	Reference Start Date	Reference End Date
Induction Phase	First dose of induction chemotherapy treatment	Date of end of study or death date, whichever comes later, for subjects who were not randomized; Randomization Date – 1 day for subject who were randomized.
Maintenance Phase	Randomization	Date of end of study or death date, whichever comes later

Listings will include all subject data as collected in the clinical database but data corresponding to different treatment phases will be presented separately.

Re-Baseline Tumor Scan:

Per the study protocol, following completion of Induction Phase chemotherapy treatment, re-evaluation of inclusion/exclusion criteria, safety parameters, and HRQoL assessments is to be

performed at the Re-Baseline visit, which should occur within the 10-day Re-baseline period. Additionally, a Re-Baseline tumor scan must be performed in order to assess eligibility for randomization into the Maintenance Phase (see Section 5) and to establish the "baseline" state to which Maintenance Phase tumor scans shall be compared, when assessing overall response.

The Re-Baseline tumor scan is the scan evaluated by the independent radiologist to determine the subject's eligibility for randomization into the Maintenance Phase. This scan is clearly identified in the data transfer from the independent radiologist. The investigator assessment of this scan will be the baseline for RECIST v1.1 assessments during the Maintenance Phase.

Tumor Response for Efficacy Analyses:

For all endpoint analyses that employ use of tumor response, the tumor response will be determined according to RECIST 1.1, as per investigator assessment, and is measured relative to the Re-Baseline scan.

Definition of baseline:

For the Induction Phase, baseline is defined as the last measurement taken prior to the first dose of open-label chemotherapy treatment during the Induction Phase.

For the Maintenance Phase, baseline is defined differently based on analysis type and randomized treatment. For subjects randomized to receive active study medication (avelumab or continuation of Induction chemotherapy), baseline for safety analyses is defined as the last measurement taken prior to the first dose of randomized treatment. For subjects randomized to receive BSC-only, baseline for safety analyses is defined as the last measurement taken on or before the Maintenance Phase Week 1 Day 1 visit. For all subjects, baseline for efficacy analyses is defined as the last measurement prior to randomization. Baseline for heart rate and QTc assessments will be derived from the visit where both heart rate and QT are not missing.

If an assessment is planned to be performed on Week 1, Day 1 per the protocol and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was performed prior to study treatment administration if assessment time point is not collected or is missing. If assessment time points are collected, the observed time point will be used to determine pre-dose on study day 1 for baseline calculation. Unscheduled assessments will be used in the determination of baseline. However, if time is missing, an unscheduled assessment on study day 1 will be considered to have been obtained after study treatment administration.

If both central and local labs are collected in a study, the baseline will be derived based only on the central lab collected data.

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 Maintenance Phase baseline in the context of HRQoL assessments:

Maintenance Phase baseline for HRQoL assessments is defined as the last available assessment conducted prior to or on the date of randomization, occurring no more than 28 days prior to randomization. If no such assessment exists, Maintenance Phase baseline will be the last available assessment after randomization and prior to the first dose of Maintenance Phase treatment.

Definition of Maintenance Phase baseline in the context of ECOG PS assessments:

Maintenance Phase baseline for ECOG PS assessments is defined as the last available assessment conducted prior to or on the date of randomization, occurring no more than 14 days prior to randomization. If no such assessment exists, Maintenance Phase baseline will be the last available assessment conducted after randomization and on or prior to the Maintenance Phase treatment start date.

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 Induction Phase Study Day:

Induction Phase Study Day is defined relative to the date of first administration of induction chemotherapy. Induction Phase Study Day 1 corresponds to the day of first administration of induction chemotherapy and the day before is defined as Induction Phase Study Day -1 (i.e., no Induction Phase Study Day 0 is defined).

Definition of Maintenance Phase Study Day:

Maintenance Phase study day is defined relative to the date randomization into Maintenance Phase. Maintenance Phase study day 1 corresponds to the day of randomization and the day before is defined as Maintenance Phase Study day -1 (i.e., no Maintenance Phase study day 0 is defined).

Definition of study treatment, in the context of subjects randomized to receive BSC-only:

Subjects randomized to receive BSC-only will be included in the overall counts for the control treatment arm, unless otherwise stated. "Study treatment" for subjects randomized to receive BSC-only, is considered to begin on the date of the Week 1 Day 1 assessment visit and continue until the

date given on the "End of Assessment" eCRF page, the date given on the "Study Termination" eCRF page, start day of new anti-cancer drug therapy – 1 day, or date of death, whichever comes first.

Definition of treatment day:

Treatment day is defined relative to the start date of study treatment within a given Treatment Phase. Treatment Day 1 is defined as the first day study treatment within a given Phase; Treatment Day -1 is defined as the day prior to the start of the given treatment. Treatment Day 0 is not defined. For example, for the Induction Phase, Treatment Day 1 is defined as the first day of Induction Phase chemotherapy and the day before is defined as Treatment Day -1. Similarly, for subjects randomized to active treatment during Maintenance Phase, Treatment Day 1 is defined as the day the first dose of Maintenance Phase study drug was given. For subjects randomized to receive BSC-only during Maintenance Phase, Treatment Day 1 is defined as the date of the Week 1 Day 1 assessment visit.

Definition of On-Treatment period:

For the Induction Phase, the on-Treatment period is defined as the time from the first dose of induction chemotherapy treatment through the earliest of (the date of the last dose of Induction chemotherapy + 30 days, start day of new anti-cancer drug therapy – 1 day, date of Maintenance Phase study treatment start – 1 day).

For subjects randomized to the avelumab arm or the control arm and receive chemotherapy, the on-treatment period is defined as the time from the first dose of randomized treatment through the earliest of (last dose of study treatment + 30 days, start day of subsequent anti-cancer drug therapy – 1 day). For subjects randomized to the control arm and receive BSC-only, study treatment is considered to start on the day of the Week 1 Day 1 assessment visit and end on the earliest date given from among the following: date given on the Study Termination eCRF page and date of death. Thus, the on-treatment period for these subjects is defined as the time from the Week 1 Day 1 assessment visit to the earlier of (end of assessment date + 30 days, start of new anti-cancer drug therapy – 1 day).

The date of subsequent anti-cancer drug therapy is derived as outlined in Section 13.2.

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) for use in Induction Phase demographics:
  - (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) / 12
- 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.

- Age (years) for use in Maintenance Phase demographics:
  - (date of randomization - date of birth + 1) / 365.25
  - In case of missing day, day only: Age [years]: (year/month of randomization – year/month of birth) / 12
  - In case only year of birth is given: Age [years]: (year of randomization - 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)$^{0.5}$

Reporting conventions are detailed in the sub-section titled "Presentation of continuous and qualitative variables", earlier in this section.

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. 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 HRQoL data and safety endpoints (laboratory, ECG and vital signs), only those scheduled, per-protocol visits/time points that have at least 5 subjects in both treatment arms will be included in the summary tables and figures. The exception is Discontinuation/End-of-Treatment visit which will be included in the summary table/figure regardless of the number of subjects who may have completed visit. Tables for safety end-points will only summarize on-treatment records.

Missing data and imputation rules:

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

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

Missing statistics, e.g. when they cannot be calculated, will be presented as 'nd' or 'n/a'. For example, if N=1, the measure of variability (StDev) cannot be computed and would be presented as 'nd'.

For laboratory assessments indicating that the sample tested below the lower limit of quantitation (LLOQ) for the analyte in question, half the lower limit of quantitation (½*LLOQ) will be used as the analysis value. For laboratory assessments testing above the upper limit of quantitation (ULOQ), the analysis value will be imputed as {ULOQ + 1*(the level of precision of ULOQ)}. For example, if an assessment tests at a value ">7.0", the level of precision is 0.1 and the analysis value will be 7.1. The imputed analysis values will be used in all analyses (e.g. summary statistics) and will be used for determination of toxicity grading. Listings will present the standardized unit values as provided in the raw data, indicating the assessment value was not quantifiable (e.g. ">7.0").

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/JAN/2015.
- If both the day and month of the AE onset date are missing but the onset year is equal to the start of 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 Induction Phase treatment.
- If the day of medication date is missing, but the month and year are equal to the start of Induction Phase treatment, then the medication date will be replaced by the start of Induction Phase treatment. For example, if the medication start date is --/JAN/2015, and Induction Phase 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 Induction Phase treatment, then the medication date will be replaced by the start of Induction Phase treatment. For example, if the medication start date is --/---/2014, and Induction Phase 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.
- In the case that the imputed start date occurs after the medication end date, the medication end date will be used as the start date.
- 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 dates for start date of subsequent anti-cancer therapy (drug therapy, radiation, surgery) will be imputed as follows and will be used for determining censoring dates for efficacy analyses and in the derivation of the end of on-treatment period.

If only day is missing, it will be imputed as the last day of the month unless the end date of subsequent anti-cancer therapy is before that date. In that case, the incomplete anticancer 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

- For the Maintenance Phase, if the study medication start date is completely or partially missing, it is assumed that the first dose of trial medication is given on the randomization date. The randomization date will replace any incomplete or missing dates of the first dose of Maintenance Phase study medications.
- If the date of the last dose of study drug is unknown or partially missing, it will be imputed as follows:
  - If the last date of the last study drug administration is incomplete, the date will be taken from the treatment termination eCRF page, if available.
  - 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's treatment is considered to be ongoing and the last dosing date on or prior to the cut-off date for the analysis will be used as the last dosing date
  - If the last date of study drug administration 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 the last dose date is imputed as follows:
    - If day is missing and 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)], then day will be assigned to the last day of the month
    - If only Year is available and Year < [Year of min (EOT date, death date)], then date will be set to 31DECYYYY
    - In all other cases, the date will be imputed as min (EOT date, death date)

Date of last contact

The last contact date will be the latest complete date among the following dates collected on the eCRF:

- All subject assessment dates (blood draws (laboratory, CCI), 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" -
- Study drug start and end dates
- Randomization date
- Date of discontinuation from the "Study Termination" eCRF page (not use if reason for discontinuation is lost to follow-up)

Death date

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

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

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 complete 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.

11 Trial Subjects

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

11.1 Disposition of Subjects and Discontinuations

Subject disposition summary will be created for the Induction Phase and the Maintenance Phase separately.

For the Induction Phase, the following will be summarized overall and by Induction Phase study treatment for all screened subjects:

- Number of subjects screened
- Number of subjects screen-failed or not treated during Induction Phase
- Number of subjects discontinued during the Induction Phase
- Number and percentage of subjects who completed the Induction Phase treatment
- Primary reasons for treatment discontinuation during the Induction Phase
- Number of subjects completed Induction Phase (overall only)
- Number of subjects completed Induction Phase but not randomized (overall only) and primary reason

Subjects who discontinue Induction Phase chemotherapy due to completion of regimen/ randomization to the avelumab or BSC-Only arms will not have the reason presented in the table.

For the Maintenance Phase, the following will be summarized overall and by treatment arm (as randomized) for all randomized subjects:

- Number of subjects randomized
- Number and percentage of subjects randomized but not treated (among those randomized to chemotherapy or avelumab)
- Number and percentage of subjects randomized to the control arm who received BSC-Only
- Number and percentage of randomized subjects still on study treatment
- Number and percentage of randomized subjects who discontinued all study drugs
- Number and percentage of subjects who discontinued any one study treatment and primary reason by study drug, i.e. avelumab, oxaliplatin, 5-FU, capecitabine.
- Number and percentage of subjects who discontinued treatment but are still in follow-up

- Number and percentage of subjects who discontinued the study and primary reasons
- Number and percentage of subjects who re-initiated avelumab treatment
- Number and percentage of subjects who discontinued avelumab treatment after re-initiation and primary reasons
- Number and percentage of subjects in the following analysis sets:
  - Safety-Induction
  - FAS
  - PP
  - Safety-Maintenance
  - HRQoL
  - PD-L1+
- The number of subjects randomized by region and country
- Cross tabulation: subjects randomized (avelumab, oxaliplatin + 5-FU, oxaliplatin + capecitabine, or BSC-Only, Total) versus subjects treated (avelumab, oxaliplatin + 5-FU, oxaliplatin + capecitabine, BSC-Only, or Not Treated).

In addition, the following will be summarized:

- Number and percentage of subjects in each Analysis Set will be broken down as follows:
  - Overall
  - By region (EEA (required by EudraCT), Eastern Europe, Western Europe, North America, Latin America, Asia, Australasia, Africa, and the Middle East)
  - By country within region
  - By center within country.

11.2 Protocol Deviations

Protocol deviations will be listed and summarized by treatment group based on the 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, but are not limited to:

- Subjects that are enrolled in the study despite not satisfying the inclusion/exclusion criteria;
- Subjects that develop withdrawal criteria while on the study but are not withdrawn;
- Subjects that receive the wrong treatment or an incorrect dose;
- Subjects that receive an excluded concomitant medication;

- Deviation from GCP.

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

Reporting of subjects randomized more than once

If a subject is randomized more than once with deviating information on the randomization factors, the following approach will be taken for the analysis:

- There will only be one subject ID in the analysis dataset.
- The subject's clinical data captured on the eCRF will be used for all analyses with the exception of randomization strata. According to the intention-to-treat principle, date of informed consent, treatment assignment, randomization strata, and randomization date will be taken from first randomization and transferred to subject ID on-treatment, irrespective of the treatment assignment at subsequent randomization.
- For safety analyses, the subject will be assigned to the group as treated.

All important protocol deviations should be documented in CDISC datasets whether identified through sites monitoring, medical review or programming. Clinically Important Protocol Deviations (CIPDs) are specified in Appendix I. Occurrences of CIPDs will be presented in the summary table and data listing.

12 Demographics and Other Baseline Characteristics

12.1 Demographics

Analysis sets: FAS, PD-L1+, Safety-Induction, Safety-Maintenance, HRQoL

Demographic characteristics at Screening will be summarized using the following information from the Screening/Baseline Visit eCRF pages. For the Safety-Induction version of the table, summaries will be displayed overall and by Induction-Phase Chemotherapy. For the FAS, HRQoL, and Safety-Maintenance versions of the table, summaries will be displayed overall and by Maintenance Phase treatment arm.

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 years
- Pooled Geographical Region:
  - North America
  - Europe
  - Asia
  - Rest of the World (Australasia, Latin America, Africa, and/or the Middle East 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
  - Middle East
  - Australasia
  - Asia
  - Africa
- EEA (Yes/No)
- Eastern Cooperative Oncology Group Performance Status (ECOG 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.

[redacted]

Demographic characteristics at Re-Baseline will be summarized overall and by treatment arm for FAS, PD-L1, HRQoL, and Safety-Maintenance analysis sets. ECOG PS, and physical measurements, will be summarized using updated values collected at the Re-Baseline visit; age will be re-calculated using randomization date as a reference.

A listing of demographics and baseline characteristics will include the following information: subject identifier, Induction and Maintenance Phase treatment groups, sex, race, ethnicity, height (cm), and each of the following as measured at Screening and Re-Baseline: age, weight (kg), BMI (kg/m^2), BSA (m^2), and ECOG PS.

12.2 Medical History

Analysis sets: FAS, PD-L1+, Safety-Induction, Safety-Maintenance, HRQoL

Medical history reported at the time of the Screening procedures will be coded using the most current available version of Medical Dictionary for Regulatory Activities (MedDRA) and will summarize data collected on the "Medical History" eCRF page. Medical history will be summarized by the number and percentage of subjects who report specific MedDRA preferred term (PT) events, grouped and additionally summarized by MedDRA primary system organ class (SOC). Each subject will be counted only once within each applicable PT and SOC.

Medical history will be displayed in frequency tables, ordered by decreasing overall counts for primary SOC and then, within SOC, by decreasing overall frequency of PT. SOC/PT's with the same overall frequency will be presented in alphabetical order. Treatment arm will be based on the analysis set used.

12.3 Other Baseline Characteristics

12.3.1 Disease Characteristics

Analysis sets: FAS, PD-L1+, Safety-Induction, Safety-Maintenance, HRQoL

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

- Site of primary tumor
- Sub-site of tumor
- Tumor histopathologic / cytologic type
- Time since initial cancer diagnosis (months), defined as (date of first dose of Induction Phase chemotherapy – date of initial cancer diagnosis)/30.4375

- Time since documented locally advanced or metastatic disease (months), defined as (date of first dose of Induction Phase chemotherapy – date of documented locally advanced or metastatic disease)/30.4375
- TNM classification at initial diagnosis
- TNM classification at study entry
- Tumor response at Re-Baseline

Baseline characteristics with respect to vital signs, physical examinations, ECGs, and hematology/biochemistry will be part of Section 16 (Safety Evaluation).

12.3.2 Prior Anti-Cancer Therapies

Analysis sets: FAS, PD-L1+, Safety-Induction, Safety-Maintenance, HRQoL

Anti-cancer therapies administered prior to enrollment in the Induction Phase 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 and percentage of subjects in each of the following anti-cancer therapy categories will be tabulated:

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

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

- at least one prior anti-cancer drug therapy
- any prior anti-cancer therapy regimens: missing / 1 / 2 / ≥3
- Type of prior anti-cancer therapy
- Type of prior anti-cancer therapy: Cytotoxic therapy / Endocrine therapy / Monoclonal antibody therapy / Small molecules / Immunotherapy / Other
- Intent of Therapy: Neoadjuvant / Adjuvant / Other
- Best response: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Non-CR/Non-PD / Progressive Disease (PD) / Not assessable / Unknown. Best response is derived from the last treatment regimen

The prior anti-cancer drugs will also be summarized based on the number and percentage of subjects by the drug class and preferred term. A subject will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. The summary will be sorted on decreasing frequency of drug class and decreasing frequency of drug name within a given drug class. In case of equal frequency regarding drug class or drug name, alphabetical order will be used.

Listings of prior anti-cancer treatments and procedures from randomized subjects will also be provided as follows. These will include analysis sets a subject belongs to, the subject identification number, relative day of prior anti-cancer therapy stop date to induction chemotherapy treatment start date (- xx days), and all the relevant collected data-fields on the corresponding eCRF pages. Listings will be displayed by randomized treatment arm.

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

CCI [redacted]

[redacted]

[redacted]

- [redacted]
- CCI [redacted]
- CCI [redacted]
- [redacted]

[redacted]

CCI [redacted]

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

CCI [redacted]

[redacted]

CCI [redacted]

13 Previous or Concomitant Medications/Procedures

Analysis sets: FAS, PD-L1+, Safety-Induction, Safety-Maintenance

13.1 Prior, Concomitant, and Subsequent Medications/Procedures

Premedications are medications that are given prior to the administration of a study medication, as required per the study protocol, during either the Induction or Maintenance Phases.

Concomitant medications are medications, other than study medications and premedications, which are started prior to first dose date of study treatment and continued during the on-treatment period, or are started during the on-treatment period for Induction Phase and Maintenance Phase, respectively.

Concomitant procedures are medical interventions/procedures which occur during the on-treatment period.

Prior medications are medications, other than study medications and premedications, which are started before first dose date of study treatment during the Induction Phase.

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

Prior and concomitant medications will be summarized utilizing data collected on the "Concomitant Medications Details" and "Prior Anti-Cancer Drug Therapies Details" eCRF pages. Premedications will be summarized using the records provided on the "Premedication Details" eCRF page. Concomitant procedures are recorded on the "Concomitant Procedures Details" eCRF page.

Summaries of prior, concomitant, and premedications will include the number and percentage of subjects by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term, as classified by the latest available WHO Drug dictionary. The summary will be presented within treatment arm and overall. A subject will be counted only once within a given drug class and within

[redacted]

a given drug name, even if he/she received the same medication at different times. If any prior or concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes. The summary tables will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. In case any specific medication does not have ATC classification level 2 coded term, it will be summarized under "Unavailable ATC classification" category. A listing of concomitant medications will be created with the relevant information collected on the "Concomitant Medications Details" eCRF page for the Safety-Maintenance analysis set.

Pre-medications for study drug will also be presented in a listing.

A summary of subjects who underwent concurrent procedures, and the reasons for these procedures, will be summarized by treatment arm and overall. This table will summarize data collected on the eCRF page "Concomitant Procedures Details". Additional details provided on the eCRF page will be presented in a listing.

13.2 Subsequent Anti-Cancer Therapies/Procedures

Analysis sets: FAS, PD-L1+

"Subsequent anti-cancer therapy" is defined separately, relative to the Induction and Maintenance Phases. Subsequent anti-cancer therapy relative to the Induction Phase applies only to subjects who were not randomized into the Maintenance Phase. Subsequent anti-cancer therapy relative to the Maintenance Phase is only applicable to randomized subjects (including BSC-only treatment, as randomized).

Anti-cancer treatment after discontinuation will be provided with data retrieved from "Anti-Cancer Treatment after Discontinuation Details", "Radiotherapy after Discontinuation", and "Surgery after Discontinuation" eCRF pages. The listing will include all subjects in the Safety-Induction Analysis Set.

The earliest date of start of new anti-cancer drug therapy, defined as therapies recorded on the "Anti-Cancer Treatment after Discontinuation Details" eCRF page, will be used for the definition of the on-treatment period. The earliest date of start of any new anti-cancer therapy (including drug therapy, radiotherapy, and surgery) will be used for censoring for efficacy analyses.

Number and percentage of subjects with any anti-cancer treatment after discontinuation of the Maintenance Phase (including drug therapy, radiotherapy, and surgery) will be tabulated overall and by type of therapy based on data collected from the aforementioned eCRF pages including classification as cytotoxic therapy / endocrine therapy / monoclonal antibodies therapy / small molecules / immunotherapy / Other.

Summary statistics will be created for best response across all anti-cancer drug therapies taken prior to the Maintenance Phase for the Safety-Induction Analysis Set and also after the start of the Maintenance Phase for the Safety Maintenance Analysis Set 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 Maintenance Phase treatment discontinuation, the best overall response among all subsequent anti-cancer drug therapies will be selected for inclusion in the summary.

Summary of subsequent anti-cancer drug therapies, relative to the Maintenance Phase, will include the number and percentage of subjects receiving each specific therapy, broken down by ATC Classification level 2 and preferred term. This is the same approach as is used for summarizing prior and concomitant medications.

14 Treatment Compliance and Exposure

Analysis sets: Safety-Induction Analysis Set, Safety-Maintenance Analysis Set

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

All subjects enrolled in this trial will start with a 12-week chemotherapy Induction Phase during which subjects receive oxaliplatin and either 5-FU or capecitabine (according to the dosing schedule in Section 6.2.2 and Table 2 in the CTP).

Subjects who experience a CR, PR, SD, or Non-CR/ Non-PD at the end of Induction Phase will be randomized to the Maintenance Phase.

Best Supportive Care (BSC): Best supportive care is defined as: treatment administered with the intent to maximize quality of life without a specific antineoplastic regimen.

Maintenance Phase treatment arms are as follows:

- **Avelumab arm:** Subjects randomized to the avelumab 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).
- **Chemotherapy/BSC arm:** Following the completion of the Induction Phase, site investigators will determine whether subjects are eligible to receive further chemotherapy during the Maintenance Phase. Subjects randomized to the Chemotherapy/BSC-only arm will receive the following:
  - **Oxaliplatin + 5-FU:** If subjects are deemed eligible to receive further chemotherapy and received oxaliplatin + 5-FU in the Induction Phase, they will receive the same regimen as they received in the Induction Phase, e.g. Oxaliplatin at 85 mg/m^2 IV on Day 1 plus leucovorin 200 mg/m^2 IV on Day 1 followed by 5-FU at 2600 mg/m^2 IV continuous infusion over 24 hours on Day 1, once every 2 weeks. Please refer to the CTP for alternative dosage/administration schedules.

- **Oxaliplatin + Capecitabine arm:** If subjects are deemed eligible to receive further chemotherapy and received oxaliplatin + capecitabine in the Induction Phase, they will receive the same regimen as they received in the Induction Phase, i.e. oxaliplatin at 130 mg/ m^2 IV on Day 1 with capecitabine at 1000 mg/m^2, twice daily for 2 weeks followed by a 1-week rest period once every 3 weeks.
- **BSC-Only:** Subjects who are not deemed eligible to receive further chemotherapy will receive BSC alone with no active therapy.

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

- Avelumab: total dose administered/weight [mg/kg]
- Chemotherapy treatment: total dose administered/BSA [mg/m^2]

where weight is taken from the last non-missing value on or prior to the day of each administration cycle of trial medication, and BSA is derived from the height value at Screening and the last non-missing weight value on or prior to the day of each administration cycle of trial medication.

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

Treatment Duration

For subjects on Avelumab arm, the duration of treatment (weeks) during the study is defined as:

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

For subjects on chemotherapy arm (oxaliplatin + 5-FU or oxaliplatin + capecitabine), the duration of treatment (weeks) is defined as follows for both the Induction Phase and the Maintenance Phase, while the first dose and last dose date are relatively to the first and last dose date in the Induction and Maintenance Phase, respectively.

- Oxaliplatin + 5-FU arm:
  - Treatment duration for oxaliplatin (weeks) = (last dose date – first dose date + 14) / 7
  - Treatment duration for 5-FU (weeks) = (last dose date – first dose date + 14) / 7
- Oxaliplatin + capecitabine arm:
  - Treatment duration for oxaliplatin (weeks) = (last dose date – first dose date + 21) / 7

  Since capecitabine is taken as a tablet twice daily for 2 weeks followed by a 1-week rest period once every 3 weeks, treatment duration is calculated as follows:
  - Treatment duration of capecitabine (weeks) = (last dose date – first dose date + 7) / 7

Duration of Regimen is defined for subjects who receive both oxaliplatin and 5-FU or capecitabine. For these subjects, Duration of Regimen is equivalent to the treatment duration of whichever chemotherapy is discontinued first. For example, if a subject receives oxaliplatin and 5-

FU and discontinues oxaliplatin before discontinuing 5-FU, Duration of Regimen will be equal to the treatment duration for oxaliplatin. In the case a subject discontinues oxaliplatin and capecitabine on the same day (and the subject does not receive further capecitabine doses) the treatment duration for oxaliplatin will be selected as the Duration of Regimen for this subject.

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 oxaliplatin, 5-FU, and capecitabine 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 (DI) will be calculated for each subject across all cycles. Each cycle is defined by a 2-week period for avelumab and oxaliplatin + 5-FU and by a 3-week period for Oxaliplatin + Capecitabine.

The dose intensity of avelumab (mg/kg/cycle), oxaliplatin (mg/m^2/cycle), 5-FU (mg/m^2/cycle), and capecitabine (mg/m^2/cycle) is defined as

$$\mathrm{DI} = \left(\frac{\text{cumulative dose}}{\frac{\text{duration of therapy (in weeks)}}{\text{number of weeks per cycle}}} \right)$$

Relative Dose Intensity

The relative dose intensity (RDI) is defined as the actual dose intensity divided by the planned dose per cycle and expressed in %.

The summary of treatment exposure and compliance will be calculated for both Induction Phase and Maintenance Phase, and will include the following information:

- Duration of therapy (weeks)
- Total number of infusions received (avelumab and IV chemotherapy)
- Total number of Capecitabine doses received

The following will be included, broken down by separate chemotherapy treatment in the chemotherapy arm:

- 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%, 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 Reductions

Dose reduction will be calculated for the chemotherapy arm only in the Maintenance Phase. Per study protocol, there will be no dose reductions for the avelumab treatment arm. A dose reduction is defined as an actual non-zero cumulative dose < 90% of the planned cumulative dose for a given treatment cycle. For subjects receiving Capecitabine or 5-FU as a push followed by a drip, any skipped doses within the cycle will be counted as part of the planned cumulative dose for the cycle (i.e. will contribute to the denominator of the calculation). Dose reductions are counted on a per-cycle basis and do not consider whether previous doses were reduced. 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 chemotherapy/BSC arm.

Dose Delays

Dose delays will be calculated for Maintenance Phase IV study drugs only. 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 considered as a "No delay".

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.

Dose Interruptions

An interruption is defined as an IV medication which is not administered (i.e. 0 dose) on one or more planned dosing days. Interruptions are identified on the [Study Drug] Administration eCRF pages (e.g. "Has the subject received a dose of Oxaliplatin during this visit?"). Please note that dose interruptions, as defined here, are not the same as drug interruptions captured on the Treatment Administration Modification Details eCRF page. The number and percentage of patients with dose

interruptions, the corresponding reasons, the number of interruptions observed, and the length of interruption will be summarized. Length of dose interruptions is measured in cycles.

Infusion Rate Reductions

Infusion rate reduction will be calculated for the Avelumab treatment arm in the Maintenance Phase only. Infusion rate reductions 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 a listing.

15 Endpoint Evaluation

A hierarchical test strategy will be used to control the overall type I error rate for the primary endpoint of OS, between analyses in all randomized subjects, and PD-L1+ subjects and key secondary endpoints of PFS and BOR in all randomized subjects in the following hierarchical order:

1: H_{01}: OS in the FAS Analysis Set and H_{02}: PD-L1+ Analysis Set in the Maintenance Phase

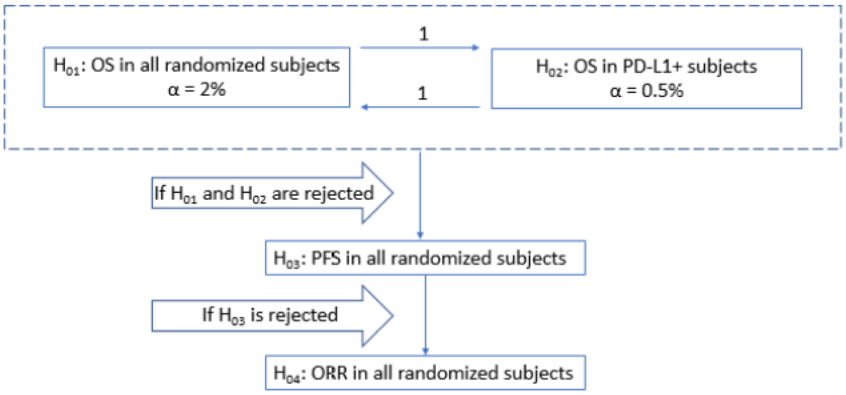
2: H_{03}: PFS in the FAS Analysis Set in the Maintenance Phase

3: H_{04}: BOR in the FAS Analysis Set in the Maintenance Phase

The overall 2.5% (1-sided) type I error will be split between OS in all randomized subjects and PD-L1+ subjects with 2.0% (1-sided) and 0.5% (1-sided), respectively. One interim analysis is planned for OS in all randomized subjects. A Lan-DeMets alpha spending function for O'Brien-Fleming boundaries will be used to control the overall type I error rate of OS in all randomized subjects at 2% between the interim and primary analysis (see Section 7.2).

For the primary endpoint, a closed testing procedure with weighted Bonferroni tests, described by Hommel et al. (2), as illustrated in Figure 1, will be used for determining significance of the dual hypothesis primary testing at the primary analysis.

Figure 1 Testing strategy



At the primary analysis, if any of the two null hypotheses of OS (evaluated in all randomized subjects and in PD-L1+ subjects) is rejected, the corresponding alpha level will be recycled between both populations. The alpha recycling approach is as follows:

1. If the null hypothesis of OS in all randomized subjects is rejected at the primary analysis, the 2% (1-sided) alpha allocated to OS in all randomized subjects will be recycled to OS in PD-L1+ subjects, then 2.5% (1-sided) alpha level will be used for OS in PD-L1+ subjects.

2. If the null hypothesis of OS in PD-L1+ subjects is rejected, the 0.5% alpha allocated to OS in PD-L1+ subjects will be recycled to OS in all randomized subjects. In this case, the alpha for OS in all randomized subjects is 2.5% (1-sided) and the 1-sided boundary at the primary analysis would be recalculated by keeping the alpha level used at interim analysis unchanged. Thus, assuming the planned number of events are observed at each analysis, the 1-sided boundaries would then be 0.0072 at the interim analysis and 0.0233 at the primary analysis.

Table 3 summarizes the operating characteristics and efficacy boundaries, in the event that planned number of events are observed for OS in all randomized subjects at both the interim and primary analyses.

Table 3 **Planned Efficacy Boundaries of OS in all randomized subjects and in PD-L1+ subjects**

Look	Hypothesis Test	Info Fraction	Approx. Cum. Events of OS	Cumulative Type I Error (1-sided)	Efficacy Boundaries	
					Z-scale	p-scale (1-sided)
Interim	OS in all randomized subjects	~75%	267	0.0072	2.45	0.0072
Primary	OS in all randomized subjects (if the test for OS in PD-L1+ subjects is not significant)	100%	356	0.0200	2.10	0.0178
	OS in all randomized subjects (if the test for OS in PD-L1+ subjects is significant)	100%	356	0.0250	1.989	0.0233
	OS in PD-L1+ subjects (if the test for OS in all randomized subjects is not significant)	100%	112	0.0050	2.576	0.0050
	OS in PD-L1+ subjects (if the test for OS in all randomized subjects is significant)	100%	112	0.0250	1.960	0.0250

PFS, followed by BOR, will be tested if and only if both null hypotheses for OS in all randomized subjects and PD-L1+ subjects are rejected at the primary analysis. In that case, PFS will be tested with full 2.5% (1-sided) type I error at the primary analysis. BOR will be tested with full 2.5% (1-sided) type I error at the primary analysis if and only if the null hypothesis of PFS is rejected. All analyses including OS in PD-L1+ subjects, CCI [redacted]

No formal adjustment for multiplicity will be undertaken for non-key secondary endpoints, sensitivity analyses, or CCI [redacted] analyses.

Statistical analyses will be performed using SAS® version 9.4 or higher.

15.1 Primary Endpoint Analyses

Analysis sets: FAS, PD-L1+, PP

The primary endpoint for maintenance therapy is OS. The primary analysis sets will be the FAS Analysis set and PD-L1+ Analysis Set. The following analysis will be conducted in the same manner

for both FAS and PD-L1+ Analysis sets. For OS, the difference between the two treatment arms will be compared using a stratified, 1-sided, log-rank test. The stratification factor will be region (Asia versus non-Asia). Control of type I error and statistical significance is discussed above (Section 5). The following null hypothesis will be tested:

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

where $\lambda(t)$ represents the hazard rate (corresponding to each endpoint) at time t for treatment arms A (avelumab) and B (chemotherapy/BSC).

15.1.1 Primary Efficacy Analysis of Overall Survival

Analysis sets: FAS, PD-L1+, PP

The primary efficacy analysis of OS in all randomized subjects and in PD-L1+ subjects will be performed on the FAS and PD-L1+ analysis sets, respectively. 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 (i.e. the date of last contact, as defined in Section 10) as of the data cut-off date for the analysis.

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

For the endpoint of OS, the alpha level used for all randomized subjects and PD-L1+ subjects will be determined as specified in Sections 6, 7.27.1, 5, based on the number of events observed at the time of analysis and the significance of the test in each of the two populations.

The treatment effect will be estimated using a Cox Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), 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 group (0= chemotherapy/ BSC, 1= avelumab) and β is the unknown regression parameter. The hazard ratio of avelumab versus chemotherapy/ BSC (with chemotherapy/ BSC as reference), together with the corresponding Wald two-sided 95% confidence intervals (CIs), will be presented.

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

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median survival time with two-sided 95% CIs. In particular, the survival rate at 3, 6, 9, 12, 15, 18, 21, and 24 months will be estimated with

corresponding two-sided 95% CIs. Survival rate for additional months may be displayed, in the event that such a value is estimable at months 27 and beyond. The CIs for the median will be calculated according to Brookmeyer and Crowley (3) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (7) (conftype=loglog default option in SAS Proc LIFETEST). 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 > 14 weeks prior to the analysis cut-off date (duration of 14 weeks is based on the assessment schedule of every 12 week for survival follow-up interval + 2-week window)

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

15.1.2 Sensitivity Analyses and Model Checks

Analysis sets: FAS, PD-L1+

Sensitivity Analyses
The following sensitivity analyses will be performed to explore the robustness of the primary analysis results and will therefore be performed on both the FAS and PD-L1+ analysis sets. CCI [redacted] Unadjusted 95% CI's will be used for all sensitivity analyses of the primary endpoint. The sensitivity analyses will include the following:

- Unstratified analyses will be performed comparing OS between the treatment arms;
- If the actual number of events is 10% more than the planned number at the primary analysis and the increase is not due to the requirement for at least 18 months of follow-up, i.e. if there are > 392 events in all randomized subjects observed at the prospectively determined cut-off date at the primary analysis, the primary analysis will be repeated, using the first cut-off date at which exactly 356 events in all randomized subjects were observed in the study.
- The primary analyses will be repeated, censoring those subjects in FAS who received subsequent anti-cancer therapy at the date of the first dose of subsequent anti-cancer therapy minus 1 day.

[redacted]

- If the model checks (described later in this section) show large departures from the proportional hazards assumption, the treatment effect on OS outcomes will be additionally explored using restricted mean survival time, discussed in detail in the following subsection.

Restricted Mean Survival Time (RMST)

In the case of large departures from proportional hazards, the primary endpoint of OS will also be analyzed based on restricted mean survival time (RMST) differences (Royston and Parmar 2011 (4)). 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.

The RMST to cut-off time τ can be interpreted as the expected survival time among all patients followed to, at most, time τ. The RMST analysis hinges on the selection of the cut-off point (τ). It is noted that the cut-off point should not exceed the minimum of the largest observed time (event or censoring) for both treatment arms, in order for the RMST to be adequately estimated for both and comparison between them to be appropriate. τ should be clinically meaningful and closer to the end of the study follow-up, so that the majority of survival outcomes will be observed within the time interval. Thus, the selection of τ should ensure that the RMST evaluation does not exceed the maximum evaluable time point, but should also seek to maximize τ, in order to provide a meaningful assessment of treatment effect. To avoid arbitrary selection of the common cut-off τ (used for both treatment arms), the following analyses will be performed:

- τ_1 = minimum of (largest observed survival time for avelumab + BSC arm, largest observed survival time for chemotherapy/BSC-only arm).
- τ_2 = minimum of (largest survival event time for avelumab + BSC arm, largest survival event time for chemotherapy/BSC-only arm).

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 a 2-sided p-value will be generated.

Methods for Evaluating the Validity of Model Assumptions

Proportional Hazards Assumption:

- The proportional hazards assumption will be checked visually for the primary analysis by plotting log(-log(OS)) versus log(time), within each randomization stratum by treatment group.

- Schoenfeld residuals including a LOESS curve will be plotted to graphically investigate any 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 approximately parallel to the x-axis.

15.1.4 Subgroup Analysis of Primary Endpoint

Analysis sets: FAS, PD-L1+

Subgroup analyses will be performed on the primary endpoint for all subgroup levels defined in Section 9 "Subgroup Analysis Sets". CCI [redacted] No adjustment for multiplicity will be performed. The two treatment arms will be compared using a two-sided unstratified log rank test for each subgroup level. The unstratified HR of avelumab over chemotherapy/BSC-only and its corresponding 95% CI will be computed at each

subgroup level. In the case of a low number of subjects within a category (<5% of the randomized population), the categories may be pooled.

The HR and its corresponding 95% CI, overall and for each subgroup level will be presented in a forest plot.

To assess the heterogeneity of treatment effect on OS across the subgroup levels, two Cox regression models will be fitted for each combination of subgroup/primary endpoint. Both models will use subgroup and treatment arm as explanatory variables; the treatment-by-subgroup interaction will be included in the second.

- 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 subgroup HR (analysis subgroup level over reference category), point estimate of the regression parameter, and corresponding 95% CI of the interaction model parameter.

15.1.5 Time of Follow-up for Primary Endpoint

Analysis sets: FAS, PD-L1+

In order to compare median time of follow-up for overall survival, Kaplan Meier estimates will be calculated for both treatment groups using the following censoring rules (reverse censoring indicator):

Subjects alive or lost to follow-up	Time from randomization to last date known to be alive	No censoring
Subjects who died	Time from randomization to date of death	Censored

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with the median time of FU.

15.2 Secondary Endpoint Analyses

Based on the hierarchical testing strategy described in Section 15, at the primary analysis, PFS and BOR will be tested in primary analysis if, and only if, of hypothesis testing for OS in both all randomized subjects and PD-L1+ subjects reaches statistical significance.

15.2.1 Progression-Free Survival

Analysis sets: FAS, PP Analysis Set

PFS is defined as the time from the date of randomization to the date of the first documentation of objective progression of disease (PD) or death due to any cause, whichever occurs first. The tumor response will be determined according to RECIST 1.1 per investigator assessment and is measured relative to the Re-Baseline scan.

PFS time (in months) = (Date of PD or death or censoring - 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 a new anti-cancer therapy (of any kind) prior to an event, or for subjects with an event after two or more missing tumor assessments. Subjects who do not have a Maintenance Phase baseline tumor assessment (as defined in Section 10) 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 DR analysis are presented in Table 4.

Table 4 Outcome and Event Dates for PFS Analyses

Scenario	Date of event/censoring	Outcome
No Maintenance Phase baseline assessment [a]	Date of randomization	Censored [b]
PD or death ≤ 12 weeks after Maintenance Phase randomization or ≤ 2*(scheduled time between tumor assessments)[d]	Date of progression or death	Event
PD or death > 2*(scheduled time between tumor assessments)[d]	Date of last adequate tumor assessment	Censored
PD or death after 2 or more missing or inadequate post-baseline tumor assessments	Date of last adequate tumor assessment [c]	Censored
No PD and no death	Date of last adequate tumor assessment [c]	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] Subjects whose baseline scan is outside of the RECIST defined window (28 days prior to randomization) will be treated as if they had no baseline and will be censored on the day of randomization

[b] However, if the subject dies ≤12 weeks after randomization the death is an event with date on death date

[c] If there are no adequate post-baseline assessments prior to PD, death, or subsequent anti-cancer therapy, then the observation will be censored on the date of randomization.

[d] Per protocol, subjects are to receive tumor scans once every six weeks during the first 12 months of the Maintenance Phase and every 12 weeks thereafter.

The analysis of PFS will be analogous to that for OS time as described in Section 15.1.1. There is no interim analysis planned for PFS.

The treatment effect on PFS time will be estimated using the same approach as OS analysis. The hazard ratio of avelumab versus chemotherapy/BSC (with chemotherapy/BSC as reference) together with the corresponding Wald two-sided 95% CIs will be presented.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median survival time with two-sided 95% CIs. In particular, the progression-free survival rate at 3, 6, 9, 12, 15, 18, 21, and 24 months will be estimated with corresponding two-sided 95% CIs. The CIs for the median will be calculated using the same approach as OS analysis.

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:

- Administrative Censoring: Ongoing in the study without an event
- Non-Administrative Censoring:
  - No baseline assessment
  - No post-baseline assessment
  - Start of new anti-cancer therapy
  - Event after 2 or more missing 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;
- Two or more missed or inadequate post-baseline tumor assessments immediately preceding analysis cut-off date.

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

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.

Sensitivity Analysis

The following sensitivity analyses will be performed to explore the robustness of the primary analysis results. CCI Unadjusted 95% CI's will be used for all sensitivity analyses. The sensitivity analyses will include the following:

1. Unstratified analyses will be performed, comparing PFS time between the treatment arms;
2. Counting all PD and deaths as events regardless of missing assessments or timing of the event;
3. PFS based on investigator assessment wherein initiation of subsequent anticancer therapies is not used as a censoring reason for PFS;
4. In the case or large deviations from proportional hazards, treatment effect on PFS outcomes will be additionally explored using restricted mean survival time, discussed in detail in Section 15.1.2.

15.2.2 Best Overall Response and Objective Response

Analysis sets: FAS, PP, HRQoL Analysis Set

The confirmed BOR is defined as the best confirmed response obtained among all tumor assessment visits completed after the date of randomization until documented disease progression. Requirements for confirmation of responses is detailed below. The tumor response at each assessment visit will be determined according to RECIST 1.1, as assessed by the investigator relative to the Re-Baseline scan.

Only tumor assessments performed on or before the start of any further anti-cancer treatment will be considered in the assessment of BOR. Clinical deterioration ("progressive disease based on clinical assessment" as captured on the treatment discontinuation eCRF pages) will not be considered as documented disease progression in the context of BOR/tumor response.

Subjects without measurable disease at the Re-Baseline visit (i.e. subjects who experienced a complete response at completion of the Induction Phase) may have subsequent tumor assessments which show no evidence of disease (i.e. no new lesions, no non-target lesions, sum of diameters of

the target lesion ≤ 5 mm). In such cases, these patients will be assigned a timepoint response of No Evidence of Disease (NED), however a timepoint response of NED is not a valid response for use as a Best Overall Response. Therefore, these subjects cannot achieve a BOR other than NE or PD, and these subjects will be included in the analysis as non-responders. For subjects who have a timepoint Overall Response of NED which occurs at least 42 days after randomization, BOR will be defined as NE with reason "NED maintained post Re-Baseline". Scans showing NED within the 42 day window will not be considered for purposes of BOR determination.

The following requirement is taken into account 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 4 weeks after the initial documentation of CR or PR and before progression
- It is assumed that tumor growth is a continuous process, but due to measurement variability there might be fluctuations in tumor assessments. Therefore, confirmation of response must not necessarily be at the next scan, but could be at any subsequent scan before PD. For instance, if a subject has PR-SD-PR or PR-NE-PR at consecutive tumor assessments, the BOR would qualify for PR.
- The minimum duration for a BOR of stable disease (SD) is defined as at least 6 weeks (42 days) after randomization and before progression

The 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 per investigator assessment in the Maintenance Phase.

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

Table 5 **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

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

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

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

The 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 ORR will be tested by the General Association Statistic of the Cochran-Mantel-Haenszel test (CMH) with the stratum of region taken into account. The null hypothesis of no association in any region is tested against the alternative, which specifies that there is an association between treatment effect and ORR in any given region. The CMH test will be performed with one-sided alpha level of 0.025 at the primary analysis only.

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 $\leq$ 10 in at least one of the stratum within each treatment arm), a Fisher's exact test will be used.

The stratified odds ratio in terms of objective response (OR) 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 divided by the odds of OR with chemotherapy. 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 alpha level of 5% two-sided, 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 target-lesion disease at screening), and NE will be tabulated. Patients with BOR of NE will be summarized by reason for having NE status. The following reasons will be used:

- No post-baseline assessments
- All post-baseline assessments have overall response NE
- No post-baseline assessments due to death within 6 weeks after randomization
- No post-baseline assessments due to other reasons
- 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)
- NED maintained post Re-Baseline

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:

1. 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;
2. 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 9 "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.

Subgroup Analysis

For the subgroup analysis of BOR, the association of treatment 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.

15.2.3 Subject-reported Outcomes/Health-Related Quality of Life

Analysis sets: HRQoL

- 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. Details on these questionnaires are provided in the following sections. The analyses specified in this section will be performed for HRQoL analysis set. CCI [redacted] 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 Maintenance Phase 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 and at the best possible score (see sub-sections below, for details on best and worst scores) will be reported. In addition, the best (biggest improvement), worst (largest decline), and last observed post-Maintenance Phase baseline change from Maintenance Phase baseline values will be summarized.

Change from Maintenance Phase 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, 95% CI's 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 Maintenance Phase 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 symptom scales. Covariates will include the Maintenance Phase baseline score for the domain score being evaluated and the randomization stratification factor. The questionnaire completion time point (Maintenance Phase-study day), 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. 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 Maintenance Phase 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 Maintenance Phase 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 (on the y-axis) between the treatment arms.

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.

15.2.3.1 European Quality of Life (EuroQoL) - 5 Dimensions - 5 Levels (EQ-5D-5L)

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 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 and 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 – it ranges from -0.594 (worst health state) to 1.000 (best health state).

15.2.3.2 European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30

The QLQ-C30 self-assessment questionnaire incorporates five functional scales (Physical, Role, Cognitive, Emotional, and Social), a global health status / HRQoL 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 (Table 6).

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 and a high score for the global health status / HRQoL scale represents a high overall quality of life, 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 / HRQoL scale**:

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

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 / HRQoL scale, which are 7-point questions with Range = 6.

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

Scale	Items
Global Health Status / HRQoL scale	
Global Health Status / HRQoL scale	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

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.

15.2.3.3 European Organization for Research and Treatment of Cancer (EORTC) Gastric Cancer Module QLQ-STO22

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) as shown in Table 7. A higher score represents more problems/ worse symptom severity.

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

Scale	Items
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
Taste	45
Body Image	49
Hair Loss	51, 52

The same scoring methodology will be used for the QLQ-STO22 as for the QLQ-C30.

CCI

CCI

CCI

[redacted]

The primary endpoint of OS time between the two treatment arms will be analyzed using a two-sided unstratified log rank test by PD-L1 expression status; the unstratified HR and its corresponding 95% CI will be computed by PD-L1 expression status as well. Details of the analysis are specified in Section 16.1.4.

The confirmed ORR by treatment group will be calculated along with the two-sided 95% CI using the Clopper-Pearson method by PD-L1 expression status. The association of treatment effect and ORR will be tested by PD-L1 expression status with the same CMH method as described in Section 15.2.2.

16 Safety Evaluation

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 summarized for the on-treatment period of the Induction and/or Maintenance Phase and the analysis sets specified in Table 8. The safety endpoints will be tabulated using descriptive statistics.

Table 8 Safety Analyses by Analysis Set and Study Phase

	Safety – Induction Analysis Set	**Safety-Maintenance Analysis Set**	
	Induction Phase	**Induction Phase**	**Maintenance Phase**
AEs			
Overall Summary by Category	Y	Y	Y
All TEAEs	Y	Y	Y
Related TEAEs	Y	Y	Y
Serious TEAEs	Y	Y	Y
NCI CTC AE Toxicity Grades	Y	Y	Y
AEs Leading to Discontinuation	Y		Y
AEs Leading to Death	Y		Y
irAEs			Y
IRRs			Y
Labs			
Descriptive Statistics		Y	Y
NCI CTCAE Toxicity	Y	Y	Y
Shifts Relative to Normal Range			Y
Vital Signs			
Descriptive Statistics			Y
Change Categories			Y
ECG			
Descriptive Statistics			Y
PCSA Criteria	Y	Y	Y

16.1 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. In this study, there are two on-treatment periods, one for Induction Phase and one for Maintenance Phase.

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

- **Related Adverse Events:** adverse events with relationship to study treatment (as recorded on the AE eCRF page, Relationship with study treatment = Related) reported by the investigator and those of unknown relationship (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 (irAEs):** immune related adverse events are identified via customized MedDRA PT queries and additionally classified by medical assessment according to case definition (two level approach). Details are included in Appendix IV.
- **Infusion Related Reactions (IRRs):** IRRs are identified based on a list of MedDRA PTs. The detailed criteria of the timing relationship to infusion are specified in Table 14 of Appendix IV.

Unless otherwise specified, AEs will be summarized by number and percentage of subjects with the AE in the category of interest as described above, by treatment group, primary System Organ Class (SOC) and Preferred Term (PT) and by decreasing frequency based on the avelumab treatment arm.

Each subject will be counted only once within each SOC or PT. 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.

Within one adverse event, changes in toxicity grade, seriousness, or outcome are recorded as separate AE entries in the eCRF with start and end dates corresponding to the dates of change. In such cases, the start date of one entry will equal the end date of a previous entry. In the case that records report an event with the same preferred term in immediately consecutive periods and the severity of the reported event does not worsen, the records will be considered as one event in the analysis. All events will be maintained as separate records in the database, in order to maintain the full detailed history of the events and preserve traceability. 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. In the case that a subsequent record indicates a worsening of an ongoing event, the subsequent, worse record will be treated as a separate adverse event. If the subsequent, worse record is outside of the on-treatment period, it will not appear on the summaries/listings of TEAEs.

16.1.1 All Adverse Events

Analysis sets: Safety-Induction Analysis Set, Safety-Maintenance Analysis Set

Adverse events will be summarized by worst severity per subject according to the National Cancer Institute – Common Terminology Criteria for Adverse Events (NCI-CTCAE version 4.03), using the latest version of MedDRA PT as event category and MedDRA primary SOC body term 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 table of AEs 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
  - 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.

16.1.2 Adverse Events Leading to Treatment Discontinuation

Analysis sets: Safety-Induction Analysis Set, Safety-Maintenance Analysis Set

The frequency (number and percentage) of subjects with each of the following will be presented for TEAEs leading to permanent discontinuation by treatment group:

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

A listing of all TEAEs leading to treatment discontinuation will also be provided, including the relevant information.

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

16.2.1 Deaths

Analysis sets: Safety-Induction Analysis Set, Safety-Maintenance Analysis Set

All deaths, deaths within 30 days after last dose of randomized study treatment, deaths within 60 days after the date of first dose of randomized study treatment, and primary reason for death will be tabulated based on information from the "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
  - Progressive disease and/or disease related condition
  - 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 randomized 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 randomized study treatment
- Flag for death within 60 days of first dose of randomized study treatment

16.2.2 Serious Adverse Events

Analysis sets: Safety-Induction Analysis Set, Safety-Maintenance Analysis Set

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

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

A listing of SAEs will also be provided with the relevant information, including a flag for SAEs with onset outside of the on-treatment period.

16.2.3 Other Significant Adverse Events

Analysis sets: Safety-Maintenance Analysis Set

The following tables will be created for Maintenance Phase treatment-emergent irAEs by treatment group. Separate summary tables will present potential irAEs (those identified solely by MedDRA preferred term) and medically assessed irAEs (which meet the criteria outlined in Appendix IV). For the potential irAE tables, both treatment arms will be presented with counts, as applicable. Only the counts for the avelumab arm will be displayed in the medically assessed irAE summary tables. "Cluster" is a compilation of PTs that are categorized by immune-related event of special interest.

- The overall summary of irAEs table will include the following categories:
  - All irAEs
  - irAEs leading to death, by Cluster and PT
  - irAEs, by Cluster and PT

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

A listing of all potential and reviewed irAEs will be provided with the relevant information and a flag for irAEs with onset outside of the on-treatment period.

The frequency (number and percentage) of subjects with each of the following will be presented 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 leading to death, by PT
- Related IRRs leading to death, by PT
- IRRs, by PT
- IRRs, Grade ≥ 3, by PT
- Related IRRs, by PT
- Related IRRs, Grade ≥ 3, by PT
- IRRs leading to permanent treatment discontinuation, by PT

- Related IRRs leading to permanent treatment discontinuation, by PT
- Serious IRRs, by PT
- Related serious IRRs, by PT
- Time related to first onset of an IRR (infusion 1, infusion 2, infusion 3, infusion 4 or later)

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

16.3 Clinical Laboratory Evaluation

Analysis sets: Safety-Induction Analysis Set, Safety-Maintenance Analysis Set

16.3.1 Hematology and Chemistry Parameters

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

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 simple descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each nominal visit over time (unscheduled measurements would therefore not be included in these summaries as described in Section 10). Changes from baseline at each nominal visit over time will also be presented in boxplots. End of Treatment and Safety Follow-Up visits will be summarized separately. The changes

computed will be the differences from baseline. Qualitative data based on reference ranges will be described according to the categories (i.e. Low, Normal, High).

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.

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

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

16.3.2 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 and hematology 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.

16.4 Vital Signs

Analysis set: Safety-Maintenance Analysis Set

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

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

For each subject the maximum on-treatment increase/decrease from baseline will be calculated and categorized by vital sign measurement (see Table 9). Missing values will define a separate category. A summary of maximum shift from baseline by category will be provided by treatment arm. A summary of maximum change from baseline per subject will also be provided.

Table 9 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 bpm; ≥ 20 bpm	≤5 breaths/min, >5 – 10 breaths/min, >10 breaths/min
Respiration rate decrease from baseline <20 bpm; ≥ 20 bpm	≤5 breaths/min, >5 – 10 breaths/min, >10 breaths/min

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

16.5 Other Safety or Tolerability Evaluations

Analysis sets: Safety-Induction Analysis Set, Safety-Maintenance Analysis Set

16.5.1 ECG

A 12-lead Electrocardiogram (ECG) assessment will be performed at the Screening, Re-Baseline and at the Discontinuation/End-of-Treatment visits. ECG summaries will include all ECG assessments from the Maintenance Phase on-treatment period. All ECG assessments will be listed, and those collected outside the on-treatment period will be flagged in the listing.

For each of the ECG parameters (heart rate, QTcB, QTcF, QRS interval, PR interval), descriptive statistics at baseline (see Section 10 for definition), at each post-baseline time point and changes from baseline at each post-baseline time point will be summarized by treatment group.

Frequency (number and percentage) of subjects with the worst potentially clinically significant abnormalities (PCSA) for ECG parameters during the on-treatment period will be summarized. Each subject will be counted only once within each category. As ECG assessments are only performed

during Screening, Re-Baseline and at the Discontinuation/End-of-Treatment visit, the denominator to calculate percentages for each PCSA category is the number of subjects with a Discontinuation/End of Treatment visit. The PCSA criteria are provided in Table 10.

Table 10 PCSA Criteria for ECG Test Results

Test	PCSA Criteria
Heart Rate	≤ 50 bpm and decrease from baseline ≥ 20 bpm ≥ 120 bpm and increase from baseline ≥ 20 bpm
PR Interval	≥ 220 msec and increase from baseline ≥ 20 msec
QRS Interval	≥ 120 msec
QTcF and QTcB absolute	Interval >450 msec and interval ≤ 480 msec Interval >480 msec and interval ≤ 500 msec Interval >500 msec
QTcF and QTcB change from baseline	Increase from baseline > 30 msec and ≤ 60 msec; Increase from baseline > 60 msec

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

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

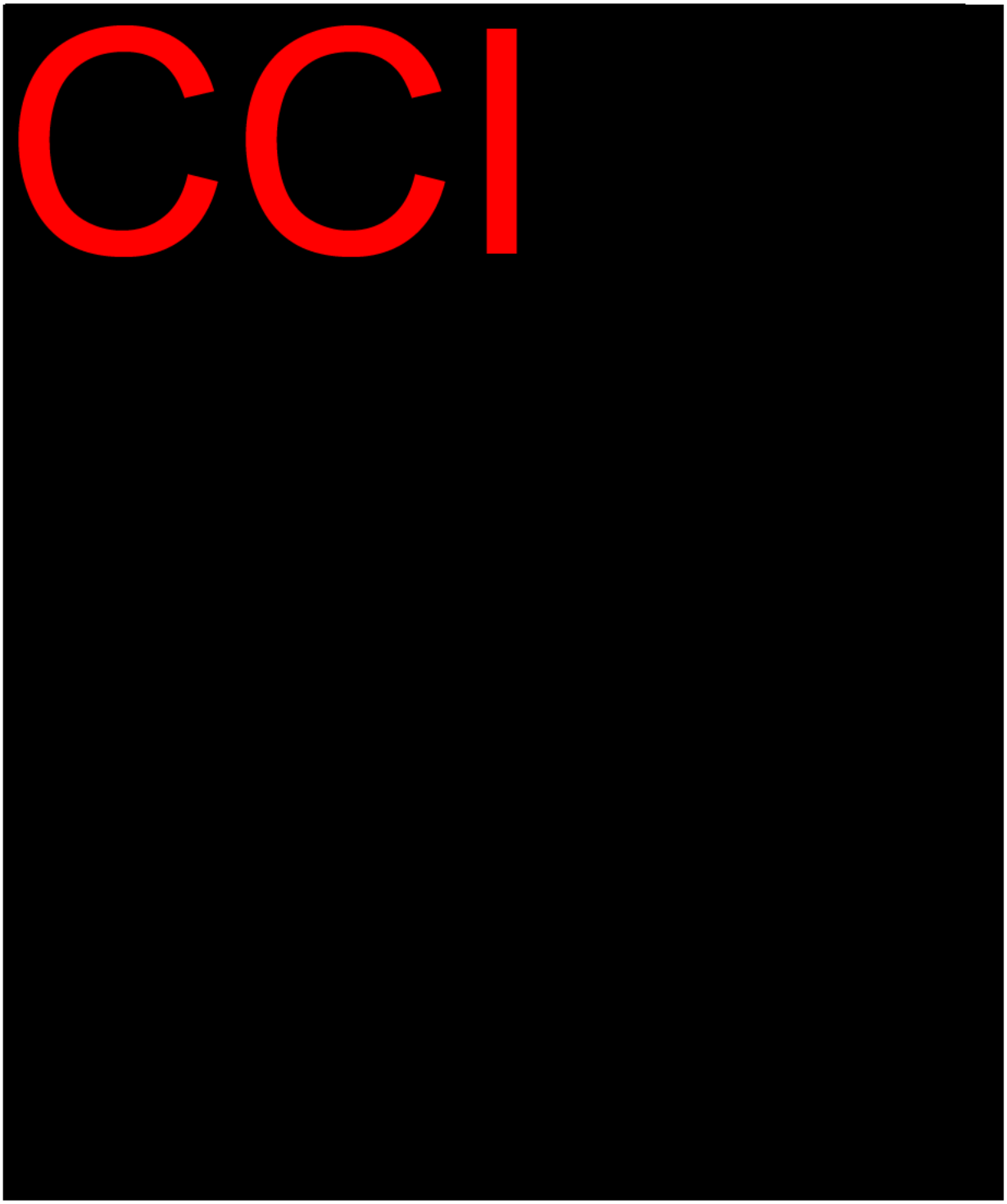
Unscheduled ECG measurements will not be used in computing the descriptive statistics for change from baseline at each post-baseline time point. However, they will be used in the analysis of notable ECG changes and the shift table analysis of notable QT parameters.

16.5.2 ECOG Performance Status

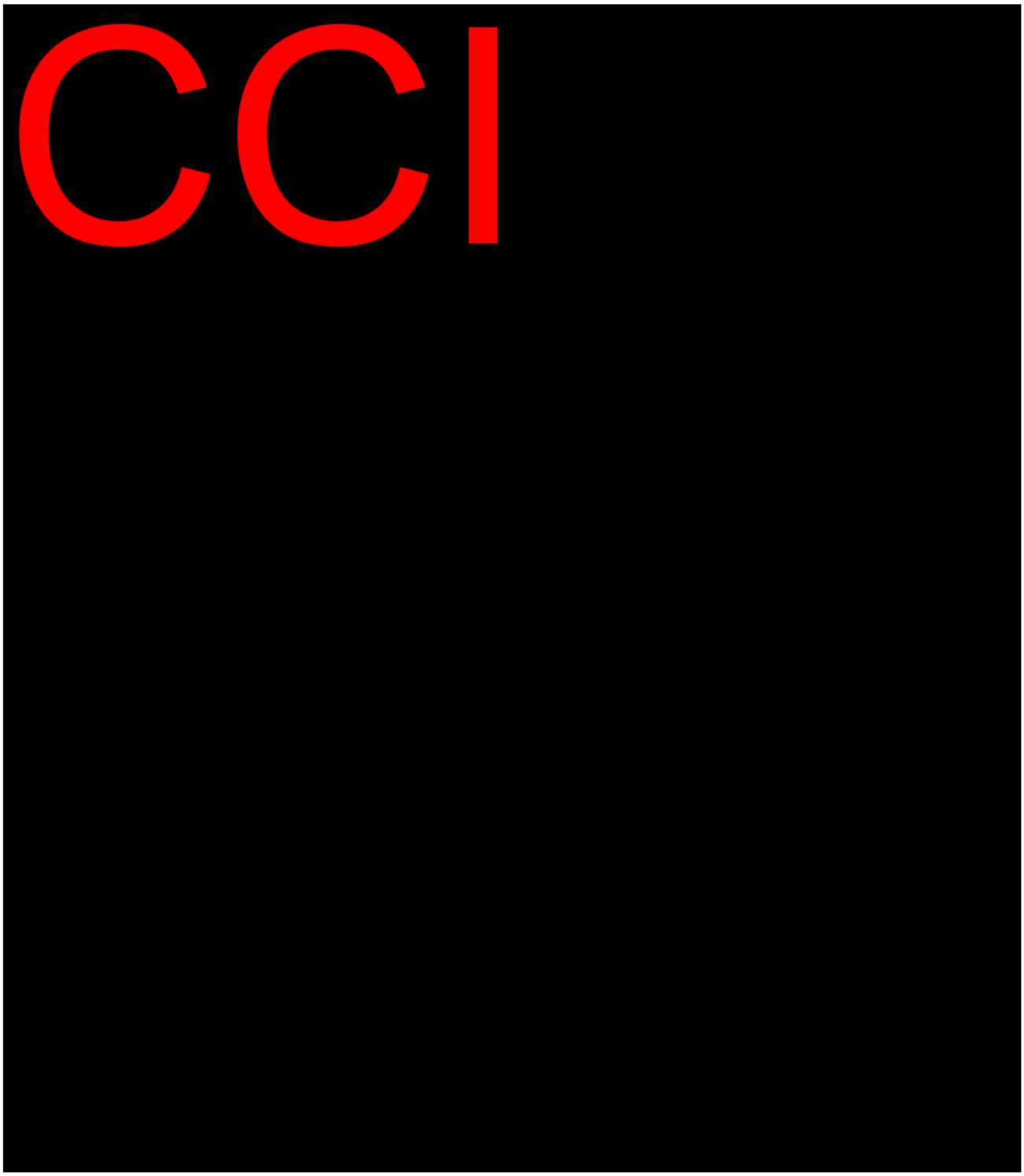
ECOG PS score at Induction Phase baseline and Maintenance Phase baseline will be presented by number and percentage for each treatment arm. In addition, shifts from Maintenance Phase baseline in ECOG PS score will be presented at planned visits during the Maintenance Phase, the Discontinuation visit, and End of Treatment visit by number and percentage for each treatment arm. Please refer to Section 10 for the definition of Maintenance Phase baseline for ECOG PS. Similarly, shifts from baseline to highest ECOG PS score during the Induction Phase and Maintenance Phase 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



[redacted]



[redacted]

17 References

1. Wassmer, G. (2006), Planning and Analyzing Adaptive Group Sequential Survival Trials. Biom. J., 48: 714-729.

2. Hommel G, Bretz F, and Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. Statist. Med. 2007; 26: 4063-4073.

3. Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982; 38:29-41.

4. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Statistics in Medicine 2011; 30:2409-2421.

5. Eisenhauer, E.A, Therasse, P., Bogaerts. J. et. al. (2009): New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer 45, 228-247.

6. Cochran WG. Some Methods for Strengthening the Common Chi-square Tests. Biometrics 1954; 10:417-51.

7. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons 1980.

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

9. Tamhane AC, Wu Y, Mehty CR. Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. Statistics Med 2012; 31:2027-40.

10. Tamhane AC, Mehta CR, Liu L. Testing a Primary and a Secondary Endpoint in a Group Seuqential Design. Biometrics 2010; 66:1174-84.

11. Hung J, Wang SJ, O'Neill R. Statistical Considerations for Testing Multiple Endpoints in Group Sequential or Adaptive Clinical Trials. J Biopharm Statistics 2007; 17:1201-10.

12. Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Statistics Med 2009; 29:219-28.

18 Appendices

Appendix I Clinically Important Protocol Deviations

Description of Protocol Deviation	Deviation Code	Clinically Important PD?
Inclusion / Exclusion criteria		
Subject did not meet inclusion criterion 3, i.e. subject without evaluable PD-L1 expression assessment	INCEXC02	
Subject did not meet inclusion criterion 4, i.e. subject had no measurable disease as per RECIST 1.1	INCEXC03	
Subject did not meet inclusion criterion 5, i.e. subject without histologically confirmed unresectable locally advanced or metastatic adenocarcinoma of the stomach or GEJ	INCEXC04	Yes
Subject did not meet inclusion criterion 8, i.e. subject does not show adequate hematological function as defined in the study protocol.	INCEXC05	
Subject did not meet inclusion criterion 9, i.e. subject does not show adequate hepatic function, as defined in the study protocol.	INCEXC06	
Subject did not meet inclusion criterion 10, i.e. subject does not show adequate renal function, as defined in the study protocol	INCEXC07	
Subject did not meet inclusion criterion 12, i.e. subject refused use of effective contraception, if the risk of conception exists	INCEXC08	
Subject met exclusion criterion 1, i.e. subject 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 (CTLA-4)	INCEXC09	Yes
Subject met exclusion criterion 2, i.e. subject receiving concurrent anticancer treatment (including immune therapy or cytokine therapy except for erythropoietin)	INCEXC10	Yes
Subject met exclusion criterion 3, i.e. subject received prior chemotherapy for unresectable locally advanced or metastatic adenocarcinoma of the stomach or GEJ	INCEXC11	Yes
Subject met exclusion criterion 4, i.e. subject's disease harbors a HER-2 positivity by IHC (≥3+) or FISH (IHC 2+, FISH +)	INCEXC12	
Subject met exclusion criterion 5, i.e. major surgery for any reason, except diagnostic biopsy, within 4 weeks of enrolment and/or if the subject has not fully recovered from the surgery within 4 weeks of enrolment	INCEXC13	
Subject met exclusion criterion 6, i.e. subject received immunosuppressive agents and these drugs were not tapered off before initiation of the trial treatment	INCEXC14	

Description of Protocol Deviation	Deviation Code	Clinically Important PD?
Subject met exclusion criterion 7, i.e. subject has brain metastases not meeting the exceptions laid out in the study protocol	INCEXC15	Yes
Subject met exclusion criterion 8, i.e. subject had 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.	INCEXC16	
Subject met exclusion criterion 10, i.e. subject has significant acute or chronic infections (e.g. HIV, AIDS, HCV, HBV, TB)	INCEXC17	
Subject met exclusion criterion 11, i.e. subject has an active autoimmune disease that might deteriorate when receiving an immunostimulatory agent	INCEXC18	
Subject met exclusion criterion 15, i.e. subject is pregnant or lactating	INCEXC19	
Any other inclusion/exclusion criteria not met and not covered above	INCEXC20/ INCEXC21	
Informed Consent/ Subject information		
Subject did not sign ICF and was enrolled.	INFCON01	
Site Personnel who signed the ICF or performed any study related procedure and is not listed on the Site Delegation Log.	INFCON02	
ICF is signed after study procedure was done, with the exception of assessments such as CT/MRI that were performed as part of normal site procedures. Exempted assessments must have been performed within the 28-day screening window.	INFCON03	
Incorrect version of ICF signed	INFCON04	
Missing Date or Time recorded in the ICF document.	INFCON06	
New ICF version not signed in a timely manner	INFCON07	
Any other informed consent deviation not covered above	INFCON08/ INFCON09	
Subject Visit Schedule		
Subject visits not performed/completed	VISITS01	
No survival information / no collection of additional cancer medication post disease progression. NOTE: if the subject withdrew fully from consent and did not agree to provide follow-up information, there is no PD	VISITS02	
Subject visits performed outside of window (-4/+2 day), except those due to safety reasons e.g. treatment delay due to toxicity	VISITS03	
Investigational Product		
IP preparation, storage and administration of prepared infusion was not according to Protocol/Pharmacy manual	INVPRO01	
Aveluamb infusion rate not performed per protocol (e.g. not between 50-80 min or at a reduced rate if required per safety management)	INVPRO02	

Description of Protocol Deviation	Deviation Code	Clinically Important PD?
IP storage temperature out of range and site used IP or patient received expired IP	INVPRO04	
Subject received wrong treatment per IXRS	INVPRO05	
Subject received an incorrect Avelumab dose	INVPRO06	
Avelumab premedication was administered but not according to the protocol requirements prior to IP dose	INVPRO07	
Chemotherapy premedication was not administered prior to dose or not according to label or local guidelines	INVPRO08	
Avelumab premedication was not administered at all, prior to IP dose	INVPRO09	
IP distributed without IXRS call	INVPRO10	
Subject received an incorrect Physician's choice chemotherapy dose (≥ +/- 5 % assigned dose)	INVPRO11	
Missed chemotherapy dose due to reason other than Adverse Event	INVPRO12	
Missed avelumab dose due to reason other than Adverse Event	INVPRO13	
Concomitant Medications/ Procedures		
Subject took immunotherapy, immunosuppressive drugs (that is, chemotherapy or systemic corticosteroids except for short-term treatment of allergic reactions or for the treatment of irAEs), or other anti-cancer or experimental pharmaceutical products.	CONMED02	
Subject randomized to receive avelumab received treatment with growth factors (granulocyte colony stimulating factor or granulocyte macrophage colony stimulating factor).	CONMED03	
Subject received herbal remedies with immunostimulating properties (for example, mistletoe extract) or known to potentially interfere with major organ function (for example, hypericin) or have abused alcohol or other drugs during the trial	CONMED05	
Subject had major surgery during the study	CONMED06	
Subject received radiation therapy (except for superficial lesions or bone-directed), while on study treatment.	CONMED07	
Other deviation on concomitant medication not covered above	CONMED08/ CONMED09	
Protocol Procedures		
Subject randomized >10 days after re-baseline period started	PROCED02	
Source data not available/ not eligible for CRA review	PROCED03	
Study procedures not done/omitted	PROCED06	
Subject met criteria for permanent withdrawal from trial treatment but treatment was not discontinued	PROCED07	
Contraception not used per protocol requirements	PROCED08	

Description of Protocol Deviation	Deviation Code	Clinically Important PD?
SAEs not reported within the mandated timeframe to Sponsor (e.g. within 24 hours) or SAE follow-up information not provided in a timely manner	PROCED10	
Screening or re-baseline CT/MRI scans not performed	PROCED11	
Screening or re-baseline images (lung/abdomen/pelvis, brain, bone scan if clinically indicated) not performed within required time window. Screening: 28 days (brain 6 weeks). Re-baseline: 10 days	PROCED12	
6-weekly tumor evaluation/staging imaging scan (CT/MRI) not performed at all	PROCED13	
6 weekly Tumor Evaluation/Staging imaging scan (CT/MRI) performed >10 days outside window	PROCED14	
Paper QoL used and the data were moved from the paper QoL to the tablet by the site staff	PROCED16	
Subject randomized despite having disease progression assessed by the independent radiologist	PROCED20	Yes
Laboratory		
Subject enrolled or randomized without all safety lab results available (including pregnancy test for females)	LABPRCO01	
Subject dosed without all safety lab results available (including pregnancy test for females) or not reviewed by the Investigator prior to dosing	LABPRCO02	
Randomization		
Randomization interrupted in IWRS resulting in erroneous application of the randomization code	RANDOM01	Yes
Patient randomized more than once	RANDOM02	
Other		
Any other protocol deviation which is deemed to be significant for a specific subject or for the trial but has not been pre-specified in this table	OTHPRC01/ OTHPRC02	

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

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

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 Definition for Assessment of Immune-Related AEs and IRRs

Immune-Related Adverse Events

Immune-related adverse events (irAEs) will be identified programmatically. AEs which satisfy all of the following criteria will be flagged as immune-related:

1) The AE preferred term matches a preferred term on the list of pre-selected MedDRA terms.

2) The AE onset occurs after the first study drug administration and no more than 90 days after last dose.

3) On the AE eCRF page, the question "Were Corticosteroids, Immunosuppresants, or hormonal therapy (e.g. Thyroid) applied?" has the answer "Yes" selected.

4) On the AE eCRF page, either:

a. The question "Does any of the following provide a clear etiology for the event?" has the answer "No" selected, indicating that the AE is not attributable to underlying cancer disease/PD, prior or concomitant medications/procedures, nor another medical condition such as an infection or pre-existing disease.

OR

b. The AE eCRF indicates that a biopsy was performed and the question "Is the histopathology/biopsy consistent with an immune-mediated event?" has the answer "Yes" selected.

In the case that criteria (1) through (3) are met, and entries for condition (4) are missing, the following rules will be applied:

- If the answer to "Does any of the following provide a clear etiology for the event?" (4a) is missing, the event will be considered as irAE (irrespective of biopsy results).
- If the answer to "Is the histopathology/biopsy consistent with an immune-mediated event?" (4b) is missing, or if no biopsy was performed, and condition (4a) is not satisfied (i.e. "Yes" is selected as the answer to the question "Does any of the following provide a clear etiology for the event?"), the event will be considered as a non-irAE.

IRRs

Infusion related reactions are identified based on a list of MedDRA PTs and criteria on the timely relationship according to Table 15.

Table 15 Criteria for infusion related reactions

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

ELECTRONIC SIGNATURES

Document: ctp-emr100070-007-sap-v4

Signed By	Event Name	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'GMT'Z)
PPD	Task Completed (Approval eSign): Approved	Business Approval	PPD
PPD	Task Completed (Approval eSign): Approved	Business Approval	PPD


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI


CCI