Official Title of Study:

A Randomized, Multicenter, Open-Label, Phase 3 Study of Nivolumab plus Ipilimumab or Nivolumab in Combination with Oxaliplatin plus Fluoropyrimidine versus Oxaliplatin plus Fluoropyrimidine in Subjects with Previously Untreated Advanced or Metastatic Gastric or Gastroesophageal Junction Cancer
(CheckMate 649: CHECKpoint pathway and nivoluMab clinical Trial Evaluation 649)

NCT Number: NCT02872116

Document Date (Date in which document was last revised): 19-MAY-2021

STATISTICAL ANALYSIS PLAN
FOR CLINICAL STUDY REPORT

RANDOMIZED, MULTICENTER, OPEN-LABEL, PHASE 3 STUDY OF NIVOLUMAB PLUS IPILIMUMAB OR NIVOLUMAB IN COMBINATION WITH OXALIPLATIN PLUS FLUOROPYRIMIDINE VERSUS OXALIPLATIN PLUS FLUOROPYRIMIDINE IN SUBJECTS WITH PREVIOUSLY UNTREATED ADVANCED OR METASTATIC GASTRIC OR GASTROESOPHAGEAL JUNCTION CANCER

PROTOCOL CA209649

VERSION # 5.0

DATE: 19-MAY-2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 BACKGROUND AND RATIONALE

This document provides the statistical analyses that will be conducted to support the clinical study report (CSR) of the study and reflects the revised protocol 09.

CA209649 is a Phase 3, randomized, open-label study of nivolumab plus ipilimumab or nivolumab in combination with oxaliplatin plus fluoropyrimidine vs. oxaliplatin plus fluoropyrimidine in subjects with previously untreated advanced or metastatic gastroesophageal junction (GEJ) or gastric cancer (GC). This study will determine if nivolumab plus ipilimumab improves overall survival (OS) vs. the standard of care, oxaliplatin plus fluoropyrimidine, in subjects with GC/GEJ cancer with enriched PD-L1 expression by combined positive score (CPS); or whether nivolumab in combination with oxaliplatin plus fluoropyrimidine improves OS or PFS vs. the standard of care, oxaliplatin plus fluoropyrimidine, in subjects with GC/GEJ cancer with enriched PD-L1 expressing by CPS. Additional objectives include further characterization of efficacy, safety and tolerability, pharmacokinetics (PK), patient-reported outcomes, and potential predictive biomarkers of nivolumab in combination with ipilimumab or nivolumab in combination with oxaliplatin plus fluoropyrimidine in subjects with GC/GEJ whose tumors do or do not express PD-L1.

Research Hypotheses:

In subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 5, the administration of nivolumab in combination with oxaliplatin plus fluoropyrimidine (nivo+chemo) will improve PFS or OS compared to oxaliplatin plus fluoropyrimidine (chemo).

In subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 5, the administration of nivolumab plus ipilimumab (nivo+ipi) will improve OS compared to oxaliplatin plus fluoropyrimidine (chemo).

Schedule of Analyses:

The primary analysis of PFS of nivo+chemo vs. chemo in randomized subjects with PD-L1 CPS ≥ 5 will be conducted after a 12-month follow-up (12 months after the last subjects was randomized in the study). The primary analysis of OS of nivo+chemo vs. chemo in randomized subjects with PD-L1 CPS ≥ 5 includes one interim and final analysis. The interim OS analysis will be conducted after a 12-month follow-up, concurrently with the primary PFS analysis. The final OS analysis of nivo+chemo vs. chemo will be conducted after a 24-month follow-up, ie, 12 months after the OS interim analysis.

The analysis of OS of nivo+ipi vs. chemo in randomized subjects with PD-L1 CPS ≥ 5 will be conducted after 24 months follow-up after the last subjects was randomized in the study. Because the nivo + ipi arm was closed to enrolment earlier, this translates to approximately 36 months follow-up for subjects randomized to nivo+ipi or chemo. This analysis will be conducted at the time of the OS final analysis of nivo+chemo vs. chemo.

The timing of the efficacy analyses is described in Table 1-1.

Table 1-1: Timing of Analyses

	Endpoints		
Timing of Analysis	**PFS nivo+chemo vs. chemo in randomized subjects with PD-L1 CPS ≥ 5**	**OS nivo+chemo vs. chemoin randomized subjects with PD-L1 CPS ≥ 5**	**OS nivo+ipi vs. chemo in randomized subjects with PD-L1 CPS ≥ 5**
12 months after last subject randomized	X	X (IA)	
24 months after last subject randomized[a]		X (FA)	X

[a] Twenty-four months follow-up from last patient randomized to nivo+chemo vs. chemo, corresponds to 36-month minimum follow-up for nivo+ipi vs chemo.

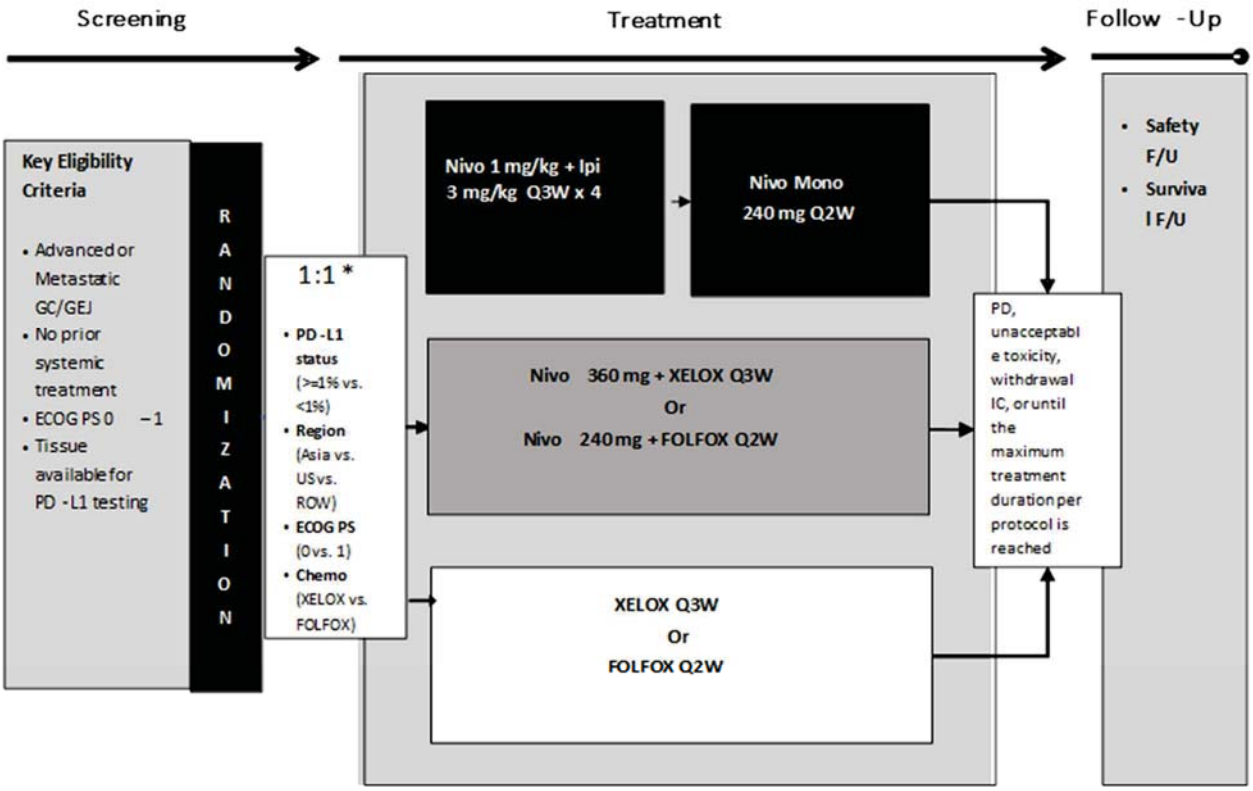
2 STUDY DESCRIPTION

2.1 Study Design

The study design described in this section reflects the revised protocol 09 (including protocol amendment 29).

This is a Phase 3, randomized, open-label, three-arm study of nivolumab plus ipilimumab or nivolumab in combination with oxaliplatin plus fluoropyrimidine vs. oxaliplatin plus fluoropyrimidine in subjects with previously untreated advanced or metastatic GC or GEJ cancer.

Figure 2.1-1: Study Scheme



* Nivolumab + Ipilimumab arm is closed to enrollment as of 05-June-2018.

The accrual was complete in May 2019 and the sample size is 2031 subjects randomized, with 83 subjects randomized in 1:1 ratio to nivo + ipi vs. chemo, followed by ~1098 subjects randomized in 1:1:1 ratio to the 3 arms until the closure of enrollment of the nivo+ipi arm, and finally 850 subjects randomized to nivo+chemo vs. chemo (see also Table 5.1-1).

The treatment administration is as follows:

- Nivo+Ipi Arm: nivolumab 1mg/kg administered IV over 30 minutes followed by ipilimumab 3mg/kg administered IV over 30 minutes on Day 1 of each treatment cycle every 3 weeks for 4 doses (Cycles 1 to 4), followed by nivolumab 240mg administered IV over 30 minutes on Day 1 of each treatment cycle every 2 weeks (Cycle 5 and beyond)

Subjects treated in the nivolumab 240mg Q2W have an option to switch nivolumab 480mg Q4W, the nivolumab 480mg should be given 2 weeks following the last administration of nivolumab 240 mg dose.

- Nivo+Chemo Arm (XELOX or FOLFOX):
  - Nivolumab plus XELOX:
    - Nivolumab 360mg IV over 30 minutes on Day 1 of each treatment cycle, every 3 weeks
    - Oxaliplatin 130mg/m^2 IV on Day 1 of each treatment cycle + capecitabine 1000mg/m^2 orally twice daily on Days 1 to 14 of each treatment cycle, every 3 weeks
  - Nivolumab plus FOLFOX:
    - Nivolumab 240mg IV over 30 minutes on Day 1 of each treatment cycle, every 2 weeks
    - Oxaliplatin 85mg/m^2 + leucovorin 400mg/m^2 + fluorouracil 400mg/m^2 IV on Day 1 of each treatment cycle, and fluorouracil 1200 mg/m^2 IV continuous infusion over 24 hours daily or per local standard on Days 1 and 2 of each treatment cycle, every 2 weeks

Nivolumab 480mg Q4W is permitted when all chemotherapy components are discontinued and nivolumab is continued alone. Subjects must be in the study for a minimum of 6 months from Cycle 1, Day 1 before switching to 480mg Q4W nivolumab. Nivolumab 480mg should be given 2 weeks or 3 weeks following the last administration of nivolumab 240mg Q2W or 360mg Q3W, respectively.

- Chemo Arm (XELOX or FOLFOX):
  - XELOX: Oxaliplatin 130mg/m^2 IV on Day 1 of each treatment cycle + capecitabine 1000mg/m^2 orally twice daily on Days 1 to 14 of each treatment cycle, every 3 weeks
  - FOLFOX: Oxaliplatin 85mg/m^2 + leucovorin 400mg/m^2 + fluorouracil 400mg/m^2 IV on Day 1 of each treatment cycle, and fluorouracil 1200mg/m^2 IV continuous infusion over 24 hours daily or per local standard on Days 1 and 2 of each treatment cycle, every 2 weeks.

The treatment will be administered until disease progression (PD) (unless treatment beyond progression is permitted, in nivo + chemo and nivo + ipi arms), unacceptable toxicity, or subject withdrawal of consent, whichever occurs first.

The treatment with nivolumab will be given for up to 24 months in the absence of disease progression or unacceptable toxicity. Chemotherapy will be given as per the study dosing schedule. Upon fulfillment of protocol pre-specified criteria, and providing written informed consent, treatment with nivolumab could be reinitiated as per the initial schedule for subsequent disease progression and administered for up to 1 additional year.

The study will consist of 3 phases: screening, treatment, and follow-up.

Screening Phase

- Begins by establishing the subject's initial eligibility and signing of the informed consent form (ICF).
- Subject is enrolled using the Interactive Response Technology (IRT) system.

- The choice of chemotherapy regimen XELOX or FOLFOX must be decided before placing the randomization call to the IRT system.
- Subjects must have PD-L1 immunohistochemistry (IHC) testing, with results, performed by the central lab during the Screening period. Without evaluable PD-L1 result, the subject may not be randomized.
- Subject is assessed for study eligibility according to the inclusion and exclusion criteria. All screening assessments and procedures must be performed within 28 days prior to randomization unless otherwise specified.

Treatment Phase

- Begins after the randomization call to the IRT system. The subject is randomly assigned in a 1:1:1 ratio to the nivo+chemo, nivo+ipi or the chemo arm up to protocol amendment 20. After amendment 20, the subject is randomized to 1:1 the nivo+chemo or the chemo arm as enrollment to nivo+ipi was closed.
- Administration of study treatment is to begin within 3 calendar days of randomization.
- The treatment phase ends when the subject is discontinued from study therapy.
- Treated subjects will be evaluated for tumor assessments every 6 weeks (± 7 days) up to and including Week 48, then every 12 weeks (± 7 days) thereafter.

Dosing in the Nivo+Ipi Arm:

- On the day of infusion, nivolumab is to be administered first. The second infusion will always be ipilimumab, and will start at least 30 minutes after completion of the nivolumab infusion.
- Nivolumab (as flat dose) is allowed to be administrated alone in case ipilimumab has to be discontinued due to toxicity. Ipilimumab is also allowed to be administrated alone in case nivolumab has to be discontinued due to toxicity.
- Treatment beyond initial, investigator-assessed, RECIST 1.1-defined progression is permitted if the subject has investigator-assessed clinical benefit and is tolerating treatment.

Dosing in the Nivo+Chemo Arm:

- On the day of infusion, nivolumab is to be administered first. The administration procedures of chemotherapy will follow local standards.
- Nivolumab is allowed to be administrated alone in case the chemotherapy has to be discontinued due to toxicity. Chemotherapy alone is also allowed to be administered in doubtlet or monotherapy in case nivolumab has to be discontinued due to unacceptable toxicity.
- Nivolumab treatment beyond initial, investigator-assessed, RECIST 1.1-defined progression is permitted if the subject has investigator-assessed clinical benefit and is tolerating treatment. Chemotherapy alone is not allowed to treat beyond progression.

- No cross-over is allowed between XELOX and FOLFOX.

Dosing in the Chemo Arm:

- The administration procedures will follow local standards.
- No cross-over is allowed between XELOX and FOLFOX.

Follow-up Phase

- Begins when the decision is made to discontinue a subject from study therapy.
- After completion of the first 2 follow-up (FU) visits (30 days [± 7 days] from last dose and 84 days [± 7 days] from FU1), subjects will be followed by clinic visit or phone contact every 3 months (±14 days) or more frequently as needed for survival status and subsequent anti-cancer therapy.
- Subjects who discontinue treatment for reasons other than PD will continue to have tumor assessments (until PD) every 6 weeks (± 7 days) up to and including Week 48, then every 12 weeks (±7 days) thereafter.
- Subjects will be followed for drug-related toxicities until these toxicities resolve, return to baseline or are deemed irreversible. All drug-related toxicities will be documented for a minimum of 100 days after the last dose of study drug.

2.2 Treatment Assignment

After the subject's initial eligibility is established and informed consent has been obtained, the subject must be enrolled into the study by entering information into the IRT system to obtain the subject number. Every subject that signs the ICF must be assigned a subject number in the IRT.

Once enrolled in the IRT, subjects that have met all eligibility criteria (the required tumor tissue received and result obtained by the central laboratory) will be ready to be randomized through the IRT.

Randomization will be stratified by the following factors:

- PD-L1 expression level (tumor cells) (≥ 1% vs. < 1% or indeterminate)
- Region: Asia vs. US vs. RoW
- ECOG performance status: 0 vs. 1
- Planned chemotherapy regimen (XELOX vs. FOLFOX)*

* Chemotherapy was added as a randomization stratification factor with Amendment 08. Information has been collected in IRT even for subjects randomized prior to amendment 08.

2.3 Blinding and Unblinding

The trial is open label, however the Sponsor will handle the study as a blinded study and no analyses or summaries by randomized treatment arms will be produced during the conduct of the study by the Sponsor.

The final PFS analysis in randomized subjects with PD-L1 CPS ≥ 5 and interim OS analyses in randomized subjects with PD-L1 CPS ≥ 5 will be reviewed by the Data Monitoring Committee (DMC). The DMC will also review interim OS analyses in randomized subjects with PD-L1 CPS ≥ 1 and the interim OS analysis in all randomized subjets. The analyses, by unblinded arms, will be provided by the DMC independent reporting statistician to the DMC members.

If the formal efficacy interim analysis meets the pre-specified statistical criteria for at least one of the primary endpoints (PFS in PD-L1 CPS ≥ 5 or OS in PD-L1 CPS ≥ 5), the DMC Chairman will notify the BMS Head of Oncology Clinical Development by releasing the DMC closed report with unblinded treatment arms. The BMS Head of Oncology Clinical Development may in turn choose to release the closed report to additional members of the BMS restricted team after reviewing the closed DMC report. The BMS Head of Oncology Clinical Development will inform the DMC members and the independent statistician by email on whether the restricted BMS team has decided that the BMS study team should be unblinded to the study results.

The closed report, which will be shared with the BMS restricted team, will only contain the aggregated OS results by treatment arms for the CPS cutoffs which meet statistical significance. It will not contain OS results for the CPS cutoffs that do not meet statistical significance unless it is requested by health authorities. The PFS results will be included in the closed report if at least one of the PFS endpoint and OS endpoint meets statistical significance as this is the final and unique analysis for PFS.

In addition, the Blinded Independent Central Review (BICR) Committee will perform the central imaging review without knowledge of the treatment group assignment.

The Lab vendor performing the tumor cell PD-L1 and CPS PD-L1 scoring is also blinded to the treatment group assignment.

2.4 Protocol Amendments

Global amendments incorporated in the protocol with relevant changes are described in Table 2.4-1.

Table 2.4-1: Relevant Protocol Amendments

Amendments	Date of Issue	Summary of Major Changes
Amendment 29	16-sep-2019	• Based on the data from the Keynote 062 study, PFS and OS Kaplan-Meier curves showed an observed delayed separation, the timing of PFS and OS analyses are updated with minimum follow up 12 and 24 months. In addition, to reduce variability of efficacy results, PFS population was expanded to all randomized subjects with PD-L1 CPS ≥ 5.

Table 2.4-1: Relevant Protocol Amendments

Amendments	Date of Issue	Summary of Major Changes
		• A 480mg Q4W nivolumab dosing option for subjects who receive nivolumab alone after treatment with nivolumab in combination with ipilimumab or FOLFOX/XELOX is allowed.
Amendment 26	15-Nov-2018	• Based on recent internal data, added approximately 356 subjects into randomization, in total of approximately 2005 subjects will be randomized to keep sample size for primary analyses of PFS and OS in PD-L1 CPS ≥ 5 subjects for nivolumab in combination with chemotherapy vs. chemotherapy.
Amendment 23	14-Sep-2018	• Incorporates the combined positive score (CPS) for PD-L1 expression, the primary population is now subjects with PD-L1 expression ≥ 5 by CPS rather than by the tumor proportion score (TPS) for subjects with nivolumab in combination with oxaliplatin and fluoropyrimidine compared to oxaliplatin and fluoropyrimidine. • The planned analysis of overall survival (OS) for nivolumab plus ipilimumab arm is changed to a secondary objective. • Objective response rate (ORR) has been changed to a secondary endpoint. • The primary, secondary, and exploratory objectives have been updated to reflect changes in subject population definition. • Second disease progression (PFS2) and time to secondary subsequent therapy (TSST) have been added as exploratory analyses. • Statistical assumptions and considerations revised to reflect changes in study population and revised objectives.
Amendment 20	11-Jun-2018	• Per recommendation of the Data Monitoring Committee (DMC), as of 05-June-2018, the nivolumab plus ipilimumab arm is now closed. Subjects randomized to this arm prior to or on 05-June-2018 will continue to receive treatment with study drugs per protocol, and the study data will remain blinded until planned primary analysis.
Amendment 19	29-May-2018	• Added that randomization of an additional 300 subjects beyond the originally planned 1349 subjects across the 3 study arms will allow for more robust analysis of the treatment effect of nivolumab in combination with ipilimumab or chemotherapy across different PD-L1 cutoffs in 1L GC/GEJ cancer.
Amendment 08	07-Dec-2016	• Added a new randomization arm, N+C to the study. • Added chemotherapy as randomization stratification factor. • Added Blinded Independent Central Review (BICR) of tumor images. • Clarified some study procedures, including limiting the treatment period with nivolumab to 2 years.

2.5 Blinded Independent Central Review of Progression

The clinical management of subjects during the study will be based upon local radiologic tumor measurements. Tumor assessments for each subject should be submitted to the radiology vendor as they are performed, on an ongoing basis. The blinded, independent radiologists will review all available tumor assessments for that given subject and determine if RECIST 1.1 criteria for progression have been met.

If clinically acceptable, subsequent therapy should begin only after RECIST 1.1 progression has been assessed by blinded independent central review. When RECIST 1.1 progression is assessed by the investigator (whether assessed before or after the start of subsequent therapy), confirmation by the BICR must be requested. Tumor assessments may be discontinued when the independent radiologist assesses the subject to have met RECIST 1.1 criteria for progression.

In addition, subjects receiving treatment beyond progression must continue tumor assessments until treatment has been discontinued.

2.6 Data Monitoring Committee

An independent DMC will be utilized to provide oversight of safety and efficacy considerations. Additionally, the DMC will provide advice to the sponsor regarding actions the committee deems necessary for the continuing protection of subjects enrolled in the study. The DMC will be charged with assessing such actions in light of an acceptable benefit/risk profile for nivolumab plus ipilimumab and nivolumab plus chemotherapy. The DMC will act in an advisory capacity to BMS and will monitor subject safety and evaluate the available efficacy data for the study. The oncology therapeutic area of BMS has primary responsibility for design and conduct of the study.

The DMC will also be responsible for conducting formal interim efficacy analysis of the study for final PFS analysis and interim analysis of OS for nivo + chemo vs. chemo.

3 OBJECTIVES

3.1 Primary

Nivolumab in combination with oxaliplatin plus fluoropyrimidine vs. oxaliplatin plus fluoropyrimidine:

To compare OS in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 5.

To compare PFS, as assessed by BICR in subjects with advanced or metastatic GC or GEJ cancer with PD-L1CPS ≥ 5.

3.2 Secondary

Nivolumab plus ipilimumab vs. oxaliplatin plus fluoropyrimidine:

- To compare OS in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 5 or all randomized subjects.
- To compare the time to symptom deterioration (TTSD) as assessed using the Gastric Cancer Subscale (GaCS) of the Functional Assessment of Cancer Therapy-Gastric (FACT-Ga) in

subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 5 or all randomized subjects.

- To evaluate OS in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 10 or subjects with PD-L1 CPS ≥ 1.
- To evaluate PFS, as assessed by BICR, in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects.
- To evaluate ORR, as assessed by BICR, in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects.

Nivolumab in combination with oxaliplatin plus fluoropyrimidine vs. oxaliplatin plus fluoropyrimidine:

- To compare OS in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 1 or all randomized subjects
- To evaluate OS in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 10.
- To evaluate PFS, as assessed by BICR, in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 10, 1 or all randomized subjects.
- To evaluate ORR, as assessed by BICR, in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 10, 5, 1, or all randomized subjects.

3.3 Exploratory

Nivolumab in combination with oxaliplatin plus fluoropyrimidine vs. oxaliplatin plus fluoropyrimidine:

- To assess TTSD using GaCS of FACT-Ga in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 5 or all randomized subjects.

For both comparisons: Nivolumab in combination with oxaliplatin plus fluoropyrimidine vs. oxaliplatin plus fluoropyrimidine and Nivolumab plus ipilimumab vs. oxaliplatin plus fluoropyrimidine:

- To assess PFS and ORR, as assessed by the investigator in subjects with advanced or metastatic GC or GEJ cancer across PD-L1 CPS cut-offs.
- To evaluate Duration of Response (DOR) as assessed by BICR and by investigator, in subjects with advanced or metastatic GC or GEJ.
- To evaluate the durable response rate (DRR) (defined as objective response lasting continuously > 6 months) as assessed by BICR and by investigator, in subjects with advanced or metastatic GC or GEJ cancer.
- To evaluate at 18, 24, and 36 months survival rates in subjects with advanced or metastatic GC or GEJ cancer.

- To evaluate second disease progression (PFS2) or time to second subsequent therapy (TSST) in subjects with advanced or metastatic GC or GEJ cancers.
- To assess PFS, ORR as assessed by either BICR or investigator, OS in subjects with advanced or metastatic GC or GEJ cancer across PD-L1 tumor proportion score (TPS) cut-offs.
- To assess the overall safety and tolerability of nivolumab plus ipilimumab or nivolumab in combination with oxaliplatin plus fluoropyrimidine and oxaliplatin plus fluoropyrimidine in subjects with advanced or metastatic GC or GEJ cancer.
- To explore potential biomarkers predictive of or associated with clinical efficacy (OS, PFS and ORR) and/or incidence of AEs of nivolumab plus ipilimumab or nivolumab in combination with oxaliplatin plus fluoropyrimidine including but not limited to microsatellite instability (MSI) status, tumor mutational burden (TMB), [redacted] and inflammatory signatures in subjects with advanced or metastatic GC or GEJ cancer.
- To assess TTSD as assessed using GaCS of FACT-Ga in subjects with advanced or metastatic GC or GEJ cancer with PD-L1 CPS ≥ 10, or 1.
- To assess changes from baseline in the subject's overall health status using the 3-level version of the EQ-5D (EQ-5D-3L) index and visual analog scale (EQ-VAS).
- To assess the subject's cancer-related quality of life using the FACT-Ga questionnaire and selected components, including the GaCS and 7-item version of the FACT-General (FACT-G7).

4 ENDPOINTS

A summary of the efficacy endpoints is presented in Table 4-1

Table 4-1: Summary of Key Efficacy Endpoints

	Nivolumab plus Chemotherapy vs Chemotherapy	Nivolumab plus Ipilimumab vs Chemotherapy
Primary endpoints	• PFS by BICR in subjects with PD-L1 CPS ≥ 5 **(controlled for Type I error)** • OS in subjects with PD-L1 CPS ≥ 5 **(controlled for Type I error)**	
Secondary endpoints	• OS in subjects with PD-L1 CPS ≥ 1, and in all randomized subjects **(controlled for Type I error)** • OS in subjects PD-L1 CPS ≥ 10 • PFS by BICR in subjects with PD-L1 CPS ≥ 10, 1 or all randomized subjects	• OS in subjects with PD-L1 CPS ≥ 5 and in all randomized subjects. **(controlled for Type I error)** • OS in subjects PD-L1 CPS ≥ 10,1 • ORR, PFS by BICR in subjects with PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects. • TTSD in subjects with PD-L1 CPS ≥ 5 and in all randomized subjects **(controlled for Type I error)**

Table 4-1: Summary of Key Efficacy Endpoints

	Nivolumab plus Chemotherapy vs Chemotherapy	Nivolumab plus Ipilimumab vs Chemotherapy
Exploratory Endpoints	• ORR by BICR in subjects with PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects • ORR, PFS by investigator in subjects with PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects • OS, PFS[a], ORR[b] in subjects with PD-L1 TPS across cut-offs • OS rates at 18, 24, and 36 months • PFS 2 or TSST on next line treatment • DOR • DRR • TTSD in subjects with PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects • Biomarkers	• ORR, PFS by investigator in subjects with PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects • OS, PFS[c], ORR[d] in subjects with PD-L1 TPS across cutoffs • OS rates at 18, 24, and 36 months • PFS 2 or TSST on next line treatment • DOR • DRR • TTSD in subjects with PD-L1 CPS ≥ 10, or 1 • Biomarkers

[a] Per BICR and investigator assessments

4.1 Primary Endpoints for Nivolumab plus Chemotherapy

For nivo+chemo versus chemo.

4.1.1 Overall Survival

Overall Survival is defined as the time from the date of randomization to the date of death from any cause. For subjects that are alive, their survival time will be censored at the date of last contact date (or "last known alive date"). Overall survival will be censored at the date of randomization for subjects who were randomized but had no follow-up.

Primary objective of the study is the comparison of OS between nivo+chemo and chemo in PD-L1 CPS ≥ 5 subjects.

4.1.2 Progression-Free Survival

Two definitions are used for analysis of Progression-Free Survival. The primary definition accounts for subsequent therapy by censoring at the last evaluable tumor assessment on or prior to the date of subsequent therapy. This definition will be used to run the primary analysis of PFS (per BICR) for this trial and to declare statistical significance for the PFS endpoint.

The secondary definition is irrespective of subsequent therapy and will be presented as supportive analysis of the primary PFS analysis. This definition is considered primary by EMA.

Clinical deterioration in the absence of unequivocal evidence of progression (per RECIST v1.1 criteria) is not considered progression for purposes of determining PFS.

PFS rates at fixed time points (eg, 6 months, depending on the minimum follow-up) are defined as the probability that a subject has not progressed and is alive at time *T* following randomization.

The first on-study tumor assessment is scheduled to be conducted at 6 weeks (± 1 week) following first dose. Subsequent tumor assessments are scheduled every 6 weeks (± 1 week) up to and including Week 48, then every 12 weeks (± 7 days) regardless of treatment schedule until disease progression.

4.1.2.1 Primary Definition of Progression-Free Survival (Accounting for Subsequent Therapy)

The primary definition of PFS (PFS truncated at subsequent therapy) is defined as the time between the date of randomization and the date of first documented tumor progression, based on BICR assessments (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first.

The following censoring rules will be applied for the primary definition of PFS:

- Subjects who did not progress or die will be censored on the date of their last evaluable tumor assessment.
- Subjects who did not have any on-study tumor assessments and did not die will be censored on their date of randomization.
- Subjects who received subsequent anti-cancer therapy prior to documented progression will be censored at the date of the last evaluable tumor assessment conducted on or prior to the date of initiation of the subsequent anti-cancer therapy.
- Subjects who did not have a documented progression and received subsequent anti-cancer therapy will be censored at the date of the last evaluable tumor assessment conducted on or prior to the initiation of the subsequent anti-cancer therapy.

In the above, subsequent anti-cancer therapy is all systemic therapies, surgery with reason of tumor resection palliative or tumor resection curative, radiotherapy with reason curative or palliative.

Censoring rules for the primary definition of PFS (PFS truncated at subsequent therapy) are presented in Figure 4.1.2.1-1 and in Table 4.1.2.1-1.

Figure 4.1.2.1-1: PFS Primary Definition

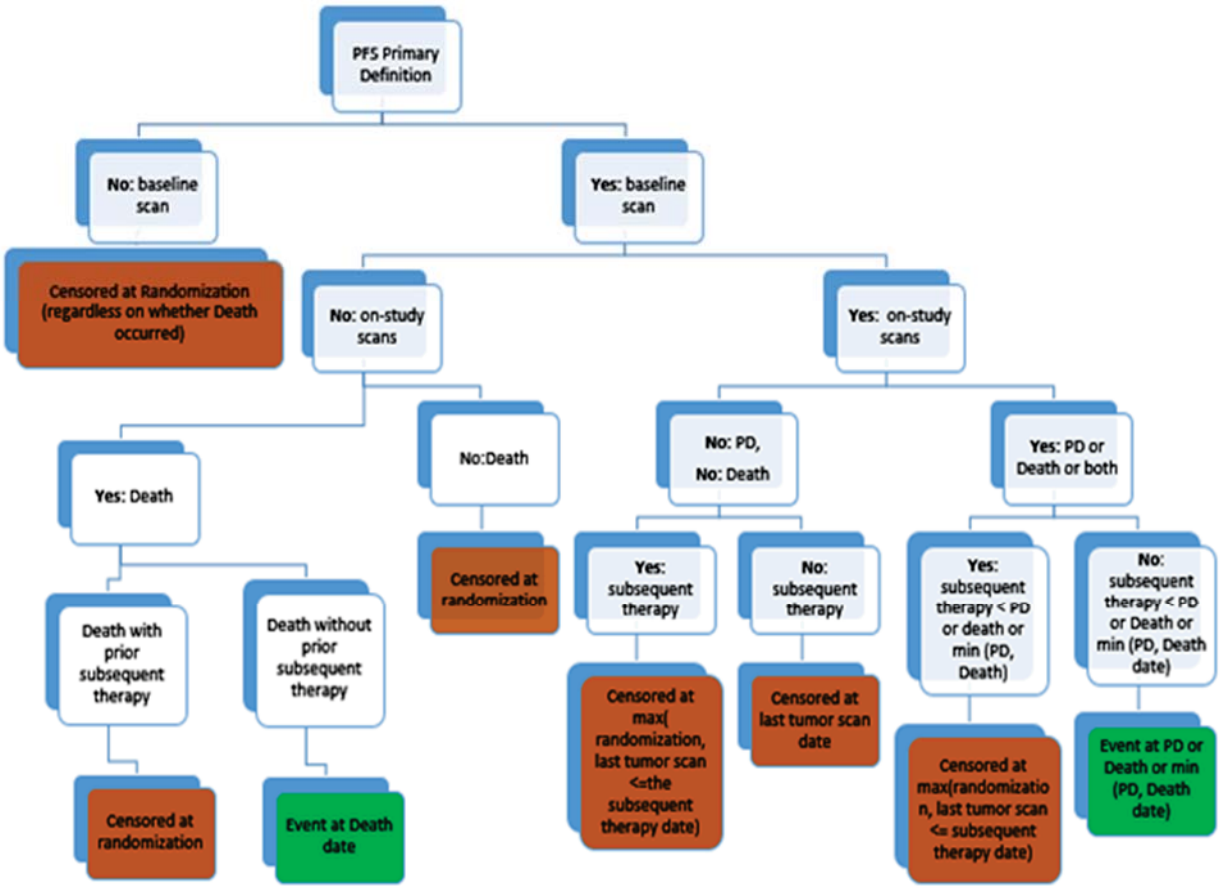


Table 4.1.2.1-1: Censoring Scheme Used in Primary Definition of PFS

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments*	Date of randomization	Censored
No on study tumor assessments and no death*	Date of randomization	Censored
Subsequent anti-cancer therapy started without prior progression per RECIST v1.1 or death	Date of last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy	Censored
Documented progression per RECIST v1.1 and no subsequent anti-cancer therapy started before	Date of the first documented progression per RECIST v1.1 (excludes clinical progression)	Progressed
No progression and no death, and no new anti-cancer therapy started	Date of last evaluable tumor assessment	Censored
Death without progression per RECIST v1.1 and no prior subsequent anti-cancer therapy	Date of death	Progressed

* Tumor assessments and death if any, occurring after start of subsequent anti-cancer therapy are not considered.

Primary objective of the study is the comparison of PFS between nivo+chemo and chemo in PD-L1 CPS ≥ 5 subjects.

4.1.2.2 Secondary Definition of Progression Free Survival (Irrespective of Subsequent Therapy)

The secondary definition of PFS (ITT definition) is defined as the time between the date of randomization and the date of first documented tumor progression, based on BICR assessments (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first.

The following censoring rules will be applied for the secondary definition of PFS:

- Subjects who did not progress or die will be censored on the date of their last evaluable tumor assessment.
- Subjects who did not have any on study tumor assessments and did not die will be censored on their date of randomization.

Censoring rules for the secondary definition of PFS (ITT definition) are presented in Figure 4.1.2.2-1 and in Table 4.1.2.2-1.

Figure 4.1.2.2-1: PFS Secondary Definition

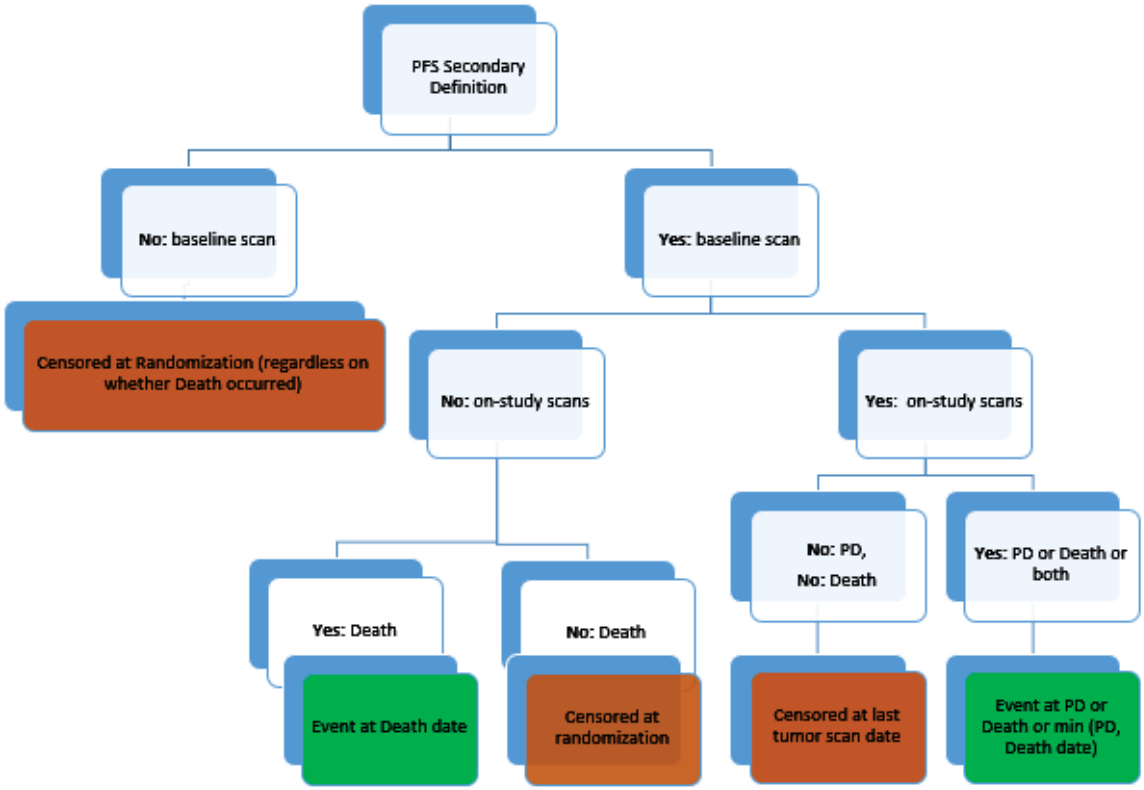


Table 4.1.2.2-1: Censoring Scheme for Secondary Definition of PFS

Situation	Date of Progression of Censoring	Outcome
No baseline tumor assessment	Date of randomization	Censored
No on-study tumor assessments and no death	Date of randomization	Censored
Documented progression per RECIST v1.1	Date of first documented progression per RECIST v1.1 criteria (excludes clinical progression)	Progressed
No progression and no death	Date of last evaluable tumor assessment	Censored
Death without progression per RECIST v1.1	Date of death	Progressed

4.2 Secondary Endpoints

4.2.1 For Nivo + Chemo versus Chemo

4.2.1.1 Overall Survival

Secondary objectives are the comparison of OS between nivo+chemo and chemo in PD-L1 CPS ≥ 1 subjects and in all randomized subjects. OS will also be evaluated for subjects with CPS ≥ 10. Overall survival is as defined in Section 4.1.1.

4.2.1.2 Progression-Free Survival

Secondary objectives are the evaluation of PFS per BICR for nivo+chemo and chemo in PD-L1 CPS ≥ 1 subjects, CPS ≥ 10 subjects and in all randomized subjects. Progression free survival is as defined in Section 4.1.2.

4.2.1.3 Objective Response Rate

Objective Response Rate (ORR) is defined as the number of randomized subjects who achieve a best overall response of confirmed complete response (CR) or confirmed partial response (PR) based on BICR assessment (using RECISTS v1.1) divided by the number of randomized subjects.

Best Overall Response (BOR) is defined as the best response, as determined by the BICR, recorded between the date of randomization and the date of objectively documented progression per RECIST v1.1 criteria or the date of subsequent anti-cancer therapy (including all systemic therapies, surgery with reason of tumor resection palliative or tumor resection curative, radiotherapy with reason curative or palliative), whichever occurs first. For subjects without documented progression or subsequent therapy, all available response designations will contribute to the BOR determination.

ORR will be evaluated for nivo + chemo and chemo in PD-L1 CPS ≥ 10, 5, 1 or all randomized subjects.

4.2.2 For Nivo + Ipi versus Chemo.

4.2.2.1 Overall Survival

Secondary objectives are the comparison of OS between nivo+ipi and chemo in PD-L1 CPS ≥ 5 subjects and in all randomized subjects. OS will also be evaluated for subjects with CPS ≥ 10 and CPS ≥ 1 subjects. Overall survival is as defined in Section 4.1.1.

4.2.2.2 Progression Free Survival

Secondary objectives are the evaluation of PFS per BICR for nivo+ipi and chemo in PD-L1 CPS ≥ 10, CPS ≥ 5, CPS ≥ 1and in all randomized subjects. Progression free survival is as defined in Section 4.1.2.

4.2.2.3 Objective Response Rate

Secondary objectives are the evaluation of ORR per BICR for nivo+ipi and chemo in PD-L1 CPS ≥ 10, CPS ≥ 5, CPS ≥ 1 and in all randomized subjects. Objective response rate is as defined in Section 4.2.1.3.

4.2.2.4 Time to Symptom Deterioration

Time to symptom deterioration is defined as the time from randomization until a clinically meaningful decline from baseline in GaCS score. A clinically meaningful deterioration is defined as a reduction of 8.2 points in the GaCS score. Only events before treatment discontinuation will be taken into account: patients who do not deteriorate before treatment discontinuation will be censored at the time of their last on-treatment GaCS assessment. Those who have no GaCS assessments after randomization and those who were never treated will be censored on the date of randomization. For more details on the GaCS scale, see Section 4.3.3. Secondary objectives are the comparison of TTSD between nivo+ipi and chemo in PD-L1 CPS ≥ 5 subjects and all randomized subjects.

4.3 Exploratory Endpoints

4.3.1 Efficacy Exploratory Endpoints

All the following exploratory endpoints may be explored on the different CPS population (ie, CPS ≥ 10, CPS ≥ 5, CPS ≥ 1 and all randomized).

Progression-Free Survival

Progression-free Survival as assessed by investigator is defined similarly as the corresponding endpoints as assessed by BICR.

Objective Response Rate

Objective Response Rate (as assessed by investigator) is defined similarly as the corresponding endpoint as assessed by BICR.

Duration of Response

Duration of response is defined as the time between the date of first confirmed documented response (CR or PR) and the date of the first documented tumor progression as determined by the

BICR or investigator (per RECIST v1.1 criteria) or death due to any cause, whichever occurs first. Censoring rules for DOR follow the censoring rules for PFS primary definition.

Durable Response rate

Durable Response Rate is defined as the number of subjects with confirmed documented response (CR or PR) as determined by the BICR or investigator (per RECIST v1.1 criteria) whose response lasted for at least 6 months, divided by the number of randomized subjects.

Survival Rate

OS rates at fixed time points (eg, 18, 14, 16 months, depending on the minimum follow-up) are defined as the probability that a subject is alive at time *T* following randomization. Overall survival rates at other timepoints are defined similarly.

PFS2/TSST

PFS on next-line therapy (PFS2) is defined as the time from randomization to objectively documented progression, per investigator assessment, on the start of the next line of therapy or the start of the second next line therapy or to death from any cause, whichever occurs first. Subjects who are alive and without progression on the next line of therapy will be censored at last known alive date.

The following censoring rules will be applied for PFS2:

- Subjects who did not receive subsequent anti-cancer therapy (ie, next-line therapy):
  - Subjects who died, the death date is the event date;
  - Else the subject's PFS2 is censored at the last known alive date.
- Subjects who received subsequent anti-cancer therapy (ie, next-line therapy):
  - Subjects who had a disease progression after the start of subsequent next line anti-cancer therapy, this disease progression date is the event date;
  - Else if a subject died or start of second next line therapy, the date of min (death, start date of second next line therapy) is the event date;
  - Else the subject's PFS2 is censored at the last known alive date.

4.3.2 Safety Exploratory Endpoints

The assessment of safety will be based on the incidence of adverse events (AEs), serious adverse events (SAEs), adverse events leading to discontinuation, adverse events leading to dose modification, select adverse events (select AEs) for EU/ROW Submissions, immune-mediated AEs (IMAEs) for US Submission, other events of special interest (OEOSI), and deaths. The use of immune modulating concomitant medication will be also summarized. In addition, clinical laboratory tests, and immunogenicity (ie, development of anti-drug antibody) will be analyzed.

4.3.3 Outcomes Research Exploratory Endpoints

EQ-5D

Subjects' reports of general health status will be measured using the EuroQoL EQ-5D-3L. EQ-5D-3L essentially has 2 components: the descriptive system and the visual analogue scale (VAS).

The instrument's descriptive system consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 3 levels, reflecting "no health problems," "moderate health problems," and "extreme health problems." A dimension for which there are no problems is said to be at level 1, while a dimension for which there are extreme problems is said to be at level 3. Thus, the vectors 11111 and 33333 represent the best health state and the worst health state, respectively, described by the EQ-5D-3L. Altogether, the instrument describes 35 = 243 health states. Empirically derived weights can be applied to an individual's responses to the EQ-5D-3L descriptive system to generate an index measuring the value to society of his or her current health. Such preference-weighting systems have been developed for the UK, US, Spain, Germany, and numerous other populations. For this study, EQ-5D-3L utility index values will be computed using a scoring algorithm based on the United Kingdom Time-Trade-Off (UK TTO) value set.[1]

In addition, the EQ-5D-3L includes a VAS, which allows respondents to rate their own current health on a 101-point scale ranging from 0="worst imaginable" health to 100="best imaginable" health state.[2]

A change from baseline of 0.08 for the EQ-5D-3L utility index score and of 7 for the EQ-5D-3L VAS are considered minimally important differences for the EQ-5D-3L.[3]

FACT-Ga

The FACT-Ga questionnaire and selected components, including the FACT-G7 and GaCS, will be used to assess the effects of underlying disease and its treatment on health-related quality of life (HRQoL) for patients.

As a generic cancer-related questionnaire, the FACT-Ga includes the 27-item FACT-General (FACT-G) to assess symptoms and treatment-related effects impacting physical well-being (PWB; seven items), social/family well-being (SWB; seven items), emotional well-being (EWB; six items), and functional well-being (FWB; seven items). Seven of these items comprise the FACT-G7, an abbreviated version of the FACT-G that provides a rapid assessment of general HRQoL in cancer patients.

In addition to the FACT-G, The FACT-Ga includes a 19-item disease-specific GaCS that assesses additional disease specific symptoms and impacts relating to pain, reflux, dysphagia, eating difficulties, tiredness, weakness, interference, and difficulty planning. Each FACT-Ga item is rated on a five-point scale ranging from 0 (not at all) to 4 (very much).

Scores for the PWB, FWB, SWB, and EWB subscales can be combined to produce a FACT-G total score, which provides an overall indicant of generic HRQoL. The FACT-G and GaCS scores can be combined to produce a total score for the FACT-Ga, which provides a composite measure

of general and targeted HRQoL. Higher scores indicate better HRQoL. The full FACT-Ga will be administered to subjects during the on-treatment phase and at follow up visits 1 and 2. However, to minimize subject response and administrative burden, only the FACT GaCS and FACT-G7 will be administered during the survival follow-up phase.

4.3.4 Pharmacokinetics Exploratory Endpoints

PK will be determined from serum nivolumab and ipilimumab concentrations. Samples will be collected to characterize pharmacokinetics of nivolumab and ipilimumab and to explore exposure-safety and exposure-efficacy relationships.

4.3.5 Immunogenicity Exploratory Endpoints

Serum samples collected will be analyzed by a validated immunogenicity assay. Selected serum samples may be analyzed by an exploratory orthogonal method that measures anti-nivolumab and anti-ipilimumab.

In addition, ad hoc serum samples designated for pharmacokinetic or biomarker assessments may also be used for immunogenicity analysis if required (eg, insufficient volume for complete immunogenicity assessment or to follow up on suspected immunogenicity related AE).

Immunogenicity endpoints are defined in APPENDIX 3.

5 SAMPLE SIZE DETERMINATION

5.1 Randomization Schema

The original study design (before Amendment 08) had 2 arms, with subjects being randomized in a 1:1 ratio to the nivolumab plus ipilimumab or to the oxaliplatin plus fluoropyrimidine arm (XELOX or FOLFOX; arm will be further referred to as chemotherapy arm). Amendment 08 added a new arm: nivolumab in combination with oxaliplatin plus fluoropyrimidine (XELOX or FOLFOX; arm will be further referred to as nivolumab plus chemotherapy arm). The IRT switched to a 1:1:1 randomization at all the participating sites when this third arm was ready to be opened on 27-Mar-2017. Up until that point, there were 83 subjects randomized in the 1:1 randomization to nivolumab plus ipilimumab and chemotherapy arm. In Amendment 08, it was planned to randomize 1266 subjects at 1:1:1 ratio to nivolumab plus ipilimumab, chemotherapy and nivolumab plus chemotherapy arms. The total number of subjects to be randomized was planned to be 1349.

Newly available internal and external data suggested a stronger predictability of PD-L1 CPS than TPS of immunotherapy treatment effects. However, uncertainty still exists of CPS at different cutoffs when extrapolated to first line metastatic setting in GC and GEJ. To ensure sufficient sample size for evaluating the correlation of efficacy and different CPS cutoffs and to keep continuity of enrollment of the study, on 29-May-2018, Amendment 19 was approved to allow additional 300 subjects to be randomized under 1:1:1 ratio for a total sample size of 1566 to the 3 treatment arms.

Per DMC recommendation, randomization to the nivolumab plus ipilimumab arm was stopped on 05-June-2018. The subjects who were already randomized to treatment in this arm prior to or on

05-June-2018 continued to receive study treatment per protocol, and the study data remained blinded for the planned analysis. As such, the 1:1:1 randomization was carried through on 05-Jun-2018. These changes were implemented in Amendment 20. At this point, a total of 1098 subjects were randomized under 1:1:1 ratio. In addition, there were 83 subjects randomized to nivolumab plus ipilimumab and to chemotherapy arms (prior to Amendment 08). Following Amendment 20, subjects were to be randomized in 1:1 ratio to either the nivolumab plus chemotherapy or the chemotherapy arm.

In November 2018, data suggested that the PD-L1 CPS ≥ 5 prevalence was ~27% (unpublished data), which was lower than the original estimate of 35%. Therefore, the enrollment was extended to ensure that the study was appropriately powered for PFS and OS primary endpoints in the CPS ≥ 5population. To allow for this, Amendment 26 increased the total sample size from 1649 to approximately 2005. This resizing of the total population was to ensure that the sample size of the enriched population for the primary analyses (subjects with CPS ≥ 5) was the same as was specified in Amendment 23 to maintain the same power under the statistical assumptions as stated in Amendment 23.

The accrual was completed in May 2019 and in total 2031 subjects were randomized. It is estimated that 1582 subjects were concurrently randomized to nivolumab plus chemotherapy and to chemotherapy and that 815 subjects were concurrently randomized to nivolumab plus ipilimumab and chemotherapy.

Table 5.1-1 present the randomization schema and sample size.

Table 5.1-1: Randomization Allocation and Sample Size

Protocol Periods	Randomization allocation	Number of Randomized Subjects Current / Cumulative
First patient randomized - Amendment 8	1:1 (Nivolumab +Ipilimumab: Chemotherapy)	83/83
Amendment 08 - Amendment 20	1:1:1 (Nivolumab + Ipilimumab: Chemotherapy: Nivolumab + Chemotherapy)	1098/1181
Amendment 20 to amendment 26	1:1 (Chemotherapy: Nivolumab + Chemotherapy)	468/1649
Amendment 26 to the end of enrollment	1:1 (Chemotherapy: Nivolumab + Chemotherapy)	382/2031

5.1.1 Changes of Statistical Analyses and Timing of Analyses in Protocol Amendment 29

Per protocol Amendment 23 (November 2018), the primary population of both primary endpoints PFS and OS was changed to randomized subjects with PD-L1 CPS ≥ 5 and the PD-L1 CPS scoring process started for subjects already randomized in the study to date. In March 2019, the need for additional steps for the PD-L1 CPS scoring process was established and consequently the expected

completion date of PD-L1 CPS scoring will be delayed and is now projected to complete approximately 6 months or later after the completion of enrollment.

Recent data in 1L GC Phase 3 study suggested a longer delay of separation of KM curves of immuno-oncology drug plus chemo vs chemo for PFS and OS.[4] These data suggested that a longer follow-up would warrant to fully capture of the treatment effect for the primary endpoints of PFS and OS in the N+CT vs. CT.

Considering the delay of completion of PD-L1 CPS scoring, and the need to have longer minimum follow-up for PFS and OS to ensure data maturity, Amendment 29 documents the changes on the timing of PFS and OS analyses from event driven to time driven. In addition, Amendment 29 specifically, documents the changes of the primary population of PFS.

The primary analysis of PFS in randomized subjects with PD-L1 CPS ≥ 5 will be conducted after a 12-month minimum follow-up (from the date last patient was concurrently randomized to N+CT or CT). The primary analysis of OS in randomized subjects with PD-L1 CPS ≥ 5 includes one interim and final analysis. The interim analysis will be conducted after 12-month minimum follow-up along with the primary PFS analysis. The final OS analysis will be conducted after a 24-month minimum follow-up, ie, 12 months after the OS interim analyses.

The analysis of OS in N+I vs. CT in randomized subjects with PD-L1 CPS ≥ 5 will be conducted at approximately 36-month minimum follow-up (after the last patient was randomized to N+I or CT). This analysis will be at the time of the final analysis of N+CT vs. CT.

The timing of the efficacy analyses is described in Table 1-1.

For the primary analyses of the endpoints stated above, all events observed in the locked database will be used.

As of May 2019, study enrollment has been completed. Therefore in Amendment 29, the total sample size was fixed; however, the sample size for the primary population (subjects with PD-L1 CPS ≥ 5) remained unknown as CPS scoring was still ongoing. The prevalence was assumed to be 35% consistent with sample size and power calculations in Amendment 23. Sections 8.1.4 and 8.1.5 of the Protocol were updated to reflect design characteristics for time driven analyses and using 35% prevalence for PD-L1 CPS ≥ 5.

5.2 General Assumptions of Sample Size Determinations

Primary Endpoints Family:

For the comparison of nivolumab plus chemotherapy vs. chemotherapy both PFS and OS are primary endpoints.

Type I Error Splitting for Primary Endpoints Family:

For sample size calculation purpose, the family-wise error rate of 5% will be split as follows:

1) Comparison of PFS between nivolumab plus chemotherapy and chemotherapy in subjects with PD-L1 CPS ≥ 5, with alpha of 2%

2) Comparison of OS between nivolumab plus chemotherapy and chemotherapy in subjects with PD-L1 CPS ≥ 5, with alpha of 3%

Population for Primary Endpoints:

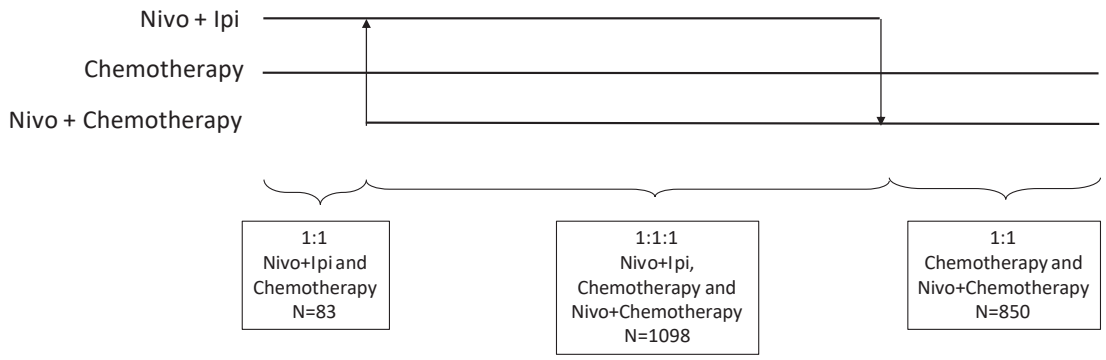
The population for the primary endpoints will be randomized subjects with PD-L1 CPS ≥ 5. For the design purpose, the prevalence was assumed to be 35% of all randomized subjects and is used the power statement in Sections 5.2.1 and 5.2.2

Concurrently Randomized:

For the comparison of nivolumab plus chemotherapy and chemotherapy only subjects who were randomized to those 2 arms concurrently will be used. This means subjects randomized to chemotherapy before the nivolumab plus chemotherapy arm was introduced will not be included in the analysis of this comparison.

For the comparison of nivolumab + ipilimumab and chemotherapy only subjects who were randomized to those 2 arms concurrently will be used. This means subjects randomized to chemotherapy after the closure of nivolumab plus ipilimumab randomization will not be included in the analysis of this comparison.

Figure 5.2-1: Randomization Schema



Assumptions for Control Arm of Chemotherapy:

In previous protocol versions (up to amendment 20), PD-L1 ≥ 1% per TPS was assumed as negative prognostic factor (ie, the median OS for chemotherapy was assumed to be smaller for subjects with tumor PD-L1 ≥ 1% per TPS than for all randomized subjects). However, for PD-L1 CPS, recent data does not suggest such a negative prognostic effect for chemotherapy[5]. Therefore, in Amendment 23, the OS and PFS distributions for the chemotherapy arm are assumed to be the same regardless of CPS cutoffs, eg, in PD-L1 CPS ≥ 5 and in all randomized subjects.

The PFS and OS data from the external data[6,7,8] were considered the most relevant to the studied chemotherapy arm (control). Therefore, the chemotherapy arm is assumed to have a 4-piece exponential distribution with a median PFS of 5.5 months (see Figure 5.2-2) and a 4-piece exponential distribution with a median OS of 11.1 months (see Figure 5.2-3).

Figure 5.2-2: PFS for Chemotherapy from External Data and Assumed Distribution for Chemotherapy

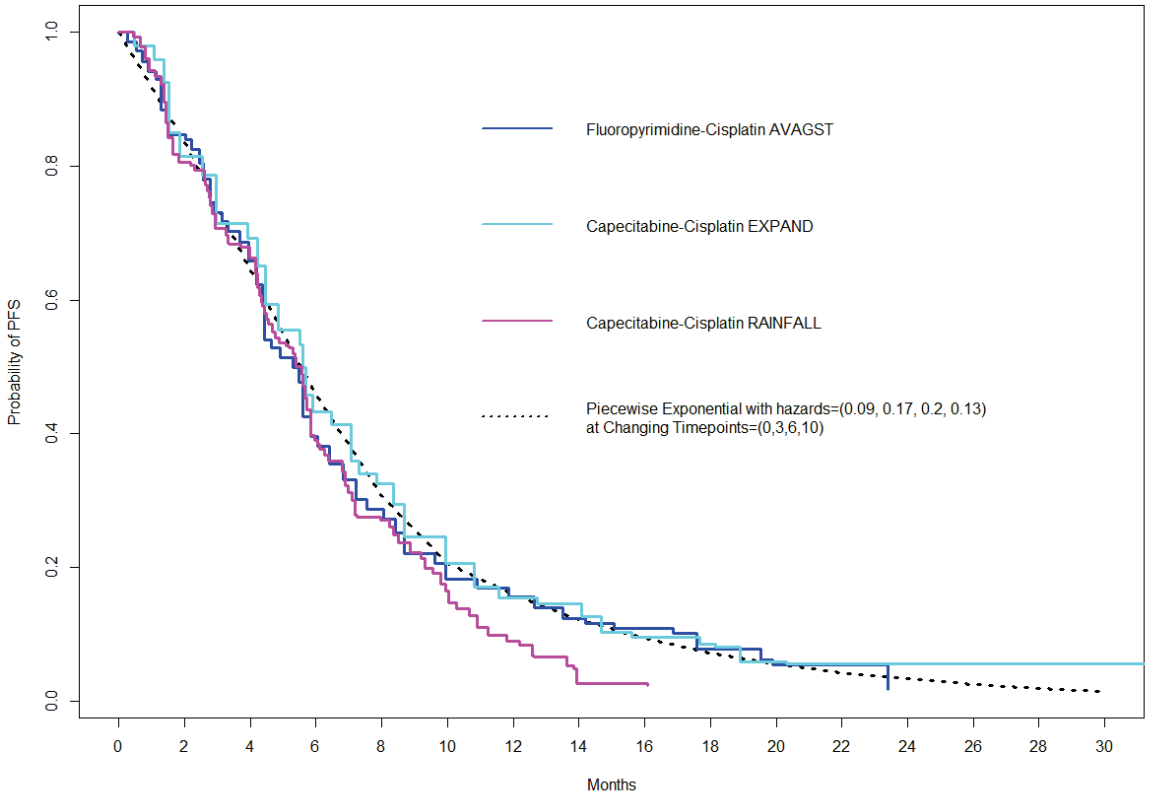
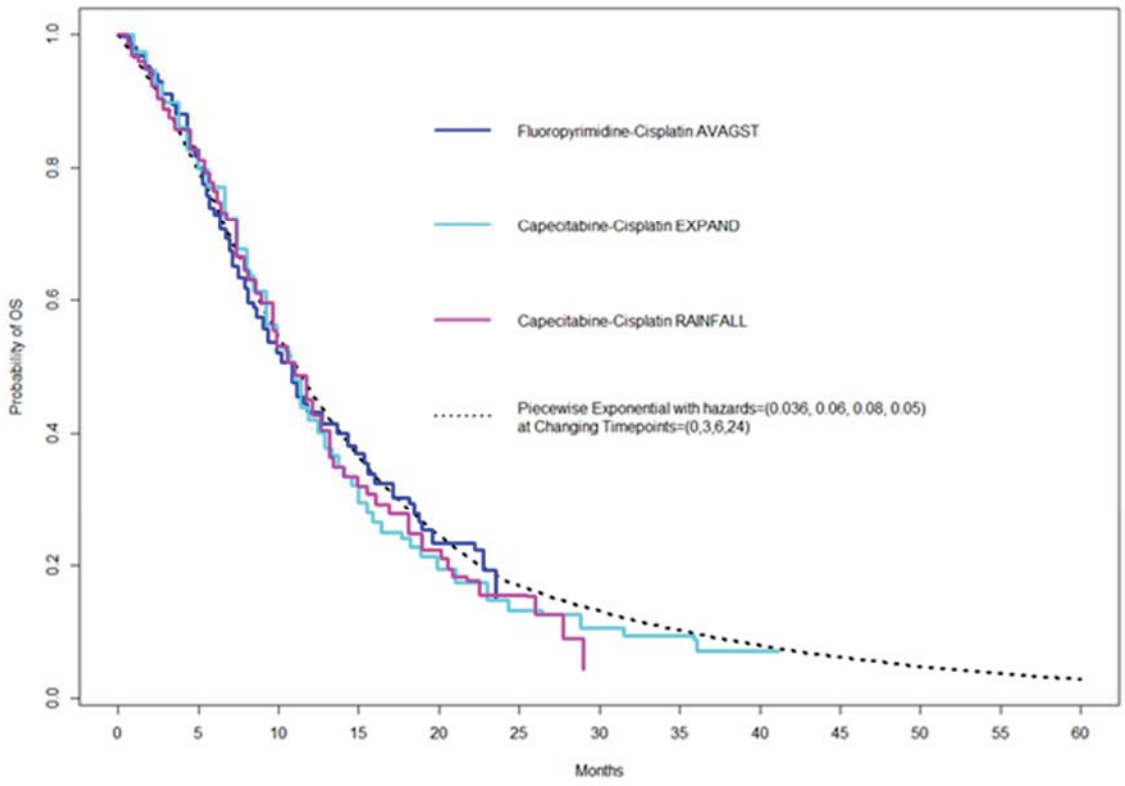


Figure 5.2-3: OS for Chemotherapy from External Data and Assumed Distribution for Chemotherapy



Statistical Software for Sample Size Calculation

Sample size calculations of the primary endpoints were based on simulations in EAST 6.4.1.

Sections 5.2.1 and 5.2.2 describe the sample size and power as stated in protocol.

5.2.1 Sample Size: Nivolumab plus Chemotherapy vs. Chemotherapy

Progression Free Survival

Concurrently randomized subjects to nivo+chemo and chemo arms with PD-L1 CPS ≥ 5 will be used for the PFS analysis. Given the delayed separation in the PFS distributions between pembrolizumab + chemotherapy and chemotherapy[4], the PFS analysis will be conducted after a minimum of 12 months follow-up after the last patient is randomized.

As such, the PFS analysis will be conducted after the following criteria are all met:

- CPS scoring by central laboratories is complete
- a minimum follow-up of 12 months

The HR is modelled as a 2-piece hazard ratio with a delayed effect of 3 and 6 months (a HR of 1 versus chemotherapy for the first 3 or 6 months) followed by a constant HR of 0.56 (see Table 5.2.1-1). With 12 months minimum follow-up it is expected that the number of events will range between 497 and 506 respectively and the power will range between approximately 99% and 60% with a Type I error of 2% (two-sided) (see Table 5.2.1-2).

The primary PFS analysis will be based on all the events observed at time of database lock.

Table 5.2.1-1: PFS Hazard Rates and Hazard Ratios (Nivolumab plus Chemotherapy vs. Chemotherapy)

Periods	Hazard Rates for Chemotherapy	Hazard Ratios (3 months delay)	Hazard Ratios (6 months delay)
0-3 months	0.09	1	1
3-6 months	0.17	0.56	1
6-10 months	0.2	0.56	0.56
> 10 months	0.13	0.56	0.56

Table 5.2.1-2: Summary of Sample Size Parameters and Schedule of Analyses for PFS (Nivolumab plus Chemotherapy vs. Chemotherapy)

	3 months Delay	6 months Delay
# subjects with PD-L1 CPS ≥ 5[a]	554	554
Hypothesized delayed period	3	6
Hypothesized HR after delayed period	0.56	0.56
Hypothesized median in control arm	5.5 months	5.5 months
Significance level (2-sided)	0.02	0.02
Enrollment Period (from start of 1:1:1)	25.4 months	25.4 months
Minimum follow-up / Expected number of events[a,b,c]	12 months /497	12 months / 506
Time of Analysis	37.4 months	37.4 months
power[b]	99%	66%
average HR[b]	0.66	0.80

[a] Based on 35% prevalence of PD-L1 CPS ≥ 5

[b] Results based on simulations.

[c] Number of events for the nivolumab plus chemotherapy vs. chemotherapy comparison in total.

Overall Survival

A delayed separation of at least 6 months between the OS distribution of pembrolizumab + chemotherapy and chemotherapy was observed.[4] To fully capture the treatment effect, the interim and final analyses of OS will be conducted after a minimum follow up after the last patient randomized.

The interim analysis will be conducted after at least 12 months minimum follow-up, and the final analysis will be conducted after at least 24 months minimum follow-up.

With 24 months minimum follow-up at final analysis it is expected that 466 events will be observed. With an average HR of 0.74 modelled as a 2-piece hazard ratio, a delayed effect with a HR of 1 versus chemotherapy for the first 6 months followed by a constant HR of 0.65 (see Table 5.2.1-3) the power will be approximately 85% with a Type I error of 3% (two-sided) (see Table 5.2.1-4).

For the OS interim and final analyses, all the events reported in the database at time of database locks will be used for the primary analyses.

Table 5.2.1-3: Assumed OS Hazard Rates and Hazard Ratios (Nivolumab plus Chemotherapy vs. Chemotherapy)

Periods	Hazard Rates for Chemotherapy	Hazard Ratio
0-3 months	0.036	1
3-6 months	0.06	1
6-24 months	0.08	0.65
>24 months	0.05	0.65

Table 5.2.1-4: Summary of Sample Size Parameters and Schedule of Analyses for OS (Nivolumab plus Chemotherapy vs. Chemotherapy)

# with PD-L1 CPS ≥ 5[a]	554
Hypothesized delayed period	6 months
Hypothesized HR after delayed period	0.65
Hypothesized median in control arm	11.1 months
Significance level (2-sided)	0.03
Enrollment Period (from start of 1:1:1)	25.4 months
INTERIM ANALYSIS #1 for OS	
Minimum follow-up/Expected number of events[a,b,c]	12 months/395
Time of analysis[d]	37.4 months

Table 5.2.1-4: Summary of Sample Size Parameters and Schedule of Analyses for OS (Nivolumab plus Chemotherapy vs. Chemotherapy)

Significance level[e]	0.0164
Power	64%
FINAL ANALYSIS	
Minimum follow-up/Expected number of events[a,b,c]	24 months/466
Time of analysis[d]	49.4 months
Significance level	0.0252
Power[b]	85%
Average HR	0.74

[a] Based on a 35% prevalence of PD-L1 CPS ≥ 5.

[b] Results based on simulations

[c] Number of events for the nivolumab plus chemotherapy vs. chemotherapy comparison in total.

[d] After first patient randomized to nivolumab + chemotherapy or chemotherapy in 1:1:1 randomization.

[e] Significance levels will be calculated based on the actual number of deaths at each interim analysis

With an assumed PD-L1 CPS ≥ 5 prevalence of 35% and under the following piecewise HR (HR of 1 vs. chemotherapy for the first 6 months followed by a constant HR of 0.65) the target number of OS events at interim and final analyses is estimated to be 395 and 466 respectively. Using the Lan-DeMets α spending function with O'Brien and Fleming type of boundary,[9] the significance levels for OS at interim and final analyses are 0.0164 and 0.0252 respectively. At time of interim analysis, the significance level will be adjusted according to actual events observed at OS interim and using the estimated final number of events. At final analysis the significance level will be calculated using the number of events in the database at time of database lock and considering the α-level already spent at interim analysis as well as the actual correlation among the test statistics.

5.2.2 Power Considerations for Nivolumab plus Ipilimumab vs. Chemotherapy

OS between nivolumab+ipilimumab vs chemotherapy will be tested as a secondary objective. This comparison will be tested if at least one primary endpoint (PFS or OS of nivolumab + chemotherapy vs. chemotherapy in PD-L1 CPS ≥ 5 subjects) meets statistical significance. The nivolumab+ipilimumab vs. chemotherapy OS comparison will be conducted at the planned final comparison of OS for nivolumab+chemotherapy vs. chemotherapy. This corresponds to approximately 36 months of minimum follow-up for the nivolumab + ipilimumab and chemotherapy arm.

With approximately 36 months minimum follow-up at final analysis it is expected that 240 events will be observed. With an average HR of 0.7 modelled as a 4-piece hazard ratio (see Table 5.2.2-

1) the power will range between 63% and 73% for Type I error of 1.5% and 3.5% (two-sided) (see Table 5.2.2-2).

Table 5.2.2-1: Assumed OS Hazard Rates and Hazard Ratios (Nivolumab plus Ipilimumab vs. Chemotherapy)

Periods	Hazard Rates for Chemotherapy	Hazard Rates for Nivo + Ipilimumab	Hazard Ratio
0-3 months	0.036	0.099	2.75
3-6 months	0.06	0.054	0.9
6-24 months	0.08	0.0328	0.41
>24 months	0.05	0.0205	0.41

Table 5.2.2-2: Summary of Sample Size Parameters and Schedule of Analyses for OS (Nivolumab plus Ipilimumab vs. Chemotherapy)

	Alpha =0.015	Alpha=0.035
# with PD-L1 CPS ≥ 5[a]	285	285
Hypothesized distribution	See Table 5.2.2-1	See Table 5.2.2-1
Hypothesized median in control arm	11.1 months	11.1 months
Significance level (two-sided)	0.015	0.035
Enrollment Period (from start of 1:1)	18.7 months	18.7 months
Time of analysis /Expected number of events[a,b,c]	54.4/240	54.4/240
Power[b]	63%	73%
Average HR[b]	0.70	0.70

[a] Based on 35% prevalence of PD-L1 CPS ≥ 5

[b] Results based on simulations.

[c] Number of events for the nivolumab plus ipilimumab vs. chemotherapy comparison in total.

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

- Baseline period:
  - Baseline evaluations or events will be defined as evaluations or events that occur before the date and time of the first dose of study treatment. Evaluations (laboratory tests, and vital signs) on the same date and time of the first dose of study treatment will be considered as baseline evaluations. Events (AEs) on the same date and time of the first dose of study treatment will not be considered as pre-treatment events.

- In cases where the time (onset time of event or evaluation time and dosing time) is missing or not collected, the following definitions will apply:
  - Pre-treatment AEs will be defined as AEs with an onset date prior to but not including the day of the first dose of study treatment;
  - Baseline evaluations (laboratory tests and vital signs) will be defined as evaluations with a date on or prior to the day of first dose of study treatment.
- If there are multiple valid assessments on or prior to the first dose of study treatment:
  - For PD-L1, among the records prior to or on first dose date (and time if collected), identify first those with quantifiable test result. If there are no records with quantifiable test result, then select those with indeterminate result ("INDETERMINATE"). If there are no records with indeterminate test result, then select those with unavailable result ("NOT EVALUABLE). If there are no records with unavailable test result, then select those with not reported or not available result (all other records). The latest record will be used as the baseline in the analyses. If there is more than one record for the latest date, then choose the one with the greatest specimen ID. This derivation will apply to tumor positive score (TPS) and composite positive score (CPS)
  - For Anti-Drug Antibody (ADA), the record related to the most recent assessment among those records where date (and time if collected) of nivolumab or ipilimumab immunoglobulin (IMG) assessment is less than or equal to the date (and time if collected) of the first nivolumab or ipilimumab dose date respectively.

- Post baseline period:
  - On-treatment AEs will be defined as AEs with an onset date and time on or after the date and time of the first dose of study treatment (or with an onset date on or after the day of first dose of study treatment if time is not collected or is missing). For subjects who are off study treatment, AEs will be included if event occurred within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of study treatment. No "subtracting rule" will be applied when an AE occurs both pre-treatment and post-treatment with the same preferred term and grade.
  - On-treatment evaluations (laboratory tests and vital signs) will be defined as evaluations taken after the day (and time, if collected and not missing) of first dose of study treatment. For subjects who are off study treatment, evaluations should be within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of study treatment.

Late-emergent drug-related AEs will be defined as drug-related AEs with an onset date greater than 100 days after the last dose of study treatment in subjects who are off study treatment.

6.2 Treatment Regimens

The treatment group "**as randomized"** will be retrieved from the IRT system

- Arm A, C in the IRT system: Experimental arm Nivolumab plus Ipilimumab (Arm A: before Amendment08; Arm C: after Amendment08)
- Arm B, D in the IRT system: Control arm Chemotherapy (Arm B: before Amendment08; Arm D: after Amendment08)

- Arm E in the IRT system: Experimental arm Nivolumab plus Chemotherapy

The treatment group "**as treated**" will be, in general, the same as the arm randomized by IRT unless a subject received the incorrect drug for **the entire period** of treatment, in which case the subject's treatment group will be defined as the incorrect drug the subject actually received.

Subjects in the nivo+chemo and in the chemotherapy arm may receive two types of chemotherapy regimen (as per investigator's choice): XELOX or FOLFOX. Chemotherapy regimen "as randomized" (referred to as "planned chemotherapy regimen") will be retrieved from the IRT system. Chemotherapy regimen "as treated" will be, in general, the same as the planned chemotherapy regimen. However, if a subject received the other chemotherapy regimen for **the entire period** of treatment, the subject's chemotherapy regimen will be defined as the regimen the subject actually received. Discrepancies between these two classifications will be reported,

Unless otherwise specified, the safety analysis will be based on the treatment group "as treated".

Unless otherwise specified, the efficacy analysis will be based on the treatment group "as randomized".

6.3 Populations for Analyses

6.3.1 Analyses of Nivolumab plus Chemotherapy versus Chemotherapy

All Enrolled Subjects: subjects who were

- randomized concurrently to nivo + chemo or chemo from 27-Mar-2017 (start date of randomization to the 3 arms)
- enrolled from 27-Mar-2017 (included) but were never randomized (enrolled means subjects who signed an informed consent form and were registered into the IRT). These subjects had equal chances to be randomized to any of the 3 arms.

All Randomized Subjects: all subjects concurrently randomized to Nivo + Chemo or Chemo from 27-Mar-2017 (included).

Randomized CPS ≥ 5 Subjects: randomized subjects with CPS ≥ 5. Similar population will be defined for other biomarker and cutoff (eg, All randomized CPS ≥ 1 Subjects).

Randomized Subjects with Measurable Disease (BICR): randomized subjects who have at least one target/measurable lesion at baseline (per BICR). Similar population will be defined based investigator assessment of baseline lesions.

Randomized CPS ≥ 5 Subjects with Measurable Disease (BICR): randomized subjects with CPS ≥ 5 who have at least one target/measurable lesion at baseline (per BICR). Similar population will be defined based investigator assessment of baseline lesions and also for other biomarker and cutoff (eg, randomized CPS≥1 Subjects).

All Treated Subjects: All randomized subjects who received at least one dose of study drug during the study.

Treated CPS ≥ 5 Subjects: randomized CPS ≥ 5 subjects who received at least one dose of study drug during the study. Similar population will be defined for other biomarker and cutoff (eg, treated CPS ≥ 1 Subjects).

PK Subjects: randomized subjects with available serum time-concentration data.

Outcome Research Subjects: randomized subjects who have an assessment at screening/baseline and at least 1 follow-up assessment.

Immunogenicity Subjects: randomized subjects who have an assessment at screening/baseline and at least 1 follow-up assessment.

Randomized Subjects with Quantifiable Biomarker: randomized subjects with quantifiable biomarker expression (eg, TPS, CPS, MSI [redacted]).

6.3.2 Analyses of Nivolumab plus Ipilimumab versus Chemotherapy

All Enrolled Subjects: subjects who were

- randomized concurrently to nivo + ipi or chemo from start of randomization
- enrolled up to 05-June-2018 (date at which the nivo + ipi arm was closed to accrual) but were never randomized (enrolled means subjects who signed an informed consent form and were registered into the IRT). These subjects had equal chances to be randomized to any of the 3 arms.

All Randomized Subjects: all subjects concurrently randomized to Nivo + Ipi or Chemo.

Randomized CPS ≥ 5 Subjects: randomized subjects with CPS ≥ 5. Similar population will be defined for other biomarker and cut-off (eg, randomized CPS ≥ 1 subjects).

Randomized Subjects with Measurable Disease (BICR): randomized subjects who have at least one target/measurable lesion at baseline (per BICR). Similar population will be defined based investigator assessment of baseline lesions.

Randomized CPS ≥ 5 Subjects with Measurable Disease (BICR): randomized subjects with CPS ≥ 5 who have at least one target/measurable lesion at baseline (per BICR). Similar population will be defined based investigator assessment of baseline lesions and also for other biomarker and cut-off (eg, randomized CPS ≥ 1 Subjects).

All Treated Subjects: All randomized subjects who received at least one dose of study drug during the study.

Treated CPS ≥ 5 Subjects: randomized CPS ≥ 5 subjects who received at least one dose of study drug during the study. Similar population will be defined for other biomarker and cut-off (eg, treated CPS ≥ 1 subjects).

PK Subjects: randomized subjects with available serum time-concentration data.

Outcome Research Subjects: randomized subjects who have an assessment at screening/baseline and at least 1 follow-up assessment.

Immunogenicity Subjects: randomized subjects who have an assessment at screening/baseline and at least 1 follow-up assessment.

Randomized Subjects with Quantifiable Biomarker: randomized subjects with quantifiable biomarker expression (eg, TPS, CPS, MSI [redacted]).

7 STATISTICAL ANALYSES

7.1 General Methods

Unless mentioned otherwise (eg, see Section 8), outputs will be provided separately for the nivo+chemo vs. chemo comparison and the nivo+ipi vs. chemo comparison. All analyses, unless otherwise noted, will be repeated for population of subjects with PD-L1 CPS ≥ 5 and for all subjects (ie, all comers). For the nivo+chemo vs. chemo comparison, analyses will be repeated for subjects with PD-L1 CPS ≥ 1 if the secondary endpoint of OS in CPS ≥ 1 randomized subjects meets the pre-specified statistical criteria.

Unless otherwise noted, the following subsections describe tabulations of discrete variables, by the frequency and proportion of subjects falling into each category, grouped by treatment (with total, as needed). Percentages given in these tables will be rounded to the first decimal and, therefore, may not always sum to 100%. Continuous variables will be summarized by treatment group (with total, as needed) using the mean, standard deviation, median, minimum and maximum values, unless specified otherwise.

Time-to-event variables (eg, time-to resolution) will be analyzed using the Kaplan-Meier technique. When specified, the median will be reported along with 95% CI using Brookmeyer and Crowley method[10] (using log-log transformation for constructing the confidence intervals).[11]

The conventions to be used for imputing missing and partial dates for analyses requiring dates are described in Section 8.

7.1.1 Adverse Events, Serious Adverse Events, Multiple Events, Select Adverse Events, Other Events of Special Interest and Immune-Mediated Adverse Events

Drug-related AEs are those events with relationship to study drug "Related", as recorded on the CRF. If the relationship to study drug is missing, the AE will be considered as drug-related.

Serious adverse events consist of AEs deemed serious by the Investigator and flagged accordingly in the CRF and clinical database.

Adverse events leading to study drug discontinuation are AEs with action taken regarding study drug(s) = "Drug was discontinued".

Adverse events leading to dose delay are AEs with action taken regarding study drug(s) = "Drug was delayed".

Adverse events leading to dose reduction are AEs with action taken regarding study drug(s) = "Dose was reduced".

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), and the most recent version of the dictionary at the time of the database lock will be used. Adverse events results will be graded for severity using NCI Common Terminology Criteria for Adverse Events (CTCAE) and the most recent version of the criteria at the time of the database lock will be used.

In the AE summary tables, unless otherwise specified, subjects will be counted only once at the Preferred Term (PT), only once at the System Organ Class (SOC), and only once at subject level for the counting of total number of subjects with an AE. The AE tables will be sorted by the SOCs and then PTs. SOC will be ordered by descending frequency overall and then alphabetically. PTs will be ordered within SOC by descending frequency overall and then alphabetically. The sorting will be done based on the 'Any Grade' column of the experimental arm when arms are presented side-by-side.

Unless otherwise specified, the AE summary tables will be restricted to on-treatment events regardless of the causality.

Analyses that take into account the multiple occurrences of a given adverse event will be conducted (see Section 7.6.9). To prepare these analyses, the CRF data will be processed according to standard BMS algorithms[12] in order to collapse adverse event records into unique records based on the preferred term. These data will be presented as the rate per 100 person-years of exposure. These analyses will take into account all on-treatment events (allowing more than 1 event per subject) and the total exposure time. The person-year exposure will be computed as the sum over the subjects' exposure expressed in years where the exposure time is defined as

- (Date of last dose of study treatment - date of first dose of study treatment + 31 days (or 101 days, depending on the analysis))/365.25, for subject who are off study treatment and were followed for at least 30 days (or 100 days, depending on the analysis) after last dose of study treatment.
- (Last known alive date - date of first dose of study treatment +1)/365.25, for subjects who are still on-treatment or who are off study treatment and were followed less than 30 days (or 100 days depending on the analysis) after last dose of study treatment.

7.1.1.1 Select Adverse Events (EU Submission)

The select Adverse Events (select AEs) consist of a list of preferred terms grouped by specific category (eg, pulmonary events, gastrointestinal events categories, etc.). AEs that may differ from or be more severe than AEs caused by non-immunotherapies and AEs whose early recognition and management may mitigate severe toxicity are included as select AEs. Categories of select AEs may include subcategories (eg, adrenal disorders, diabetes, pituitary disorders, and thyroid disorders are subcategories of the endocrine event category).

The list of MedDRA preferred terms used to identify select adverse events is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock will be provided by categories/subcategories.

In addition to the frequency and worst severity of select AEs, time-to onset, time-to resolution, and time-to resolution where immune modulating medication was initiated will be analyzed for each specific category/subcategory of drug-related select AEs when applicable.

Further details on the definitions of select adverse event, time-to onset and time-to resolution are described in APPENDIX 1.

7.1.1.2 Other Events of Special Interest

Other events of special interest (OEOSI) consist of a list of preferred terms grouped by specific category (eg, Myositis Event, Myocarditis Event, Demyelination Event, Guillain-Barre Syndrome, Pancreatitis Event, Uveitis Event, Encephalitis Event, Myasthenic Syndrome, Rhabdomyolysis Event, Graft Versus Host Disease). The list of MedDRA preferred terms used to identify OEOSI is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock by categories will be provided.

7.1.1.3 Immune-Mediated Adverse Events (US Submission)

In order to further characterize AEs of special clinical interest, analysis of immune-mediated AEs (IMAE) will be conducted. IMAEs are specific events (or groups of PTs describing specific events) that include pneumonitis, diarrhea/colitis, hepatitis, nephritis/renal dysfunction, rash, endocrine (adrenal insufficiency, hypothyroidism/thyroiditis, hyperthyroidism, diabetes mellitus, and hypophysitis), and other specific events, considered as potential immune-mediated events by investigator that meet the definition summarized below:

- those occurring within 100 days of the last dose,
- regardless of causality,
- treated with immune-modulating medication (of note, endocrine AEs such as adrenal insufficiency, hypothyroidism/thyroiditis, hyperthyroidism, diabetes mellitus, and hypophysitis are considered IMAEs regardless of immune-modulating medication use, since endocrine drug reactions are often managed without immune-modulating medication).
- with no clear alternate etiology based on investigator assessment, or with an immune-mediated component

The list of MedDRA preferred terms used to identify IMAEs is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock by categories will be provided.

7.1.2 Laboratory Tests

Clinical laboratory parameters (hematology, serum chemistry and electrolytes) will be evaluated.

Laboratory tests will be graded using the NCI Common Terminology Criteria, and the most recent version of the criteria at the time of the database lock will be used.

Clinical laboratory data will be analyzed using International System of Units (SI) and will be repeated using US conventional units.

In the laboratory summary tables, unless otherwise specified, subjects will be counted only once for each lab parameter according to their worst on treatment CTC grade (worst being the highest CTC grade). The laboratory tables and listings will be sorted by laboratory category, laboratory subcategory and laboratory test code sequence number.

7.1.3 Immunogenicity Data

Blood samples for immunogenicity analysis will be collected from subjects assigned to the experimental treatment group(s) according to the protocol schedule. Samples will be evaluated for development of Anti-Drug Antibody (ADA) by a validated electrochemiluminescent (ECL) immunoassay.

7.1.4 Efficacy

General Methods for Time-to-event Variables

Time to event distribution (eg, OS, PFS, DoR) will be estimated using Kaplan-Meier (KM) techniques.

Median survival time along with 95% (or adjusted) confidence interval (CI) will be constructed based on a log-log transformed CI for the survivor function S(t).[10,11] Rates at fixed timepoints (eg, OS at 6 months) will be derived from the KM estimate and corresponding CI will be derived based on Greenwood formula[13] for variance derivation and on log-log transformation applied on the survivor function S(t).[14]

Unless otherwise specified, the stratified log-rank test will be performed to test the comparison between time to event distributions. Unless otherwise specified, the stratified HR between 2 treatment groups along with CI will be obtained by fitting a stratified Cox model with the treatment group variable as unique covariate.

7.2 Study Conduct

A by-subject listing of batch numbers for all treated subjects will be provided.

7.2.1 Accrual

The accrual pattern will be summarized per region, country, investigational site and month. Randomization date, first dosing date, country, investigational site will be presented in a by subject listing of accrual. These analyses will be performed on the all enrolled subjects population only.

7.2.2 Relevant Deviations

The relevant Protocol Deviations will be summarized on randomized CPS ≥ 5 subjects and on all randomized subjects by treatment group and overall. The following programmable deviations from inclusion and exclusion criteria will be considered as relevant protocol deviations. Non-programmable relevant eligibility and on-treatment protocol deviations, as well as significant (both programmable and nonprogrammable) eligibility and on-treatment protocol deviations will be reported through [redacted] listings.

At Entrance:

- Subjects without inoperable, advanced or metastatic GC or GEJ or distal esophageal carcinoma and have histologically confirmed predominant adenocarcinoma.
- Subjects who have received prior systemic therapy (adjuvant and neoadjuvant is allowed).
- Subject with baseline ECOG performance status > 1.
- Subjects without any disease at baseline.
- Subjects without any PD-L1 result. ("Indeterminate" is not a deviation).

On-study:

- Subjects receiving concurrent anti-cancer therapy (surgery, chemotherapy, hormonal therapy, immunotherapy, non-palliative radiation therapy, standard or investigational agents for treatment of gastric cancer).
- Subjects treated differently as randomized (subjects who received the wrong treatment, excluding the never treated).

A subject listing will also be produced.

7.3 Study Population

7.3.1 Subject Disposition

The total number of subjects enrolled (randomized or not randomized) will be presented along with the reason for not being randomized. A subject listing for subjects not randomized will also be provided, showing the subject's race, gender, age, consent date and reason for not being randomized. The number of subjects enrolled but not randomized will be reported for the 3 arms not knowing to which arm the subject would have been randomized to. Therefore, the number of screen failures will be overestimated when reporting separately the comparison of nivo+chemo vs. chemo and the comparison of nivo+ipi vs. chemo.

Number of subjects randomized but not treated along with the reason for not being treated will be tabulated by treatment group as randomized.

Number of subjects who discontinued study treatment along with corresponding reason will be tabulated by treatment group as treated. For the nivo+ipi arm, number of subjects who discontinued study treatment during the nivo-plus-ipi phase and during the nivo mono phase (along with corresponding reason) will be also tabulated. Reason for discontinuation will be derived from subjects status CRF page.

A by-subject listing for all treated subjects will be provided showing the subject's off treatment date and whether the subject continue in the treatment period long with the reason for going off treatment period.

7.3.2 Demographics and Other Baseline Characteristics

The following baseline characteristics will be summarized by treatment group as randomized and overall. All baseline presentations identify subjects with missing measurements. Listings will also be provided.

In addition to the presentations defined previously in Section 6.3 (ie, for N+C vs C and N+I vs C), these baseline characteristics will be summarized to support the Contribution of Components (Section 8) by presenting this analysis for the All Concurrently Randomized Subjects to 3 Arms group (N~1098, see Figure 8.1-1) and the All Randomized Subjects to 3 Arms group (N~2031).

- Age (continuous)
- Age category (< 65, ≥ 65 and < 75, ≥ 75 and < 85, ≥ 85; ≥ 75, ≥ 65)
- Sex (male, female)
- Race (White, Black or African American, Asian [Asian Indian, Chinese, Japanese, Asian Other], American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Region (Asia [including China], US, RoW, Asia[excluding China], China)
- ECOG PS
- Weight (descriptive statistics)
- Primary tumor location at initial diagnosis (GC, GEJ cancer, or Esophageal adenocarcinoma)
- Disease stage at initial diagnosis (Stage I, Stage II, Stage III, Stage IV)
- Disease status (locally recurrent, metastatic, locally advanced)
- Lauren classification (intestinal type, diffuse type, mixed, unknown)
- Siewert-Stein classification (reported for GEJ only) (type I, type II, type III)
- WHO histologic classification (adenosquamous carcinoma, mucinous adenocarcinoma, papillary serous adenocarcinoma, signet ring cell, tubular adenocarcinoma, other)
- TNM classification:
  - Tumor (Tx, T0, Tis, T1, T2, T3, T4, unknown)
  - Nodes (Nx, N0, N1, N2, N3, unknown)
  - Metastasis (Mx, M0, M1, unknown)
- Smoking status (current/former, never smoker, unknown)
- All lesions at baseline (investigator and BICR): sites of disease, number of disease sites per subject, number of subjects with no measurable lesion
- Target lesions (investigator and BICR): presence of target lesions, site of target lesion, sum of diameters of target lesions
- CNS metastases (yes/no)
- Liver metastases (yes/no)
- Peritoneal metastases (yes/no)
- Time from Initial Disease Diagnosis to Randomization (< 6 months, 6 months - < 1 year, 1 - < 2 year, 2 - < 3 year, 3 -< 4 year, 4 - < 5 year, ≥ 5 year)

- Time from randomization to first dose date (≤ 3, 4 - 5, 6 - 7, 8 - 14, 15 - 21, > 21 days, Not Reported or Not Treated)
- HER-2 status at study entry (negative, positive, unknown)
- Microsatellite instability (MSI-H, MSS, invalid, unknown)
- H.pylori (no, yes, unknown)
- Baseline hemoglobin (< 10g/dL, ≥ 10g/dL)
- Baseline albumin (< LLN, ≥ LLN)

A listing of randomization scheme presenting randomized treatment group and as treated treatment group will be provided for all randomized subjects.

7.3.3 Medical History

A by-subject listing of general medical history will be provided for all randomized subjects.

7.3.4 Prior Therapy Agents

Prior cancer therapy will be summarized by treatment group as randomized and overall.

Prior anti-cancer therapy

- Setting of prior systemic therapy regimen received (adjuvant, metastatic disease, neo-adjuvant).
- Time from completion of prior adjuvant/neo-adjuvant therapy to treatment (for subjects who received prior adjuvant/neo-adjuvant therapy), (< 6 months, 6 - < 12month, ≥ 12 months)
- Prior surgery related to current cancer (yes or no).
- Prior radiotherapy (yes or no).
- Prior systemic therapy classified by therapeutic class and generic name.

7.3.5 Physical Examinations

Subjects with abnormal baseline physical examination will be listed by subject for all randomized subjects.

7.3.6 Baseline Physical Measurements

Baseline physical measurements will be listed by subject for all randomized subjects.

7.3.7 Discrepancies between IRT Stratification Factors and Other Datasets

Summary tables (cross-tabulations), by treatment group as randomized, for stratification factor (except for region) will be provided to show any discrepancies between what was reported through IRT vs. other data sources (CRF data and laboratory data) at baseline. This analysis will be provided for randomized CPS ≥ 5 subjects and all randomized subjects.

- PD-L1 expression level (≥ 1% vs. < 1% or indeterminate) (IRT vs. laboratory data)
- ECOG performance status (0 vs. 1) (IRT vs. CRF)

- Planned chemotherapy regimen (XELOX vs. FOLFOX) (randomization in IRT vs. actual treatment received) – only for subjects in the nivo+chemo and in the chemotherapy arms

7.4 Extent of Exposure

Analyses in this whole section will be performed by treatment group as treated for treated CPS ≥ 5 subjects and for all treated subjects. For chemotherapy (FOLFOX or XELOX) the Chemotherapy regimen "as treated" will be reported (see Section 6.2).

7.4.1 Administration of Study Therapy

For details on the dosing schedule per protocol, see Section 2.1.

The following parameters will be summarized (descriptive statistics) by treatment group:

- Number of cycles received (summary statistics).
- Duration of treatment
  - using a KM curve whereby the last dose date will be the event date for those subjects who are off study therapy. Subjects who are still on study therapy will be censored on their last dose date. Median duration of treatment and associated 95% CI will be provided.

The following parameters will be summarized (descriptive statistics) by study therapy and treatment group:

- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥110%.
- Cumulative dose

A by-subject listing of dosing of study medication (record of study medication, infusion details, dose change).

Key parameters used to characterize dosing data for the nivo+ipi arm are defined in Table 7.4.1-1; for subjects treated with XELOX in the nivo+chemo and chemotherapy arms in Table 7.4.1-3; for subjects treated with FOLFOX in the N+C and chemotherapy arms in Table 7.4.1-5.

Table 7.4.1-1: Administration of Study Therapy in the Nivolumab plus Ipilimumab Arm

	Nivolumab	Ipilimumab
Dosing schedule per protocol	1mg/kg Q3Wx4 then 240mg Q2W, IV Or 240mg Q2W, IV or 480 Q4W, IV	3mg/kg Q3W, IV
Dose*	Total dose administered (mg)/most recent weight (kg) Or Total dose administered (mg)	Total dose administered (mg)/most recent weight (kg)
Cumulative Dose	The sum of all doses (mg) administered to a subject during the treatment period	The sum of all doses (mg/kg) administered to a subject during the treatment period
Relative Dose Intensity (%)	See equation in Table 7.4.1-2	100x [Cum dose (mg/kg) / [(Last ipilimumab dose date in the nivo-plus-ipi phase – ipilimumab Start dose date + 21) x 3 / 21]]
Duration of Treatment	Last dose date (of the last administered study therapy) - Start dose date (of the first administered study therapy) + 1	

* Dose administered in mg at each dosing date and weight are collected on the CRF

Table 7.4.1-2: Formula for Relative Dose Intensity for Nivolumab in Nivolumab plus Ipilimumab Regimen

Overall duration of Nivo treatment	*(last dose date of nivo - start dose date of nivo +X)* *X= 14 days if last dose is 240 mg* *X= 21 days if last dose is 360 mg* (some patients take 360 mg despite protocol instruction) **X= 28 days if last dose is 480 mg**
*Overall RDI**	$\mathbf{w_1 * RDI_1 + (1 - w_1) * RDI_2}$
$\boldsymbol{RDI_1}$	**100 x[Cum dose (mg/kg))in nivo-plus-ipi phase / [(Last nivolumab dose date in the nivo-plus-ipi phase – nivolumab Start dose date + X) x 1 / 21]]**
$\boldsymbol{RDI_2}$	$\mathbf{100} \times \frac{\boldsymbol{cumulative\ dose\ in\ mono\ phase\ (mg)}}{\left\{\left[(\boldsymbol{nivo\ last\ dose\ date\ in\ mono\ phase - start\ date\ in\ the\ nivo\ mono\ phase + X}) \times \frac{\mathbf{120}}{\mathbf{7}}\right]\right\}}$ *X as above*
$\boldsymbol{w_1}$	*[(First nivolumab dose date in the mono phase - nivolumab Start dose date in nivo-plus-ipi phase)]/overall duration of Nivo treatment*

***** if patient did not start the combo phase then overall *RDI= RDI*$_1$

Table 7.4.1-3: Administration of Study Therapy for Subjects Treated with XELOX

	Nivolumab	**Oxaliplatin**	**Capecitabine**
Dosing schedule per protocol	360mg Q3WK, IV or 480 mg Q4WK, IV (nivo mono)	130mg/m^2 Q3WK, IV	1000mg/m^2 twice daily D1-14, Q3WK, oral
Dose*	Total dose administered (mg)	Total dose administered (mg) / most recent BSA	Total dose administered (mg) / most recent BSA
Cumulative Dose	The sum of all doses (mg) administered to a subject during the treatment period	The sum of all doses (mg/m^2) administered to a subject during the treatment period	The sum of all doses (mg/m^2) administered to a subject during the treatment period

Table 7.4.1-3: Administration of Study Therapy for Subjects Treated with XELOX

	Nivolumab	Oxaliplatin	Capecitabine
Relative Dose Intensity (%)	see formula in Table 7.4.1-4	100x[Cumulative dose (mg/m^2) / [Last oxaliplatin dose date – oxaliplatin Start dose date + 21) x 130/21]]	100x[Cumulative dose (mg/m^2) / [First dose of capecitabine in the last cycle – capecitabine Start dose date + 21) x 28000/21]]
Duration of Treatment	Last dose date (of the last administered study therapy) - Start dose date (of the first administered study therapy) + 1		

* Dose administered in mg at each dosing date and BSA (computed using recent weight and baseline height) are collected on the CRF.

Table 7.4.1-4: Formula for Relative Dose Intensity for Nivolumab in Nivolumab plus XELOX Regimen:

$$\mathbf{100} \times \frac{\boldsymbol{cumulative\ dose\ (mg)}}{\left\{\begin{array}{l}\left[(\boldsymbol{last\ dose\ date\ of\ nivo\ 360\ mg - start\ date\ of\ nivo} + \mathbf{21}) \times \frac{\mathbf{360}}{\mathbf{21}}\right] \\ + \left[(\boldsymbol{last\ dose\ date\ of\ nivo\ 480\ mg - start\ date\ of\ nivo\ 480mg} + \mathbf{28}) \times \frac{\mathbf{480}}{\mathbf{28}}\right]\end{array}\right\}}$$

Table 7.4.1-5: Administration of Study Therapy for Subjects Treated with FOLFOX

	Nivolumab	Oxaliplatin	Leucovorin	Fluorouracil
Dosing schedule per protocol	240mg Q2WK, IV or 480 mg Q4WK, IV (nivo mono)	85mg/m^2 Q2WK	400mg/m^2 Q2WK	**Bolus 5-FU:** 400mg/m^2 Q2WK, IV on Day 1 of the cycle **Continuous 5-FU:** 1200mg/m^2 Q2WK, IV continuous infusion on Day1-2 of the cycle
Dose*	Total dose administered (mg)	Total dose administered (mg) / most recent BSA	Total dose administered (mg) / most recent BSA	Total dose administered (mg) / most recent BSA

Table 7.4.1-5: Administration of Study Therapy for Subjects Treated with FOLFOX

	Nivolumab	Oxaliplatin	Leucovorin	Fluorouracil
Cumulative Dose	The sum of all doses (mg) administered to a subject during the treatment period	The sum of all doses (mg/m^2) administered to a subject during the treatment period	The sum of all doses (mg/m^2) administered to a subject during the treatment period	The sum of all doses (mg/m^2) administered to a subject during the treatment period
Relative Dose Intensity (%)	See formula in Table 7.4.1-6	100 x [Cumulative dose (mg/m^2) / [Last oxaliplatin dose date – oxaliplatin Start dose date + 14) x 85/14]]	100 x [Cumulative dose (mg/m^2) / [Last leucovorin dose date – leucovorin Start dose date + 14) x 400/14]]	**Bolus 5-FU:** 100 x [Cumulative dose (mg/m^2) / [First dose of fluorouracil in the last cycle – fluorouracil Start dose date + 14) x 400/14]] **Continuous 5-FU:** 100 x [Cumulative dose (mg/m^2) / [First dose of continuous fluorouracil in the last cycle – continuous fluorouracil Start dose date + 14) x 2400 / 14]]
Duration of Treatment	Last dose date (of the last administered study therapy) - Start dose date (of the first administered study therapy) + 1			

* Dose administered in mg at each dosing date and BSA (computed using recent weight and baseline height) are collected on the CRF.

Table 7.4.1-6: Formula for Relative Dose Intensity for Nivolumab Nivolumab plus FOLFOX Regimen

$$\mathbf{100} \times \frac{\boldsymbol{cumulative\ dose\ (mg)}}{\left\{\begin{array}{l}\left[(\boldsymbol{last\ dose\ date\ of\ nivo\ 240\ mg - start\ date\ in\ the\ nivo\ mono\ phase} + \mathbf{14}) \times \frac{\mathbf{240}}{\mathbf{14}}\right] \\ \quad + \left[(\boldsymbol{last\ dose\ date\ of\ nivo\ 480\ mg - start\ date\ of\ nivo\ 480mg} + \mathbf{28}) \times \frac{\mathbf{480}}{\mathbf{28}}\right]\end{array}\right\}}$$

7.4.2 Modification of Study Therapy

7.4.2.1 Dose Delays

A dose will be considered as actually delayed if the delay is exceeding the number of days in Table 7.4.2.1-1.

Table 7.4.2.1-1: Dose Delays

	Ipilimumab	Q3W Nivolumab (a)	Q2W Nivolumab (b)	Q4W Nivolumab (c)	XELOX (d)	FOLFOX (d)
	Specifications per Protocol					
Dose considered as actually delayed	If the delay is exceeding 3 days	If the delay is exceeding 3 days	If the delay is exceeding 2 days	If the delay is exceeding 3 days	If the delay is exceeding 3 days	If the delay is exceeding 2 days
Maximum delay allowed between doses	42 days	42 days	42 days	42 days	Not specified	Not specified
	Definitions for the Analysis					
Dose Delay	duration of preceding cycle in days – 21	duration of preceding cycle in days – 21	duration of preceding cycle in days – 14	duration of preceding cycle in days – 28	duration of preceding cycle in days – 21	duration of preceding cycle in days – 14
Categories of dose delays	on-time, 4 - 7 days, 8 - 14 days, 15 - 42 days, > 42 days	on-time, 4 - 7 days, 8 - 14 days, 15 - 42 days, > 42 days	on-time, 3 – 4 days, 5 - 7 days, 8 - 14 days, 15 - 42 days, > 42 days	on-time, 4 - 7 days, 8 - 14 days, 15 - 42 days, > 42 days	on-time, 4 - 7 days, 8 - 14 days, 15 - 42 days, > 42 days	on-time, 3 – 4 days, 5 - 7 days, 8 - 14 days, 15 - 42 days, > 42 days
	(a) nivo-plus-ipi phase in the nivo+ipi arm, FOLFOX-treated subjects in the nivo+chemo arm (b) nivo mono phase in the nivo+ipi arm, XELOX-treated subjects in the nivo+chemo arm (c) applicable to all arms with nivolumab, when nivolumab administered alone after treatment with nivolumab in combination with ipilimumab, FOLFOX or XELOX. (d) in the nivo+chemo and in the chemotherapy arms					

The following parameters will be summarized by study therapy and treatment group:

- Number of subjects with at least one dose delayed, the number of doses delayed per subject, the reason for dose delay and the length of dose delay.

7.4.2.2 Infusion Interruptions and Rate Changes

Each study therapy given as infusion can be interrupted and/or the IV infusion rate can be reduced. This information will be retrieved from CRF dosing pages.

For capecitabine, which is oral, interruption will be programmatically derived.

The following parameters will be summarized by study therapy and treatment group:

- Number of subjects with at least one dose infusion interruption, the reason for interruption, and the number of infusion interruptions per subject.
- Number of subjects with at least one IV infusion rate reduction, the reason for reduction and the number of infusion with IV rate reduction per subject (not for capecitabine).

7.4.2.3 Dose Reductions

Per protocol, there will be no dose escalations or reductions of nivolumab and ipilimumab. However, for chemotherapy, dose reduction is permitted according to local standard or local package insert. This information will be retrieved from CRF dosing pages.

For subjects treated with chemotherapy, the following will be summarized by treatment group:

- Number of subjects with at least one dose reduction along with the reason for the dose reduction.

7.4.3 Concomitant Medications

Concomitant medications, defined as medications other than study medications which are taken at any time on-treatment (ie, on or after the first day of study therapy and within 100 days following the last dose of study therapy), will be coded using the WHO Drug Dictionary.

The following summary tables will be provided:

- Concomitant medications (subjects with any concomitant medication, subjects by medication class and generic term).

A by-subject listing will accompany the table.

Prior medications, defined as non-study medications with a start date before consent date, and current medications, defined as non-study medications with a start date before the first date of study medication and stop date after consent date, will be coded using the UMC WHO Drug Global Dictionary.

The following summary table will be provided:

- Prior/current medications (subjects with any prior/current medication, subjects by medication class and generic term)

By-subject listings will accompany the tables.

7.4.3.1 Immune Modulating Medication

Immune modulating concomitant medications are medications entered on an immune modulating medication form or available from the most current pre-defined list of immune modulating

medications. The list of anatomic class, therapeutic class and generic name used for the selection at the time of the database lock will be provided.

The percentage of subjects who received immune modulating concomitant medication for

- management of adverse event
- premedication
- other use
- any use
- management of drug-related select adverse event (any grade, grade 3-5) by select AE category/ subcategory
- management of IMAEs (any grade, grade 3-5) by IMAE categorywill be reported separately for each treatment group (percentages of treated subjects by medication class and generic term).

For each category/subcategory of drug-related select AEs (any grade, grade 3-5) and IMAEs (any grade, grade 3-5), the following will be reported for each treatment group:

- The total immune modulating medication treatment duration (excluding overlaps), duration of high dose of corticosteroid, initial dose of corticosteroid, and tapering duration (summary statistics)

Duration represents the total duration the subject received the concomitant medication of interest. If the subject took the medication periodically, then DURATION in the summation of all use. Initial dose represents the dose of the concomitant medication of interest received at the start of the event. In the case multiple medications started on the same date, the highest equivalent dose is chosen and converted to mg/kg by dividing by the subject's recent weight.

These analyses, except the ones related to IMAEs will be conducted using the 30-day safety window. The analyses related to IMAEs will be conducted using the 100-day safety window.

7.4.3.2 Subsequent Therapy

Number and percentage of subjects receiving subsequent therapies will be summarized for randomized subjects with PD-L1 CPS≥5 subjects and all randomized subjects. Categories include:

- Immunotherapy including commercial Nivolumab (anti-PD1 agents, anti-PD-L1 agents, anti-CTLA-4 agents and others) by drug name
- Other anti-cancer agents excluding all immunotherapy (approved and investigational) by drug name
- Surgery for treatment of tumors (ie reason of tumor resection palliative or tumor resection curative)
- Radiotherapy for treatment of tumors (ie, radiotherapy with reason curative or palliative)
- Any combination of the above

A subject listing of follow-up therapy will also be produced for subjects who had any subsequent therapy.

7.5 Efficacy

Analyses of progression free survival and objective response rate will be based on the Blinded Independent Central Review (BICR) evaluation, unless noted otherwise.

Analyses in this section will be tabulated for all randomized subjects by treatment group as randomized, unless otherwise specified.

Unless stated otherwise, whenever a stratified analysis is specified, the following stratifications factors (recorded at randomization as per IRT) will be used see Table 7.5-1

Table 7.5-1: Stratification Factors Used for Stratified Analyses

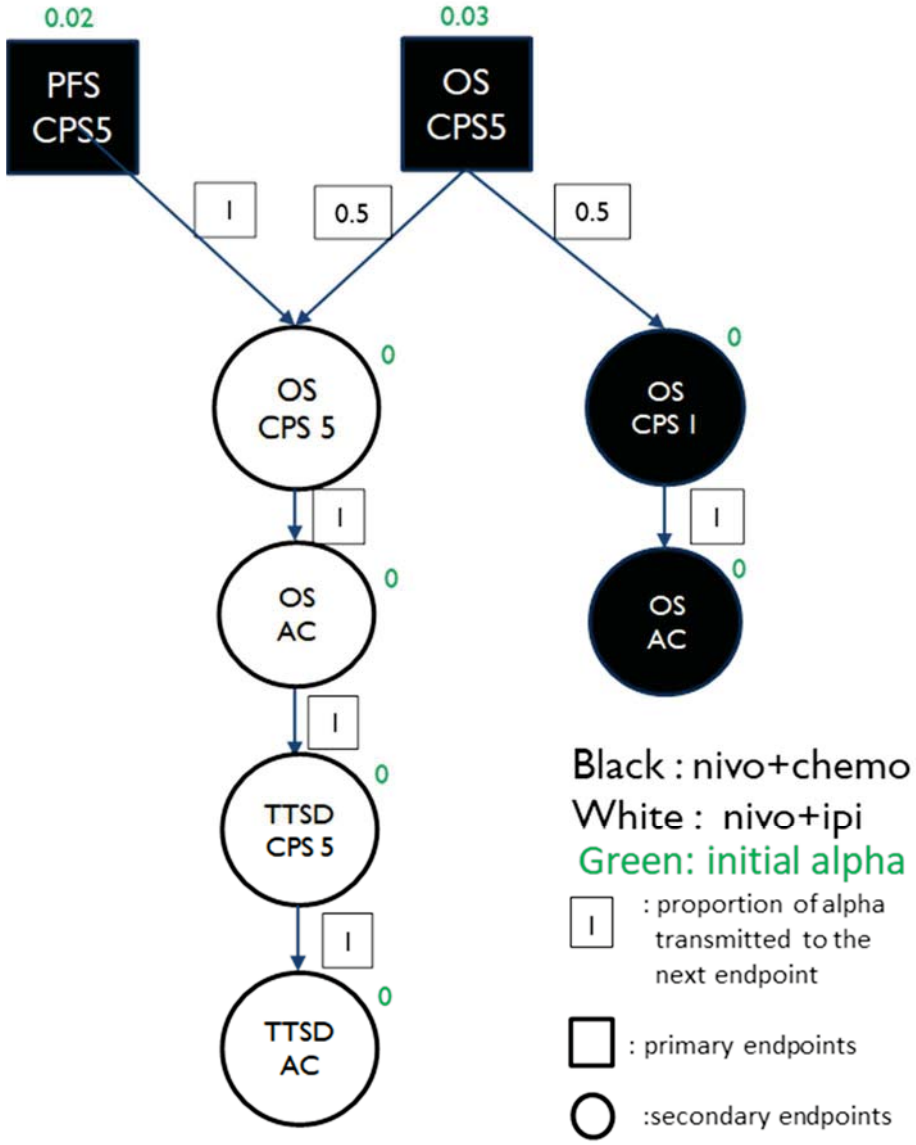
	Nivo+ipi vs. Chemotherapy	Nivo+chemo vs. Chemotherapy
	• Region (Asia vs. US vs. RoW) • ECOG performance status (0 vs. 1) • PD-L1 expression level (≥ 1% vs. < 1% or indeterminate)	• Region (Asia vs. US vs. RoW) • ECOG performance status (0 vs. 1) • PD-L1 expression level (≥ 1% vs. < 1% or indeterminate) • Chemotherapy (XELOX vs. FOLFOX)

7.5.1 Strong Control of Type I Error for Primary and Secondary endpoints:

The following testing strategy guarantees a strong control of FWER at level α of 5% for the primary and key secondary endpoints.

For the 2 primary endpoints of PFS and OS in the comparison of nivo+chemo vs. chemo in randomized CPS ≥ 5 subjects, 2% is allocated to PFS and 3% is allocated to OS. For the first secondary endpoint of OS in nivo+ipi vs. chemo in randomized CPS ≥ 5 subjects, alpha will be passed from both PFS and OS as shown in Figure 7.5.1-1, the other secondary endpoints formally tested are also illustrated in the figure[15].

Figure 7.5.1-1: Graphical Approach for Primary and Secondary Endpoints



In the graph, each vertex corresponds to a hypothesis to be tested. The square represent the hypotheses for the primary endpoints and the circles for secondary endpoints, light blue color represent comparison for nivo+ipi vs. chemo and dark blue represent comparison of nivo+chemo vs. chemo. The numbers next to the vertices are the initially allocated (endpoint-specific) significance levels. It is 0 for all secondary endpoints meaning that they cannot be tested until alpha is passed to them from the primary endpoints. The weight (in rectangular frame) associated with a directed edge (line) between any two vertices indicates the fraction of the (local)

significance level at the initial vertex that is added to the significance level at the terminal vertex, if the hypothesis at the tail is rejected.

Table 7.5.1-1 shows the significance levels for the primary and key secondary endpoints. For endpoints which have an interim analysis, the significance level will be distributed using the Lan-DeMets alpha spending function with O'Brien-Fleming boundaries. The nominal significance levels for OS in randomized PD-L1 CPS ≥ 5 subjects in Table 7.5.1-1 are those obtained with the number of events exactly as in the table and assuming a prevalence rate of 35% for PD-L1 CPS ≥ 5. At time of analysis, the significance levels will be adjusted according to actual number of events and using the estimated final number of events. For interim analyses of OS in randomized PD-L1 CPS ≥ 1 subjects and all randomized subjects, the nominal significance levels will be obtained using the same information fraction as for randomized PD-L1 CPS ≥ 5 subjects. At the final analysis the significance level will be calculated using the number of events in the database at time of database lock and considering the α-level already spent at interim analysis as well as the actual correlation among the test statistics.

Table 7.5.1-1: Significance Levels for Primary and Secondary Endpoints

Analysis Timepoint	PFS Nivo+chemo in CPS ≥ 5		OS Nivo+chemo in CPS ≥ 5		OS Nivo+ipi in CPS ≥ 5	
	Event	Sign. level	Event	Sign. level	Event	Sign. level
12 months follow-up	497[a]	0.02	395[b]	0.0164	NA	NA
24 months follow-up	NA	NA	466[b]	0.0252	240[c]	0.015[d] 0.02 0.035

[a] Estimated based on 2-piece hazard ratio with HR of 1 for the first 3 months followed by constant HR of 0.56 and assuming 35% prevalence for PD-L1 CPS ≥ 5.

[b] Estimated based on 2-piece hazard ratio with HR of 1 for the first 6 months followed by constant HR of 0.65 and assuming 35% prevalence for PD-L1 CPS ≥ 5.

[c] Estimated based on hazard ratios as in Table 5.2.1-3

[d] The significance level will depend on whether only one primary endpoint was positive or both. If only OS of nivo+chemo in CPS ≥ 5 is positive then the significance level is 0.015, if only PFS of nivo+chemo in CPS ≥ 5 is positive then the significance level is 0.02 and if both are positive then the significance level is 0.035. The other secondary endpoints for nivo+ipi vs. chemo (OS in all randomized and TTSD) will be tested hierarchically at the same significance level as OS nivo+ipi in CPS ≥ 5 subjects.

At the first analysis timepoint, PFS final analysis and OS interim analysis for nivo+chemo vs. chemo in CPS ≥ 5 subjects will be tested.

- If none of the primary endpoints can be rejected then no secondary endpoints are tested, and the study continues to proceed to the final OS analysis.
- If PFS is rejected, its entire alpha is passed to OS of nivo+ipi vs. chemo in PD-L1 CPS ≥ 5 subjects.
- If OS is rejected at interim analysis, then its alpha (0.03) will be shared equally to OS of nivo+ipi vs. chemo in PD-L1 CPS ≥ 5 subjects and to OS of nivo+chemo vs. chemo in PD-L1 CPS ≥ 1 subjects.

At the second analysis timepoint, OS final analysis for nivo+chemo vs. chemo in CPS ≥ 5 subjects will be tested.

- If this endpoint cannot be rejected then none of the other secondary endpoints for nivo+chemo will be tested. OS for nivo+ipi vs. chemo in CPS ≥ 5 subjects can be tested only if PFS for nivo+chemo in CPS ≥ 5 subjects was positive, the testing will be done at the significance level of 0.02.
- If OS of nivo+chemo vs. chemo in PD-L1 CPS ≥ 5 subjects is statistically significant then its alpha (0.03) will be shared equally to OS of nivo+ipi vs. chemo in PD-L1 CPS ≥ 5 subjects and to OS of nivo+chemo vs. chemo in PD-L1 CPS ≥ 1 subjects. The testing of OS for nivo+ipi vs chemo will be conducted with significance level of 0.015 or 0.035 (if PFS for nivo+chemo in CPS ≥ 5 subjects was positive).

7.5.2 Nivo + Chemo vs. Chemo Comparison

7.5.2.1 Analysis of Progression-Free Survival

PFS (primary definition) as assessed by BICR in randomized CPS ≥ 5 subjects will be compared between nivo+chemo vs. chemo, using a two-sided log rank test, stratified by the stratification factors as specified in Table 7.5-1. A significance level of 0.02 will be used. The estimate of the PFS hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole indicator covariate (for nivo+chemo vs. chemo arms). Ties will be handled using the exact method. A two-sided 98% CI for the hazard ratio will also be presented.

The PFS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median PFS in each treatment group will be computed via the log-log transformation method. PFS rates at fixed time points (eg, 6, 12, 18 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood[13] formula for variance derivation and on log-log transformation applied on the survivor function[14].

The source of PFS event (progression or death) will be summarized by treatment group. The status of subjects who are censored (as per primary definition of PFS) in the PFS KM analysis will be tabulated for each treatment group including the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)
- No baseline tumor assessment
- No on-study tumor assessment and no death
- Received subsequent anticancer therapy

A by-subject listing of lesion evaluations per BICR will be presented.

7.5.2.2 Supportive Analyses of Progression-Free Survival

Sensitivity analyses to examine assumptions of the stratified Log-rank test and Cox model:

1) To examine the assumption of proportional hazards in the Cox regression model, in addition to treatment, a time-dependent variable defined by treatment by time interaction will be added into the model. If the two-sided Wald Chi-square p-value of the treatment by time interaction is less than 0.1 or if a visual examination of the K-M plot of the PFS indicates a delayed separation of the curves, the following analyses and potential other exploratory analyses will be conducted in order to report the treatment effect when taking into consideration of the delayed effect.
   a) Delayed effect of immunotherapy interventions may cause a late separation in the PFS KM curves and non-proportional hazards. PFS (as determined by BICR) will be compared between treatment groups via a 2-sided max-combo test. The max-combo test statistic is the maximum of 4 different Fleming-Harrington family weighted log-rank test statistics. Z_m = max (FH (0, 0), FH (0,1), FH (1,0), F(1,1)), where FH(ρ,γ) are the test statistics from the Fleming-Harrington family of test statistics. FH (0, 0) corresponds to the log-rank test, while FH (0, 1) is more sensitive to late-difference alternatives, FH(1,0) is more sensitive to early difference with decreasing treatment effect and FH(1,1) uses weights at the median.
   b) The estimates of the PFS hazard ratios will be estimated in 2 periods. The periods will be defined by a cut-off point. The optimal cut off point will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. The optimal cut-off point will be estimated using a grid of possible cut-off points and obtained by maximizing the partial log likelihood. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio's will also be presented.
2) The stratified log rank-rank test assumes same treatment effect among the different strata. The method of Gail and Simon[16] will be used to test for a qualitative interaction between treatment and strata. This test will be conducted at α=0.10 level. The p-value reported from this specific analysis is for descriptive purposes alone.

Assess impact of Potential Prognostic Factors on Treatment Effect:

3) A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization, will be included in the model as stratification factors. However, all additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include:
   a) Age categorization (< 65, ≥ 65)
   b) Gender (Male vs. Female)
   c) Primary tumor location at initial diagnosis (GC, GEJ cancer, Esophageal Adenocarcinoma)
   d) Disease status (locally recurrent/advanced, metastatic)
   e) Lauren classification (intestinal type, diffuse type, mixed, unknown)
   f) Peritoneal metastases (yes/no)
   g) Prior surgery related to current cancer OR radiotherapy finished more than 6 months before randomization (yes/no)
   h) Number of organs with baseline lesion (target or non-target; ≤ 1, ≥ 2) outside primary location (gastric, gastroesophageal junction and esophagus)
   i) WHO histologic classification presence of signet ring cell (yes/no)

   The level of the covariate normally associated with the worst prognosis will be coded as the reference level. The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 95% CIs.

4) The identification of the primary population using PD-L1 CPS ≥ 5 was introduced close to accrual completion. Therefore the randomization was not stratified by PD-L1 CPS ≥ 5. The overall study population was stratified by
   - PD-L1 expression level (tumor cells) (≥ 1% vs. < 1% or indeterminate)
   - Region: Asia vs. US vs. RoW
   - ECOG performance status: 0 vs. 1
   - Planned chemotherapy regimen (XELOX vs. FOLFOX)

   As CPS≥5 is not a stratification factor, while pronounced imbalance is not expected due to the large sample size in this group, thorough review of data will be conducted and a sensitivity analysis may be conducted.

   The distribution of baseline characteristics (demographics, disease characteristics and the stratification factors) between the two treatment arms will be inspected for potential imbalance in the subgroup of subjects with PD-L1 CPS ≥ 5.

   The distribution of these characteristics will also be assessed between the CPS PD-L1 ≥ 5 subgroup and the overall population, regardless of treatment.

   If a pronounced imbalance between the treatment arms is observed for one or several factors (eg, >=10% difference between the frequency of the treatment arms for any level of the

categorical variables) a multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for those factors identified. The factors used in the randomization, will be included in the model as stratification factors and the factors identified with imbalance will be incorporated as covariates. For the purpose of fitting the multivariate model and easy interpretation of the results, any continuous variable identified will be dichotomized using a cut-off around median value and some sub-categories of a categorical covariate could collapsed due to small size of the sub-categories.

Assess Impact of Stratification:

5) PFS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided 98% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the PD-L1 CPS ≥ 5 randomized subjects for the 2 arms.
6) PFS using an un-stratified log-rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided 98% CIs.
7) PFS using an unstratified Cox proportional hazards model, adjusted, using as covariates only the stratification factors used in randomization. The hazard ratio associated with treatment will be presented along with the associated two-sided 98% CIs.

PFS using different sources or definition:

8) PFS by BICR using secondary definition of PFS (ITT) will be compared between nivo+chemo vs. chemo, using a two-sided log rank test, stratified by the stratification factors as specified in Table 7.5-1. In addition, for PFS by BICR using secondary definition, PFS using the investigator's assessment the primary definition and PFS using investigator's assessment and the secondary definition the hazard ratio associated with treatment with its 98% CI. The PFS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median PFS in each treatment group will be computed via the log-log transformation method. PFS rates at fixed time points (eg, 6, 12, 18 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood[13] formula for variance derivation and on log-log transformation applied on the survivor function[14].

   A by-subject listing will be presented including treatment group, PFS duration under the primary definition, PFS duration on the ITT definition (secondary definition), whether the subject was censored under the primary definition, and if censored, the reason, and whether the subject was censored under the ITT definition, and if censored, the reason.
9) A cross tabulation of PFS assessment by BICR versus PFS assessment by investigator will be presented, by treatment group. The concordance rate of events will be computed as the frequency with which investigator and BICR agree on classification of a subject as progression

vs. death vs. censored as a proportion of the total number of randomized PD-L1 CPS ≥ 5 subjects assessed by both the investigator and BICR.

A by-subject listing of PFS (primary definition) assessment per BICR and investigator will be presented.

Assessment of PFS in case of relevant protocol deviation:

10) If there are more than 10% PD-L1 CPS ≥ 5 randomized subjects with relevant protocol deviations the PFS (BICR primary definition) analysis will be conducted on subjects with no relevant protocol deviations. The hazard ratio associated with treatment will be presented along with the associated two-sided 98% CIs.
11) The stratified log rank test will be repeated on PFS (BICR) primary definition based on the first 228 events among the first 298 randomized subjects with CPS ≥ 5. The hazard ratio associated with treatment with its 98% CI will also be provided. This analysis will reflect the primary analysis per original design (see Amendment 23).
12) PFS accounting for two or more consecutively missing disease assessments prior to PFS event: This analysis will be performed only if at least 10% of PFS events have missing prior disease assessments. In case a subject has two or more consecutively missing disease assessments, the subject will be censored at the last disease assessment date prior the PFS event.
13) Analysis in which progression-free subjects who are lost to follow-up for any cause will be considered as having an event at the time of the last tumor assessment date prior to loss to follow-up. This analysis will be performed only if at least 10% of the progression-free subjects are lost to follow-up.

7.5.2.3 Subset Analyses of Progression-Free Survival

The influence of baseline and demographic characteristics on the treatment effect among randomized CPS ≥ 5 subjects will be explored via exploratory subset analyses. The median PFS (primary definition) based on KM product-limit method along with two-sided 95% CIs will be produced for the following subgroups:

- Region (Asia vs US vs. RoW; from CRF)
- Region (Asia [without China] vs. US vs. RoW [including China])
- ECOG performance status (0 vs. 1; from CRF)
- Age category (< 65, ≥ 65 and < 75, ≥ 75, ≥ 65)
- Gender (male, female)
- Race (Asian, White, Other)
- Primary tumor location at initial diagnosis (GC, GEJ cancer, Esophageal Adenocarcinoma)
- Disease stage at initial diagnosis (Stage I-II, Stage III, Stage IV)
- Disease status (locally recurrent/advanced, metastatic)
- Prior surgery related to current cancer (yes/no)
- Prior radiotherapy (yes/no)
- Lauren classification (intestinal type, diffuse type, mixed, unknown)
- WHO histologic classification presence of signet ring cell (yes/no)
- TNM classification Metastasis (M0, M1, unknown/Mx)
- At least one target lesion (yes/no)

- Number of organs with baseline lesion (target or non-target; ≤ 1, ≥ 2) outside primary location (gastric, gastroesophageal junction and esophagus)
- Time from Initial Disease Diagnosis to Randomization (< 6 months, 6 months -< 1 year, ≥ 1 year)
- Peritoneal metastases (yes/no)
- Liver metastases (yes/no)
- H.pylori (no, yes, unknown)
- HER-2 status at study entry (negative, positive, unknown)
- Baseline albumin (< LLN, ≥ LLN)
- Microsatellite instability (MSI-H, MSS, invalid, unknown)

A forest plot of the PFS unstratified hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above. The unstratified analysis comparing treatment (ie, Hazard Ratio) will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.2.4 Analysis of Overall Survival

OS in randomized CPS ≥ 5 subjects will be compared between nivo+chemo vs. chemo at the interim and final analyses, using stratified log-rank test, stratified by the stratification factors as specified in Table 7.5-1. An O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses (see Table 7.5.1-1). The stratified hazard ratio between the treatment groups will be presented along with 100* (1 - α)% CI (adjusted for interim) and will be calculated using a stratified Cox proportional hazards model, with treatment as the sole indicator covariate (for the nivo+chemo and chemo arm). Ties will be handled using the exact method.

OS will be estimated using the KM techniques. A two-sided 95% CI for median OS in each treatment group will be computed via the log-log transformation method. OS rates at fixed time points (eg, 12, 18, 24, 36 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and, corresponding CIs will be derived based on Greenwood[13] formula for variance derivation and on log-log transformation applied on the survivor function[14].

The status of subjects who are censored in the OS KM analysis will be tabulated for each treatment group using the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)

A by-subject listing will be presented including treatment group, first and last dose date, whether the subject died, and if censored, the reason, event/censored date and OS duration.

7.5.2.5 Supportive Analyses of Overall Survival

Sensitivity analyses to examine assumptions of the stratified Log-rank test and Cox model:

1) To examine the assumption of proportional hazards in the Cox regression model, in addition to treatment, a time-dependent variable defined by treatment by time interaction will be added into the model. If the two-sided Wald Chi-square p-value of the treatment by time interaction is less than 0.1 or if a visual examination of the K-M plot of the OS indicates a delayed separation of the curves, the following analyses and potential other exploratory analyses will be conducted in order to report the treatment effect when taking into consideration of the delayed effect.
   a) Delayed effect of immunotherapy interventions may cause a late separation in the OS KM curves and non-proportional hazards. OS will be compared between treatment groups via a 2-sided max-combo test. The max-combo test statistic is the maximum of 4 different Fleming-Harrington family weighted log-rank test statistics. Zm = max (FH (0, 0), FH (0,1), FH (1,0), F(1,1)), where FH(ρ,γ) are the test statistics from the Fleming-Harrington family of test statistics. FH (0, 0) corresponds to the log-rank test, while FH (0, 1) is more sensitive to late-difference alternatives, FH(1,0) is more sensitive to early difference with decreasing treatment effect and FH(1,1) uses weights at the median.
   b) The estimates of the OS hazard ratios will be estimated in 2 periods. The periods will be defined by a cut-off point. The optimal cut off point will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. The optimal cut-off point will be estimated using a grid of possible cut off points and obtained by maximizing the partial log likelihood. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio's will also be presented.
2) The method of Gail and Simon[16] will be used to test for a qualitative interaction between treatment and strata. This test will be conducted at α=0.10 level. The p-value reported from this specific analysis is for descriptive purposes alone.

Assess impact of Potential Prognostic Factors on Treatment Effect:

3) A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization will be included in the model as stratification factors. However, all additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include the same factors as in the PFS analysis.

   The level of the covariate normally associated with the worst prognosis will be coded as the reference level. The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 95% CIs.
4) The identification of the primary population using PD-L1 CPS ≥ 5 was introduced close to accrual completion. Therefore the randomization was not stratified by PD-L1 CPS ≥ 5. The overall study population was stratified by:
   - PD-L1 expression level (tumor cells) (≥ 1% vs. < 1% or indeterminate)
   - Region: Asia vs. US vs. RoW

- ECOG performance status: 0 vs. 1
- Planned chemotherapy regimen (XELOX vs. FOLFOX)

As CPS ≥ 5 is not a stratification factor, while pronounced imbalance is not expected due to the large sample size in this group, thorough review of data will be conducted and a sensitivity analysis may be conducted.

The distribution of baseline characteristics (demographics, disease characteristics and the stratification factors) between the two treatment arms will be inspected for potential imbalance in the subgroup of subjects with PD-L1 CPS ≥ 5.

The distribution of these characteristics will also be assessed between the PD-L1 CPS ≥ 5 subgroup and the overall population, regardless of treatment.

If a pronounced imbalance between the treatment arms is observed for one or several factors (eg, >=10% difference between the frequency of the treatment arms for any level of the categorical variables) a multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for those factors identified. The factors used in the randomization, will be included in the model as stratification factors and the factors identified with imbalance will be incorporated as covariates. For the purpose of fitting the multivariate model and easy interpretation of the results, any continuous variable identified will be dichotomized using a cutoff around median value and some sub-categories of a categorical covariate could collapsed due to small size of the sub-categories.

Assess Impact of Stratification:

5) OS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the randomized subjects.
6) OS using an un-stratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs.
7) OS using an unstratified Cox proportional hazards model, adjusted, using as covariates only the stratification factors used in randomization. The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs.

Assessment of OS in case of relevant protocol deviation:

8) If there are more than 10% subjects with relevant protocol deviations the OS analysis for subjects with no relevant protocol deviations will be conducted. The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs.
9) The OS analysis (stratified log rank) will be conducted based on the first 420 randomized subjects with PD-L1 CPS ≥ 5 (see Protocol Amendment 23) and all events observed in this

population. The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs.

10) The OS analysis (stratified log rank) will be conducted based on the original planned number of events (ie, first 354 deaths) observed among randomized subjects with CPS ≥ 5. The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs.

7.5.2.6 Subset Analyses of Overall Survival

The influence of baseline and demographic characteristics on the treatment effect among all randomized CPS ≥ 5 subjects will be explored via exploratory subset analyses. The median OS based on KM product-limit method along with two-sided 95% CIs will be produced for the same subgroups as in the PFS analysis (see Section 7.5.2.3).

A forest plot of the OS unstratified hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above. The unstratified analysis comparing treatment (ie, Hazard Ratio) will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.2.7 Current Status of PFS and OS Follow-up

The extent of follow-up for survival, defined as the time between randomization date and last known alive date (for subjects who are alive) or death date (for subjects who died), will be summarized descriptively (median, min, max, etc.) in months for randomized CPS ≥ 5 subjects and all randomized subjects.

The currentness of follow-up for survival, defined as the time between last OS contact (ie, last known alive date or death date) and cutoff date (defined by last patient last visit date), will be summarized in months for randomized CPS≥5 subjects and all randomized subjects. Subjects who died and subjects with last known alive date on or after data cut-off date will have zero value for currentness of follow-up.

Minimum follow-up of OS for all randomized subjects, defined as the time from cutoff date to last subject's randomization date, will be summarized in months.

Time from last evaluable tumor assessment to cut-off date in months will be summarized by treatment group and overall for randomized CPS ≥ 5 subjects and all randomized subjects. Subjects who have a PFS event will be considered as current for this analysis. The secondary definition of PFS will be used for this summary.

A by-subject listing will also be produced to accompany the subject time from last evaluable tumor assessment.

7.5.2.8 Analysis of OS by Tumor Response:

Survival by response category will be analyzed by arm using the landmark method.[17] Subjects still on study at the landmark time will be separated into two response categories according to whether they have responded before that time. This will assess whether survival from the landmark depends

on the subject's response status at the landmark. Subjects who go off protocol (eg, subjects who die) before the time of landmark will be excluded from the analysis.

The survival curves from Week 6, Week 12, Week 18, Month 6, Month 9, Month 12, by response status, for each randomized arm will be produced using the KM product-limit method. Two sided, 95% CIs for median OS will be constructed based on a log-log transformed CI for the survivor function S(t).

7.5.2.9 Secondary Efficacy Endpoints

Overall Survival in CPS ≥ 1 subjects and All Randomized Subjects

If the OS primary endpoint in randomized CPS ≥ 5 subjects is significantly superior then OS in randomized CPS ≥ 1 subjects will be compared using a two-sided log rank test and a significance level of 1.5%, stratified by the stratification factors as specified in Table 7.5-1. Similarly if OS testing in randomized CPS ≥ 1 subjects is significantly superior then OS in all randomized subjects (all comers) will be compared using a two-sided log rank test and a significance level of 1.5%, stratified by the stratification factors as specified in Table 7.5-1.

The stratified hazard ratio between the treatment groups will be presented along with 100* (1 - α)% CI (adjusted for interim). In addition, two-sided p-value will also be reported for the analysis of OS.

OS will be estimated using the KM techniques. A two-sided 95% CI for median OS in each treatment group will be computed via the log-log transformation method. OS rates at fixed time points (eg, 6, 12, 18, 24 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood[13] formula for variance derivation and on log-log transformation applied on the survivor function.[14]

Additional Analyses for OS in CPS ≥ 1 subjects and All Randomized Subjects

As OS in CPS ≥ 1 subjects and OS in all randomized subjects are tested hierarchically, the following analyses will only be conducted if the endpoint is statistically significant:

All analyses as described in Sections 7.5.2.5, 7.5.2.6 and 7.5.2.8.

Overall Survival in CPS ≥ 10 subjects

This endpoint will not be formally tested. The stratified hazard ratio between the treatment groups will be presented along with 95% CI. OS will be estimated using the KM techniques. A two-sided 95% CI for median OS in each treatment group will be computed via the log-log transformation method. OS rates at fixed time points (eg, 6, 12, 18, 24 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

PFS (BICR) in CPS ≥ 10, 1 and in all randomized:

These endpoints will not be formally tested. PFS using primary definition per BICR will be analyzed. The stratified HR with its associated two-sided 95% CI will be estimated via a stratified Cox model with treatment arm as the only covariate in the model. PFS will be estimated using the KM techniques. A two-sided 95% CI for median PFS in each treatment group will be computed via the log-log transformation method. PFS rates at fixed time points (eg, 6, 12, 18 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

Best Overall Response and ORR (BICR) in CPS ≥ 10, 5, 1 and in all Randomized Subjects

These endpoints will not be formally tested. The analyses will repeated for the 4 populations: CPS ≥ 10, CPS ≥ 5, CPS ≥ 1 and all randomized. The analyses using BICR data are considered as secondary endpoints (no formal testing). The analyses will be repeated on subjects with measurable disease at baseline (see Section 6.3)

- The number and percentage of subjects in each category of BOR (complete response [CR], partial response [PR], stable disease [SD], progressive disease [PD], or unable to determine [UTD]) will be presented, by treatment group.
- Estimates of response rate, along with its exact two-sided 95% CI by Clopper and Pearson[18] will be presented, by treatment group.
- A 2-sided, 95% CI for the difference of ORR between treatment groups will also be computed by the method of DerSimonian and Laird,[19] using a fixed-effects model (setting Δ2 equal to zero), adjusting for the stratification factors. The weighted response rate difference and 95% CI can be determined using the following formula:

$$\hat{\theta} = \frac{\sum_{i=1}^{24} \hat{\theta}_i w_i}{\sum_{i=1}^{24} w_i} \sim N(\theta, 1/\sum_{i=1}^{24} w_i)$$

  where $\hat{\theta}_i$ is the response rate difference of the i[th] stratum and $w_i = 1/var(\hat{\theta}_i)$.

- In addition, the stratified (source: IVRS) odds ratios (Mantel-Haenszel estimator) between the treatments will be provided along with the 95% CI.
- ORR based on BICR assessment per RECIST 1.1 will be compared between two treatment groups using a two-sided Cochran-Mantel-Haenszel (CMH) test, stratified by the stratification factors. This is p-value is descriptive only.

The following subject-level graphics will also be provided:

- For the responders only, time courses of the following events of interest will be graphically displayed: tumor response, progression, last dose received, and death.
- For response evaluable subjects (randomized subjects with evaluable* baseline and at least one evaluable* on-study tumor assessment),

- A bar plot showing the best % reduction from baseline in sum of diameter of target lesions based on BICR assessment for each subject will be produced (excluding assessments after PD and assessments after start of subsequent anti-cancer therapy).
- A plot of individual time course of tumor burden change per BICR assessment will be produced.

*Evaluable is defined according to RECIST v1.1

Subset Analyses of Objective Response

The influence of baseline and demographic characteristics on the treatment effect will be explored via exploratory subset analysis. BICR assessment of ORR in CPS ≥ 5 subjects will be summarized for the same subgroups than for PFS (see Section 7.5.2.3).

A forest plot of treatment effect on ORR per BICR in CPS ≥ 5 subjects will be produced. The un-weighted differences in ORR between the two treatment groups and corresponding 95% two-sided CI using the method of Newcombe[20] will be provided. The analysis comparing treatment (ie, ORR difference) will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.2.10 Exploratory Efficacy Endpoints

PFS (investigator) in CPS ≥ 10, 5, 1 and all randomized subjects

Similar analyses than for PFS (BICR) in CPS ≥ 10, 1 and in all randomized will be produced but using investigator assessment using primary definition.

ORR (investigator) in CPS ≥ 10, 5, 1 and in all randomized subjects:

Similar analyses than for Best Overall Response and ORR per BICR will be produced but using investigator assessment of response and progression.

Duration of Response (BICR and investigator) in CPS ≥ 10, 5, 1 and all randomized subjects

Duration of response (DoR) will be evaluated per BICR assessment and per investigator assessment. DoR will be estimated using the KM techniques. A two-sided 95% CI for median in each treatment group will be computed via the log-log transformation method. For the DoR analyses the population is restricted to subjects with measurable disease at baseline.

For DoR, type of event (PD, death) and status of subjects censored (received subsequent therapy, ongoing follow-up [current, non-current], off study [withdrew consent] will be presented by treatment arm.

Time to objective response (BICR) in randomized CPS ≥ 5 subjects and all randomized subjects.

Time to response will be evaluated per BICR assessment. Summary statistics will be provided for each treatment group for subjects who achieve PR or CR. For these analyses the population is restricted to subjects with measurable disease at baseline.

To assess further tumor response kinetics, time to response will be analyzed using the KM methodology. Kaplan-Meier curve will represent the cumulative rate of response over time. For

the non-responders, time to response will be censored at the maximum time of response + 1 day of all subjects in their respective treatment group. Cumulative Response Rates will be tabulated for Week 6, Week 12, Months 4, 6, 8, and 12, and overall Response Rate will be provided for each treatment group.

Durable response rate (BICR and investigator) in all randomized CPS ≥ 5 subjects

Durable response will be evaluated per BICR assessment and per investigator assessment for subjects with measurable disease at baseline. The durable response rate will be computed in each treatment group along with the exact 95% CI using Clopper-Pearson method.[18] An estimate of the difference in DRRs and corresponding 95% CI will be calculated using DerSimonian and Laird methodology and adjusted by the stratification factors as specified in Table 7.5-1. The stratified odds ratios (Mantel-Haenszel estimator) between the treatments will be provided along with the 95% CI.

PFS and ORR as assessed by either BICR or investigator and OS across TPS cut-offs.

See Section 7.8 for Biomarker Analysis.

PFS2 in randomized CPS ≥ 5 subjects and in all randomized subjects

PFS2 will be analyzed similarly to PFS:

- Median values based on KM method, along with two-sided 95% CI using Brookmeyer and Crowley method will be calculated. The estimate of standard error will be calculated using the Greenwood formula;
- PFS2 will be graphically displayed along with the median and 95% CI.

A by-subject listing of PFS and PFS2 will be provided.

TTSD in CPS ≥ 10, 5, 1 and all randomized subjects

See Section 7.9.1 in Clinical Outcome Assessment.

7.5.3 Nivo + Ipi vs. Chemo Comparison

7.5.3.1 Analysis of Overall Survival

OS in randomized CPS ≥ 5 subjects will be compared between nivo+ipi vs. chemo using stratified log-rank test, stratified by the stratification factors as specified in Table 7.5-1. The significance level α will depend on which primary endpoints was statistically significant (see Section 7.5.1). The estimate of the OS hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole indicator covariate (for nivo+ipi and chemo). Ties will be handled using the exact method. A two-sided $(1 - \alpha)$% CI for the hazard ratio will also be presented.

OS will be estimated using the KM techniques. A two-sided 95% CI for median OS in each treatment group will be computed via the log-log transformation method. OS rates at fixed time points (eg, 12, 18, 24, 36 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate

and, corresponding CIs will be derived based on Greenwood[13] formula for variance derivation and on log-log transformation applied on the survivor function.[14]

The status of subjects who are censored in the OS KM analysis will be tabulated for each treatment group using the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)

A by-subject listing will be presented including treatment group, first and last dose date, whether the subject died, and if censored, the reason, event/censored date and OS duration.

7.5.3.2 Supportive Analyses of Overall Survival

Sensitivity analyses to examine assumptions of the stratified Log-rand test and Cox model:

1) To examine the assumption of proportional hazards in the Cox regression model, in addition to treatment, a time-dependent variable defined by treatment by time interaction will be added into the model. If the two-sided Wald Chi-square p-value of the treatment by time interaction is less than 0.1 or if a visual examination of the K-M plot of the OS indicates a delayed separation of the curves, the following analyses and potential other exploratory analyses will be conducted in order to report the treatment effect when taking into consideration of the delayed effect..
   a) Delayed effect of immunotherapy interventions may cause a late separation in the OS KM curves and non-proportional hazards. OS will be compared between treatment groups via a 2-sided max-combo test. The max-combo test statistic is the maximum of 4 different Fleming-Harrington family weighted log-rank test statistics. Zm = max (FH (0, 0), FH (0,1), FH (1,0), F(1,1)), where FH(ρ,γ) are the test statistics from the Fleming-Harrington family of test statistics. FH (0, 0) corresponds to the log-rank test, while FH (0, 1) is more sensitive to late-difference alternatives, FH(1,0) is more sensitive to early difference with decreasing treatment effect and FH(1,1) uses weights at the median.
   b) The estimates of the OS hazard ratios will be estimated in 2 periods. The periods will be defined by a cut-off point. The optimal cut off point will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. The optimal cut-off point will be estimated using a grid of possible cut off points and obtained by maximizing the partial log likelihood. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio's will also be presented.
2) The method of Gail and Simon[16] will be used to test for a qualitative interaction between treatment and strata. This test will be conducted at α=0.10 level. The p-value reported from this specific analysis is for descriptive purposes alone.

Assess impact of Potential Prognostic Factors on Treatment Effect:

3) A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization will be included in the model as stratification factors. However, all

additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include the same factors as in the nivo+chemo vs. chemo analysis.

The level of the covariate normally associated with the worst prognosis will be coded as the reference level. The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 95% CIs.

4) The identification of the primary population using PD-L1 CPS≥5 was introduced close to accrual completion. Therefore, the randomization was not stratified by PD-L1 CPS ≥ 5. The overall study population was stratified by:

- PD-L1 expression level (tumor cells) (≥ 1% vs. < 1% or indeterminate)
- Region: Asia vs. US vs. RoW
- ECOG performance status: 0 vs. 1
- Planned chemotherapy regimen (XELOX vs. FOLFOX)

As CPS ≥ 5 is not a stratification factor, while pronounced imbalance is not expected due to the large sample size in this group, thorough review of data will be conducted and a sensitivity analysis may be conducted.

The distribution of baseline characteristics (demographics, disease characteristics and the stratification factors) between the two treatment arms will be inspected for potential imbalance in the subgroup of subjects with PD-L1 CPS ≥ 5.

The distribution of these characteristics will also be assessed between the PD-L1 CPS ≥ 5 subgroup and the overall population, regardless of treatment.

If a pronounced imbalance between the treatment arms is observed for one or several factors (eg, ≥ 10% difference between the frequency of the treatment arms for any level of the categorical variables) a multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for those factors identified. The factors used in the randomization, will be included in the model as stratification factors and the factors identified with imbalance will be incorporated as covariates. For the purpose of fitting the multivariate model and easy interpretation of the results, any continuous variable identified will be dichotomized using a cutoff around median value and some sub-categories of a categorical covariate could collapsed due to small size of the sub-categories

<u>Assess Impact of Stratification:</u>

5) OS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the randomized subjects.

6) OS using an un-stratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs.
7) OS using an unstratified Cox proportional hazards model, adjusted, using as covariates only the stratification factors used in randomization. The hazard ratio associated with treatment will be presented along with the associated two-sided (1 - α)% CIs.

Assessment of OS in case of relevant protocol deviation:

8) If there are more than 10% subjects with relevant protocol deviations the OS analysis for subjects with no relevant protocol deviations will be conducted. The hazard ratio associated with treatment will be presented along with the associated two-sided (1- α)% CIs.

7.5.3.3 Subset Analyses of Overall Survival

The influence of baseline and demographic characteristics on the treatment effect among randomized CPS ≥ 5 subjects will be explored via exploratory subset analyses. The median OS based on KM product-limit method along with two-sided 95% CIs will be produced for the following subgroups

- Region (Asia vs. US vs. RoW; from CRF)
- Region (Asia [without China] vs. US vs. RoW [including China])
- ECOG performance status (0 vs. 1; from CRF)
- Age category (< 65, ≥ 65 and < 75, ≥ 75, ≥ 65)
- Gender (male, female)
- Race (Asian, White, Other)
- Primary tumor location at initial diagnosis (GC, GEJ cancer, Esophageal Adenocarcinoma)
- Disease stage at initial diagnosis (Stage I-II, Stage III, Stage IV)
- Disease status (locally recurrent/advanced, metastatic)
- Prior surgery related to current cancer (yes/no)
- Prior radiotherapy (yes/no)
- Lauren classification (intestinal type, diffuse type, mixed, unknown)
- WHO histologic classification presence of signet ring cell (yes/no)
- TNM classification Metastasis (M0, M1, unknown/Mx)
- At least one target lesion (yes/no)
- Number of organs with baseline lesion (target or non-target; ≤1, ≥2) outside primary location (gastric, gastroesophageal junction and esophagus)
- Time from Initial Disease Diagnosis to Randomization (< 6months, 6months - < 1 year, ≥ 1 year)
- Peritoneal metastases (yes/no)
- Liver metastases (yes/no)

- H.pylori (no, yes, unknown)
- HER-2 status at study entry (negative, positive, unknown)
- Baseline albumin (< LLN, ≥ LLN)
- Microsatellite instability (MSI-H, MSS, invalid, unknown)

A forest plot of the OS unstratified hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above. The unstratified analysis comparing treatment (ie, Hazard Ratio) will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.3.4 Analysis of OS by Tumor Response:

Survival by response category will be analyzed by arm using the landmark method.[21] Subjects still on study at the landmark time will be separated into two response categories according to whether they have responded before that time. This will assess whether survival from the landmark depends on the subject's response status at the landmark. Subjects who go off protocol (eg, subjects who die) before the time of landmark will be excluded from the analysis.

The survival curves from Week 6, Week 12, Week 18, Month 6, Month 9, Month 12, by response status, for each randomized arm will be produced using the KM product-limit method. Two sided, 95% CIs for median OS will be constructed based on a log-log transformed CI for the survivor function S(t).

Overall Survival in All Randomized Subjects

If the OS secondary endpoint in randomized CPS ≥ 5 subjects is significantly superior then OS in all randomized subjects will be compared using a two-sided log rank test at the same significance level of α%, stratified by the stratification factors as specified in Table 7.5-1.

The stratified hazard ratio between the treatment groups will be presented along with 100* (1 - α)% CI (adjusted for interim). In addition, two-sided p-value will also be reported for the analysis of OS.

OS will be estimated using the KM techniques. A two-sided 95% CI for median OS in each treatment group will be computed via the log-log transformation method. OS rates at fixed time points (eg, 6, 12, 18, 24 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

Additional Analyses for OS in all Randomized Subjects

As OS in all randomized subjects is tested hierarchically, the following analyses will only be conducted if the endpoint is statistically significant:

All analyses as described in Sections 7.5.3.2, 7.5.3.3 and 7.5.3.4.

Overall Survival in CPS≥10, 1 subjects

These endpoints will not be formally tested. The stratified hazard ratio between the treatment groups will be presented along with 95% CI. OS will be estimated using the KM techniques. A two-sided 95% CI for median OS in each treatment group will be computed via the log-log transformation method. OS rates at fixed time points (eg, 6, 12, 18, 24, 36 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

7.5.3.5 Analysis of Progression-Free Survival

PFS (BICR) in CPS ≥ 10, 5, 1 and in all randomized subjects:

These endpoints will not be formally tested. PFS using primary definition and secondary definition per BICR will be analyzed. The stratified HR with its associated two-sided 95% CI will be estimated via a stratified Cox model with treatment arm as the only covariate in the model. PFS will be estimated using the KM techniques. A two-sided 95% CI for median PFS in each treatment group will be computed via the log-log transformation method. PFS rates at fixed time points (eg, 6, 12, 18 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

The source of PFS event (progression or death) will be summarized by treatment group. The status of subjects who are censored (as per primary definition of PFS) in the PFS KM analysis will be tabulated for each treatment group including the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)
- No baseline tumor assessment
- No on-study tumor assessment and no death
- Received subsequent anticancer therapy

A by-subject listing will be presented including treatment group, PFS duration under the primary definition, PFS duration on the ITT definition (secondary definition), whether the subject was censored under the primary definition, and if censored, the reason, and whether the subject was censored under the ITT definition, and if censored, the reason.

A by-subject listing of lesion evaluations per BICR will be presented.

7.5.3.6 Subset Analyses of Progression-Free Survival

The influence of baseline and demographic characteristics on the treatment effect among all randomized subjects will be explored via exploratory subset analyses. The median PFS (primary

definition) based on KM product-limit method along with two-sided 95% CIs will be produced for the following subgroups in subjects with CPS≥5:

- Region (Asia vs. US vs. RoW; from CRF)
- Region (Asia [without China] vs. US vs. RoW [including China])
- ECOG performance status (0 vs. 1; from CRF)
- Chemotherapy regimen (XELOX vs. FOLFOX; from CRF)
- Age category (< 65, ≥ 65 and < 75, ≥ 75, ≥ 65)
- Gender (male, female)
- Race (Asian, White, Other)
- Primary tumor location at initial diagnosis (GC, GEJ cancer, Esophageal Adenocarcinoma)
- Disease stage at initial diagnosis (Stage I-II, Stage III, Stage IV)
- Disease status (locally recurrent/advanced, metastatic)
- Prior surgery related to current cancer (yes/no)
- Prior radiotherapy (yes/no)
- Lauren classification (intestinal type, diffuse type, mixed, unknown)
- WHO histologic classification presence of signet ring cell (yes/no)
- TNM classification Metastasis (M0, M1, unknown/Mx)
- At least one target lesion (yes/no)
- Number of organs with baseline lesion (target or non-target; ≤ 1, ≥ 2) outside primary location (gastric, gastroesophageal junction and esophagus)
- Time from Initial Disease Diagnosis to Randomization (< 6 months, 6 months -< 1 year, ≥ 1 year)
- Peritoneal metastases (yes/no)
- Liver metastases (yes/no)
- H.pylori (no, yes, unknown)
- HER-2 status at study entry (negative, positive, unknown)
- Baseline albumin (< LLN, ≥ LLN)
- Microsatellite instability (MSI-H, MSS, invalid, unknown)

A forest plot of the PFS unstratified hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above. The unstratified analysis comparing treatment (ie, Hazard Ratio) will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.3.7 Current Status of PFS and OS Follow-up

The extent of follow-up for survival, defined as the time between randomization date and last known alive date (for subjects who are alive) or death date (for subjects who died), will be summarized descriptively (median, min, max, etc.) in months for all randomized CPS ≥ 5 subjects and all randomized subjects.

The currentness of follow-up for survival, defined as the time between last OS contact (ie, last known alive date or death date) and cutoff date (defined by last patient last visit date), will be summarized in months for all randomized CPS ≥ 5 subjects and all randomized subjects. Subjects who died and subjects with last known alive date on or after data cut-off date will have zero value for currentness of follow-up.

Minimum follow-up of OS for all randomized subjects, defined as the time from cut-off date to last subject's randomization date, will be summarized in months.

Time from last evaluable tumor assessment to cut-off date in months will be summarized by treatment group and overall for all randomized CPS ≥ 5 subjects and all randomized subjects. Subjects who have a PFS event will be considered as current for this analysis. The secondary definition of PFS will be used for this summary.

7.5.3.8 Other Secondary Efficacy Endpoints

Best Overall Response and ORR (BICR) in CPS ≥ 10, 5, 1 and in all randomized subjects

These endpoints will not be formally tested. The analyses will repeated for the 4 populations: CPS ≥ 10, CPS ≥ 5, CPS ≥ 1 and all randomized. The analyses using BICR data are considered as secondary endpoints (no formal testing). The analyses will be repeated on subjects with measurable disease at baseline (see Section 6.3).

- The number and percentage of subjects in each category of BOR (complete response [CR], partial response [PR], stable disease [SD], progressive disease [PD], or unable to determine [UTD]) will be presented, by treatment group.
- Estimates of response rate, along with its exact two-sided 95% CI by Clopper and Pearson[18] will be presented, by treatment group.
- A 2-sided, 95% CI for the difference of ORR between treatment groups will also be computed by the method of DerSimonian and Laird,[19] using a fixed-effects model (setting Δ2 equal to zero), adjusting for the stratification factors. The weighted response rate difference and 95% CI can be determined using the following formula:

$$\hat{\theta} = \frac{\sum_{i=1}^{12} \hat{\theta}_i w_i}{\sum_{i=1}^{12} w_i} \sim N(\theta, 1/\sum_{i=1}^{12} w_i)$$

  where $\hat{\theta}_i$ is the response rate difference of the i[th] stratum and $w_i = 1/var(\hat{\theta}_i)$.

- In addition, the stratified (source: IVRS) odds ratios (Mantel-Haenszel estimator) between the treatments will be provided along with the 95% CI.
- ORR based on BICR assessment per RECIST 1.1 will be compared between two treatment groups using a two-sided Cochran-Mantel-Haenszel (CMH) test, stratified by the stratification factors. This is p-value is descriptive only.

The following subject-level graphics will also be provided:

- For the responders only, time courses of the following events of interest will be graphically displayed: tumor response, progression, last dose received, and death.
- For response evaluable subjects (randomized subjects with evaluable* baseline and at least one evaluable* on-study tumor assessment),
  - A bar plot showing the best % reduction from baseline in sum of diameter of target lesions based on BICR assessment for each subject will be produced (excluding assessments after PD and assessments after start of subsequent anti-cancer therapy).
  - A plot of individual time course of tumor burden change per BICR assessment will be produced.
  - *Evaluable is defined according to RECIST v1.1

Time to Symptom Deterioration in in CPS ≥ 5 and all randomized Subjects

TTSD in CPS ≥ 5 and all randomized will be analyzed. See details Section 7.9.1 "Clinical Outcome Assessment".

7.5.3.9 Exploratory Efficacy Endpoints

PFS (investigator) in CPS ≥ 10, 5, 1 and in All randomized subjects

Similar analyses than for PFS (BICR) will be produced but using investigator assessment and primary definition.

ORR (investigator) in CPS ≥ 10, 5, 1 and in all randomized subjects:

Similar analyses than for Best Overall Response and ORR per BICR will be produced but using investigator assessment of response and progression.

Duration of Response (BICR and investigator) in CPS ≥ 10, 5, 1 and all randomized subjects

Duration of response (DoR) will be evaluated per BICR assessment and per investigator assessment. DoR will be estimated using the KM techniques. A two-sided 95% CI for median in each treatment group will be computed via the log-log transformation method. For the DoR analyses the population is restricted to subjects with measurable disease at baseline.

For DoR, type of event (PD, death) and status of subjects censored (received subsequent therapy, ongoing follow-up [current, non-current], off study [withdrew consent] will be presented by treatment arm.

Time to objective response (BICR) in CPS ≥ 5 and all randomized subjects

Time to response will be evaluated per BICR assessment. Summary statistics will be provided for each treatment group for subjects who achieve PR or CR. For these analyses the population is restricted to subjects with measurable disease at baseline.

To assess further tumor response kinetics, time to response will be analyzed using the KM methodology. Kaplan-Meier curve will represent the cumulative rate of response over time. For

the non-responders, time to response will be censored at the maximum time of response + 1 day of all subjects in their respective treatment group. Cumulative Response Rates will be tabulated for Week 6, Week 12, Months 4, 6, 8, and 12, and overall Response Rate will be provided for each treatment group.

Durable response rate (BICR) in CPS ≥ 5 and all randomized subjects

Durable response will be evaluated per BICR assessment and per investigator assessment for subjects with measurable disease at baseline. The durable response rate will be computed in each treatment group along with the exact 95% CI using Clopper-Pearson method.[18] An estimate of the difference in DRRs and corresponding 95% CI will be calculated using CMH methodology and adjusted by the stratification factors as specified in Table 7.5-1. The stratified odds ratios (Mantel-Haenszel estimator) between the treatments will be provided along with the 95% CI.

PFS and ORR as assessed by either BICR or investigator and OS across TPS cut-offs

See Section 7.8 for Biomarker Analysis.

PFS2 in CPS ≥ 5 and in all randomized subjects

PFS2 will be analyzed similarly to PFS:

- Median values based on KM method, along with two-sided 95% CI using Brookmeyer and Crowley method will be calculated. The estimate of standard error will be calculated using the Greenwood formula;
- PFS2 will be graphically displayed along with the median and 95% CI.

A by-subject listing of PFS and PFS2 will be provided.

TTSD in CPS ≥ 10, 5, 1 and all randomized

See Section 7.9.1 in Clinical Outcome Assessment.

7.6 Safety

Analyses in this section will be tabulated for all treated subjects by treatment group as treated, unless otherwise specified. In addition some safety selected outputs will be presented for treated CPS ≥ 5 Subjects.

For chemotherapy (FOLFOX or XELOX) the Chemotherapy regimen "as treated" will be reported (see Section 6.2).

7.6.1 Deaths

Deaths will be summarized by treatment group:

- All deaths, reasons for death.
- Deaths within 30 days of last dose received, reasons for death. This output will also be presented for CPS ≥ 5 treated subjects.
- Deaths within 100 days of last dose received, reasons for death.

A by-subject listing of deaths will be provided for the all enrolled subjects population.

7.6.2 Serious Adverse Events

Serious adverse events will be summarized by treatment group:

- Overall summary of SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of drug-related SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT. This output will also be presented for CPS ≥ 5 treated subjects.

All analyses will be conducted using the 30-day safety window.

A by-subject SAE listing will be provided for the "enrolled subjects" population.

7.6.3 Adverse Events Leading to Discontinuation of Study Therapy

AEs leading to discontinuation will be summarized by treatment group:

- Overall summary of AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of drug-related AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT. This output will also be presented for CPS ≥ 5 treated subjects.

The analyses will be conducted using the 30-day safety window.

A by-subject AEs leading to discontinuation listing will be provided.

7.6.4 Adverse Events Leading to Dose Modification

AEs leading to dose delay/reduction will be summarized by treatment group:

- Overall summary of AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of related AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

The analysis will be conducted using the 30-day safety window.

A by-subject AEs leading to dose delay/reduction listing will be provided.

7.6.5 Adverse Events

Adverse events will be summarized by treatment group.

The following analyses will be conducted using the 30 days safety window only:

- Overall summary of any AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of any AEs presented by worst CTC grade (any grade, grade 3-4, grade 5) by SOC/PT. This table will be restricted to events with an incidence greater or equal to 5% in any treatment group. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of any non-serious AEs presented by SOC/PT. This table will be restricted to events with an incidence greater or equal to 5% in any treatment group.
- Overall summary of any AEs that required immune modulating medication by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT. This output will also be presented for CPS ≥ 5 treated subjects.
- The following analyses will be conducted using the 30 days safety window and repeated using the 100 days safety window:
- Overall summary of drug-related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT. This output will also be presented for CPS ≥ 5 treated subjects.

A by-subject AE listing will be provided. A by-subject listing of any AE requiring immune modulating medications will also be provided.

7.6.6 Select Adverse Events

Unless otherwise specified, analyses will be performed by select AE category. Analyses will also be repeated by subcategory of endocrine events.

7.6.6.1 Incidence of Select AE

Select AEs will be summarized by treatment group for each category/subcategory.

The following analyses will be conducted using the 30-day safety window only:

- Overall summaries of any select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory/PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summaries of any drug-related select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory/PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summaries of any serious select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Overall summaries of drug-related serious select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Overall summaries of any select AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Overall summaries of drug-related select AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT.
- Summary of frequency of unique select AEs by Category.

A by-subject select AE listing will be provided.

7.6.6.2 Time-to Onset of Select AE

Time-to onset of drug-related select AEs (any grade, grade 3-5) will be summarized for each category/subcategory by treatment group.

Time-to onset analyses are restricted to treated subjects who experienced at least one drug-related select AE in the category/subcategory. The analyses will be conducted using the 30-day safety window.

Additional details regarding the time-to onset definition are described in time-to onset definition subsection of APPENDIX 1.

7.6.6.3 Time-to Resolution of Select AE

Time-to resolution of the following specific events will be summarized separately for each category/subcategory.

- Time-to resolution of drug-related select AE (any grade, grade 3-5) by treatment group
- Time-to resolution of drug-related select AE (any grade, grade 3-5) where immune modulating medication was initiated, by treatment group

Time-to resolution analyses are restricted to treated subjects who experienced the specific events. Time-to resolution where immune modulating medication was initiated analyses are restricted to treated subjects who experienced the specific events and who received immune modulating medication during the longest select AE.

The analyses will be conducted using the 30-day safety window.

The following summary statistics will be reported: percentage of subjects with resolution of the longest select AE, median time-to resolution along with 95% CI (derived from Kaplan-Meier estimation) and ranges.

See time-to resolution definition subsection of APPENDIX 1 for additional details.

7.6.7 Immune-Mediated Adverse Events

IMAEs will be summarized by treatment group for each immune-mediated category / PT using the 100-day safety window:

- Overall summary of non-endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT. This output will also be presented for CPS ≥ 5 treated subjects.

- Overall summary of non-endocrine IMAEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of endocrine IMAEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of non-endocrine IMAEs leading to dose delay or reduction by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Overall summary of endocrine IMAEs leading to dose delay or reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT. This output will also be presented for CPS ≥ 5 treated subjects.
- Summaries of time-to onset and time-to resolution of non-endocrine IMAEs where immune modulating medication was initiated presented by Category. This output will also be presented for CPS ≥ 5 treated subjects.
- Summaries of time-to onset and time-to resolution of endocrine IMAEs presented by Category. This output will also be presented for CPS ≥ 5 treated subjects.

A by-subject listing of IMAEs will be provided. By-subject listings of time-to resolution for longest IMAEs cluster (any grade and grade 3-5 in separate summaries) will also be provided. For new studies which collect investigator assessment of potential IMAE data, a by-subject listing of AEs considered as immune-mediated events per investigator but not qualified for IMAEs definition will also be provided.

In addition, for all nivolumab treated subjects who experienced at least one IMAE, the following data presentation will be provided:

- Summary of subjects who were re-challenged with nivolumab by IMAE category, with extended follow-up. This output will also be presented for CPS ≥ 5 treated subjects.

For these, re-challenge is considered to have occurred when last nivolumab infusion was administered after the onset of an IMAE.

7.6.8 Other Events of Special Interest

OEOSI will be summarized by treatment group for each category.

The following analyses will be conducted using the 100-day safety window:

- Overall summary of OEOSI by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT
- Overall summary of drug-related OEOSI by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT

A by-subject listing of OEOSI will be provided.

7.6.9 Multiple Events

The following summary tables will be provided:

- A table showing the total number and rate (exposure adjusted) of occurrences for all AEs.
- A table showing the total number and rate (exposure adjusted) of occurrences for AEs occurring in at least 5% of subjects in any treatment group.

A listing displaying the unique instances of all AEs, ie, after duplicates have been eliminated and overlapping and contiguous occurrences of the same event (ie, same PT) have been collapsed will be provided. No formal comparisons will be made between treatment groups.

7.6.10 Laboratory Parameters

The analysis population for each laboratory test is restricted to treated subjects who underwent that laboratory test.

A by-subject listing of differences in categorization of SI and US laboratory test results will be provided.

7.6.10.1 Hematology

The following will be summarized by treatment group as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: hemoglobin (HB), platelets, white blood counts (WBC), absolute neutrophils count (ANC) and lymphocyte count (LYMPH). This output will also be presented for CPS ≥ 5 treated subjects.

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these laboratory parameters will be provided.

7.6.10.2 Serum Chemistry

The following will be summarized by treatment group as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: ALT, AST, total bilirubin and creatinine. This output will also be presented for CPS ≥ 5 treated subjects.

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these laboratory parameters will be provided.

7.6.10.3 Electrolytes

The following will be summarized by treatment group as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: sodium (high and low), potassium (high and low), calcium (high and low) and Glucose Serum (fasting hyperglycemia and hypoglycemia regardless of fasting status). This output will also be presented for all ≥ 5 treated subjects.

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these laboratory parameters will be provided.

7.6.10.4 Additional Analyses

In addition, further analyses on specific laboratory parameters will be performed by treatment group:

Abnormal Hepatic Function Test

The number of subjects with the following laboratory abnormalities from on-treatment evaluations will be summarized by treatment group:

- ALT or AST > 3 x ULN, > 5 x ULN, > 10 x ULN and > 20 x ULN
- Total bilirubin > 2 x ULN
- Concurrent (within 1 day) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these specific abnormalities will be provided.

Abnormal Thyroid Function Test

The number of subjects with the following laboratory abnormalities from on-treatment evaluations will be summarized by treatment group:

- TSH value > ULN and
  - with baseline TSH value ≤ ULN
  - with at least one FT3/FT4 test value < LLN within 2-week window after the abnormal TSH test
  - with all FT3/FT4 test values ≥ LLN within 2-week window after the abnormal TSH test
  - with FT3/FT4 missing within 2-week window after the abnormal TSH test.
- TSH < LLN and
  - with baseline TSH value ≥ LLN
  - with at least one FT3/FT4 test value > ULN within 2-week window after the abnormal TSH test
  - with all FT3/FT4 test values ≤ ULN within 2-week window after the abnormal TSH test
  - with FT3/FT4 missing within 2-week window after the abnormal TSH test

The analyses will be conducted using the 30-day safety window.

A by-subject listing of these specific abnormalities will be provided.

7.6.11 Vital Signs

Vital signs will be listed by subject.

7.6.12 Physical Measurements

Physical measurements will be listed by subject.

7.6.13 Immunogenicity Analysis

Further details on immunogenicity background and rationale, definitions, population for analyses and endpoints are described in APPENDIX 3.

Incidence of ADA

Number (%) of subjects will be reported for the following parameters based on nivolumab treated subjects with baseline and at least one post-baseline assessment. For the nivo +chemo arm the following parameters will be reported for nivolumab and for the nivolumab + ipilimumab arm the following parameters will be reported for nivolumab and ipilimumab.

- Baseline ADA-positive
- ADA-positive
  - Persistent Positive (PP)
  - Not PP-Last Sample Positive
  - Other positive
  - Neutralizing Positive
- ADA-negative

A listing of all ADA assessments will be provided together with results for neutralizing antibody.

A spider plot of nivolumab ADA test result (titers) over time may be provided for nivolumab ADA positive subjects.

Clinical implications

Clinical implications of nivolumab or ipilimumab immunogenicity will be primarily focused on subjects with persistent ADA-positive relative to ADA-negative. Subjects with any ADA-positive samples after initiation of treatment (relative to baseline) may be used to explore clinical implications. Effect of immunogenicity on clearance of nivolumab or ipilimumab will be explored by comparison of clearance estimates (determined by PPK analysis). Effect of immunogenicity on safety will be explored by examining the frequency and type of AEs of special interest such as hypersensitivity/infusion reaction. Summary tables for incidence of overall and each of the preferred terms will be provided, if the number of subjects is of sufficient size (eg, at least 10 subjects). Otherwise, individual subject's safety profile will be examined and described based on a listing. Clinical implications on efficacy will also be explored similarly. Association between trough concentrations of nivolumab or combination drug (eg, ipilimumab) and ADA assessments may be explored, as needed.

The following data presentation will be provided:

- Swimmer plot of occurrence of ADA and NAb Occurrence in Relation to PFS per BICR, BOR per BICR and OS

7.6.14 Pregnancy

A by-subject listing of pregnancy tests results will be provided for randomized female subjects.

7.6.15 Adverse Events By Subgroup

Overall summary of any AEs and drug-related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT and for each treatment group for the following subgroups:

- Sex (Male vs. Female)
- Race
- Age (< 65 vs. 65 - < 75 vs. 75 - < 85 vs. ≥ 85 vs. ≥ 75 vs. ≥ 65)
- Region (Asia [including China], US, RoW)

7.7 Pharmacokinetic Analysis

The nivolumab concentration data obtained in this study will be combined with data from other studies in the clinical development program to develop a population PK model. This model will be used to evaluate the effects of intrinsic and extrinsic covariates on the PK of nivolumab. In addition, exposure-response analyses with selected efficacy and safety endpoints will be conducted. Results of population PK and exposure response-analyses will be reported separately.

7.8 Biomarker Analysis

These analyses will be descriptive and not adjusted for multiplicity.

The analyses will be conducted on the following populations:

- For PFS and OS analyses of nivo + chemo versus chemo, the analyses will be performed on all randomized subjects or randomized subjects with quantifiable biomarker (see Section 6.3.1)
- For OS analyses of nivolumab + ipilimumab versus chemotherapy, the analyses will be performed on all randomized subjects or randomized subjects with quantifiable biomarker (see Section 6.3.2)

7.8.1 Tumor and Immune PD-L1 (CPS)

The Combined Positivity Score (CPS) is a scoring system taking into account immunoreactivity for PD-L1 in both tumor cells and tumor infiltrating immune cells (restricted to lymphocytes and macrophages) within or directly associated with tumor cell strands. As such, CPS is calculated as the ratio of the number of PD-L1 positive cells (tumor cells, lymphocytes and macrophages) divided by the total number of viable tumor cells within the evaluated tumor area, multiplied by 100. Although the maximum of this calculation may be greater than 100, the maximum CPS is defined as 100. The CPS score will be reported on a continuous scale.

CPS PD-L1 expression missing: Subjects without an available tumor biopsy specimen for PD-L1 evaluation will be considered CPS PD-L1 expression missing.

For subjects with an available tumor biopsy specimen(s), the following will be considered:

CPS PD-L1 expression not quantifiable will be further classified as

- Indeterminate: Tumor cell membrane staining hampered for reasons attributed to the biology of the tumor biopsy specimen and not because of improper sample preparation or handling
- Not evaluable: Tumor biopsy specimen was not optimally collected or prepared (eg, PD-L1 expression is neither quantifiable nor indeterminate)

CPS PD-L1 status is a dichotomized variable and will use different cutoffs (1, 5, and 10).

7.8.1.1 Distribution of CPS PD-L1

Descriptive statistics of CPS PD-L1 expression:

- Listing of all CPS PD-L1 data, all randomized subjects.
- Summary of tumor specimen acquisition and characteristics, all randomized subjects.
- Summary statistics of CPS PD-L1 expression in all randomized subjects with quantifiable PD-L1 expression.
- Cumulative distribution plot of baseline CPS PD-L1 expression versus population percentile in all randomized subjects with quantifiable CPS PD-L1 expression.

7.8.1.2 Association between PD-L1 Status and Efficacy

CPS PD-L1 categories used for these analyses are:

- Each CPS PD-L1 status subgroup (CPS < 5 and CPS ≥ 5, CPS < 1 and CPS ≥ 1, CPS < 10 and CPS ≥ 10)
- PD-L1 indeterminate or not evaluable subgroup

For both OS and PFS (using primary definition, per BICR and per Investigator), the following analyses will be provided and within each category above:

- curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% CIs for median OS and PFS will be computed by Brookmeyer and Crowley method Two-sided log-rank test comparing treatment arms
- HR (with corresponding two-sided 95% CI) will be estimated via a Cox model with treatment arm as the only covariate in the model
- Forest plot of HR with 95% CIs
- Figure with 6 KM curves for each combination of CPS PD-L1 status subgroup and treatment group.

For DOR (BICR) the following analyses will be provided and within each category above:

- curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% CIs for median DOR will be computed by Brookmeyer and Crowley method Two-sided log-rank test comparing treatment arms

- Figure with 6 KM curves for each combination of CPS PD-L1 status subgroup and treatment group.

For ORR, the following analysis will be provided using BICR data within each category:

- An estimate of the response rate and an associated exact two-sided 95% CI (Clopper and Pearson) will be presented for each subgroup categorized by treatment arm. An associated odds ratio and 95% CI will be computed. An estimate of the difference in ORRs and corresponding 95% CI will be calculated.

7.8.1.3 Predictive Relationship between PD-L1 Status and Efficacy

For both OS and PFS (using primary definition, per BICR and per Investigator), a Cox proportional hazards regression model will be fitted with treatment, CPS PD-L1 status, and treatment by CPS PD-L1 status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- HR of treatment vs. control and its associated 95% CI for each of the CPS PD-L1 status subgroup
- HR CPS PD-L1 $\geq$ cutoff vs. < cutoff and its associated 95% CI within each treatment group

7.8.2 Tumor PD-L1

The tumor cell PD-L1 positive scoring takes into account the percent of tumor cell membrane staining in a minimum of 100 evaluable tumor cells per validated Dako PD-L1 IHC assay unless otherwise specified.

Tumor Cell PD-L1 expression missing: Subjects without an available tumor biopsy specimen for PD-L1 evaluation will be considered as Tumor Cell PD-L1 expression missing.

For subjects with an available tumor biopsy specimen(s), the following will be considered:

Tumor Cell PD-L1 expression is not quantifiable will be further classified as

- Indeterminate: Tumor cell membrane staining hampered for reasons attributed to the biology of the tumor biopsy specimen and not because of improper sample preparation or handling
- Not evaluable: Tumor biopsy specimen was not optimally collected or prepared (eg, PD-L1 expression is neither quantifiable nor indeterminate)

Tumor Cell PD-L1 status is a dichotomized variable and will use different cutoffs (1,5, and 10)

7.8.2.1 Distribution of Tumor Cell PD-L1

Descriptive statistics of Tumor Cell and CPS PD-L1 expression:

- Listing of all Tumor Cell PD-L1 IHC data, all randomized subjects.
- Summary of tumor specimen acquisition and characteristics, all randomized subjects.
- Summary statistics of Tumor Cell PD-L1 expression in all randomized subjects with quantifiable PD-L1 expression.
- Cumulative distribution plot of baseline Tumor Cell PD-L1 expression versus population percentile in all randomized subjects with quantifiable PD-L1 expression.

7.8.2.2 Association between Tumor Cell PD-L1 Status and Efficacy

Analyses will be based on all randomized subjects if not otherwise specified. PD-L1 categories used for these analyses are:

- Each PD-L1 status subgroup (< 5 and ≥ 5, < 1 and ≥ 1, < 10 and ≥ 10)
- Tumor Cell PD-L1 indeterminate or not evaluable subgroup

For both OS and PFS (using primary definition, per BICR and per Investigator), the following analyses will be provided and within each category above:

- curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% CIs for median OS and PFS will be computed by Brookmeyer and Crowley method Two-sided log-rank test comparing treatment arms
- HR (with corresponding two-sided 95% CI) will be estimated via a Cox model with treatment arm as the only covariate in the model
- Forest plot of HR with 95% CIs
- Figure with 6 KM curves for each combination of Tumor Cell PD-L1 status subgroup and treatment group.

For DOR (BICR) the following analyses will be provided and within each category above:

- curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% CIs for median DOR will be computed by Brookmeyer and Crowley method Two-sided log-rank test comparing treatment arms
- Figure with 6 KM curves for each combination of Tumor Cell PD-L1 status subgroup and treatment group.

For ORR, the following analysis will be provided using BICR data within each category.

- An estimate of the response rate and an associated exact two-sided 95% CI (Clopper and Pearson) will be presented for each subgroup categorized by treatment arm. An associated odds ratio and 95% CI will be computed. An estimate of the difference in ORRs and corresponding 95% CI will be calculated.

7.8.2.3 Predictive Relationship between Tumor Cell PD-L1 Status and Efficacy

For both OS and PFS (using primary definition, per BICR and per Investigator), a Cox proportional hazards regression model will be fitted with treatment, Tumor Cell PD-L1 status, and treatment by Tumor Cell PD-L1 status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- HR of treatment vs. control and its associated 95% CI for each of the Tumor Cell PD-L1 status subgroup
- HR Tumor Cell PD-L1 ≥ cutoff vs. < cutoff and its associated 95% CI within each treatment group.

7.8.3 MSI

7.8.3.1 MSI Distribution

MSI categories (MSI-H, MSS, Invalid, unknown) will be summarized once on all randomized and once on all randomized subjects with MSI status available.

7.8.3.2 Association between MSI Status and Efficacy

For OS and PFS ((using primary definition, per BICR and per Investigator), the following analyses will be provided within each MSI category above for all randomized subjects.

- curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% CIs for median OS and PFS will be computed by Brookmeyer and Crowley method Two-sided log-rank test comparing treatment arms
- HR (with corresponding two-sided 95% CI) will be estimated via a Cox model with treatment arm as the only covariate in the model
- Forest plot of HR with 95% CIs
- Figure with 8 KM curves for each combination of MSI status subgroup and treatment group

For DOR (BICR) the following analyses will be provided and within each category above for all randomized subjects:

- curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% CIs for median DOR will be computed by Brookmeyer and Crowley method Two-sided log-rank test comparing treatment arms
- Figure with 8 KM curves for each combination of CPS PD-L1 status subgroup and treatment group

For ORR, the following analysis will be provided using BICR data within each category for all randomized subjects:

- An estimate of the response rate and an associated exact two-sided 95% CI (Clopper and Pearson) will be presented for each subgroup categorized by treatment arm. An associated odds ratio and 95% CI will be computed. An estimate of the difference in ORRs and corresponding 95% CI will be calculated.

7.8.3.3 Predictive Relationship between MSI Status and Efficacy

For both OS and PFS, a Cox proportional hazards regression model will be fitted with treatment, MSI status, and treatment by MSI status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported. The following analyses will be produced for all randomized subjects:

- Interaction p-value
- HR of treatment vs. control and its associated 95% CI for each of the MSI status subgroup
- HR between MSI categories and associated 95% CI's within each treatment group

7.9 Clinical Outcome Assessments

All questionnaires completed at baseline and on-study will be assigned to a time-point according to the windowing criteria in the table below and included in the analysis. In case a subject has two on-study assessments within the same window, the assessment closest to the time-point will be used. In the case of two assessments at a similar distance to the time-point, the latest one will be chosen. If there are two assessments reported on the same date, the assessment corresponding to the earliest visit would be used for analysis reporting purposes. In the event where the subject has no assessment at all in a specific window, the observation will be treated as missing for that time-point.

Time Windows for EQ-5D-3L, FACT-Ga

Nominal Time-Point	Time Window
Baseline	Treated subjects: Prior to first dose on Day 1 (assessment date <= first dose date; assessment time should not be used because of its unreliability) Randomized never treated subjects: Assessment date <= randomization date
Week 7	Treated subjects only: Nominal day is 43 Day 2 thru Day 64, inclusive
Week X (Every 6 weeks thereafter)	Treated subjects only: Nominal Day of previous visit + 42 Range of window: [nominal day -20, nominal day +21] Note: any assessment that is after the last dose will be slotted to the follow-up visit (see below).

Time Windows for EQ-5D-3L, FACT-Ga

Nominal Time-Point	Time Window
Follow-Up 1[a]	Compute Day as (Assessment date - last dose date +1) if Day is in [2, 73]
Follow-Up 2	compute day as (Assessment date - last dose date +1) if day is in [74, 160]
Survival Follow-up visit i (every 3 months from follow-up 2) I=1,2,3....	Compute Day as (Assessment date - last dose date +1) Nominal day is 115 + i * 91 If Day is in [Nominal day -45, Nominal day + 45]

[a] Only FACT-GaCS, FACT G7 and EQ-5D-3L are collected in follow-up 1, follow-up 2 and survival follow-up.

The analysis of FACT-Ga and EQ-5D-3L will be restricted to randomized subjects who have an assessment at baseline and at least one post-baseline assessment.

The analyses will be conducted on the following populations:

- For PFS and OS analyses of nivo+chemo versus chemo, the analyses will be performed on randomized CPS ≥ 5 subjects, randomized CPS ≥ 1 subjects and on all randomized subjects (see Section 6.3.1)
- For OS analyses of nivo+ipi versus chemotherapy, the analyses will be performed on randomized CPS ≥ 5 subjects, randomized CPS ≥ 1 subjects and on all randomized subjects (see Section 6.3.2)

7.9.1 FACT-Ga and TTSD

TTSD is an exploratory endpoint for the comparison of nivo+chemo vs. chemo and the analyses of TTSD will be descriptive. TTSD is a secondary objective for the comparison of nivo+ipi vs. chemo in the hierarchy and will be tested formally if the preceding secondary endpoint in the hierarchy is significant. See section 4.2.2.4 for the definition of TTSD.

TTSD will be assessed for the different CPS cutoffs and for all randomized subjects.

TTSD will be compared between nivo+ipi and chemo via stratified log-rank test at a two-sided α level with α inherited from the preceding endpoint in hierarchy. The two-sided log-rank p-value will be reported.

The estimate of the TTSD hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole indicator covariate (for each pair of treatment arms separately). Ties will be handled using the exact method. A two-sided 95% for the hazard ratio will also be presented. CI In case TTSD is formally tested the 100 x (1 - α)% CI will be provided.

The TTSD function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. Medians and two-sided 95% CI in each treatment group will be computed via the log-log transformation method.

Sensitivity Analyses for TTSD

- *Stratified analysis using stratification factors as obtained from the baseline CRF pages* (instead of IRT). This analysis will be performed only if at least one stratification factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the population.
- *Analysis accounting for death*. TTSD will be defined similarly to the primary definition except that for patients who die without deterioration, death will be counted as an event.
- *Analysis accounting for assessment on/after treatment discontinuation:* TTSD will be defined similarly to the primary definition except that events (deterioration or death) and assessments that occurred on or after treatment discontinuation will be considered (no time point truncation)

For the FACT-Ga items are scored on categorical scales. Some items (score between 0 and 4) are reversed so that a higher score represents a better quality of life.

The following descriptive analyses will be conducted:

- FACT-Ga questionnaire completion rate, defined as the proportion of questionnaires actually received out of the expected number (ie, number of subjects on treatment or in follow up) will be calculated and summarized for each assessment time point by treatment group. As of follow-up visit 1 and after, limited questionnaire are collected for GaCS and Fact-G7. Completion rate will be provided for the follow up visits only for those two scales.
- Mean score and mean change from baseline in individual wellbeing scales, GaCS, FACT-G7, FACT-G, and FACT-Ga total scores will be summarized at each assessment time point using descriptive statistics (ie, N, mean with SD and 95% CI, median, first and third quartiles, minimum, maximum). As of follow-up visit 1 and after, limited questionnaire are collected for GaCS and Fact-G7. Summary statistics will be provided for the follow up visits only for those two scales.
- A line graph summarizing the mean changes from baseline will be produced for GaCS, FACT-G7, FACT-G, and FACT-Ga total scores.

7.9.2 EuroQol EQ-5D-3L

The following descriptive analyses will be conducted:

- EQ-5D-3L questionnaire completion rate, defined as the proportion of questionnaires actually received out of the expected number (ie, number of subjects on treatment or in follow up), will be calculated and summarized for each assessment time point by treatment group.
- A by-subject listing of the level of problems in each dimension, corresponding to EQ-5D-3L health state (ie, 5-digit vector) and EQ-5D-3L VAS score will be provided.
- Proportion of subjects reporting problems for the 5 EQ-5D-3L dimensions at each assessment time point will be summarized by level of problem and by treatment group. Percentages will be based on number of subjects assessed at assessment time point.
- For the EQ-5D-3L VAS scores and utility index:

- Mean score and mean change from baseline at each assessment time point will be summarized by treatment group using descriptive statistics (N, mean with SD and 95% CI, median, first and third quartiles, minimum, maximum).
- A line graph summarizing the mean changes from baseline will be produced.

7.10 COVID-19 Related Analyses

In this study, in order to evaluate the impact of COVID-19 pandemic, the following CRF pages were implemented.

- Disposition: The subjects who discontinue the study treatment or discontinue the study due to COVID-19.
- Exposure: The subjects who have study therapy modification due to COVID-19.
- Contact Information: The subjects who miss visits due to COVID-19.

Listings will be provided based on the data collected on each CRF pages and additional analyses may be performed in order to evaluate the impact of COVID-19 on this study.

8 ANALYSES TO EVALUATE THE CONTRIBUTION OF COMPONENTS IN THE N+I REGIMEN

8.1 Ipilimumab contribution

The analyses described in this section will help to evaluate the contribution of components (CoC); specifically that ipilimumab contributes to the N+I combination. With nivolumab treatment common to both regimes, an analysis that isolates the ipilimumab contribution in N+I relative to chemo in the N+C combination will be conducted.

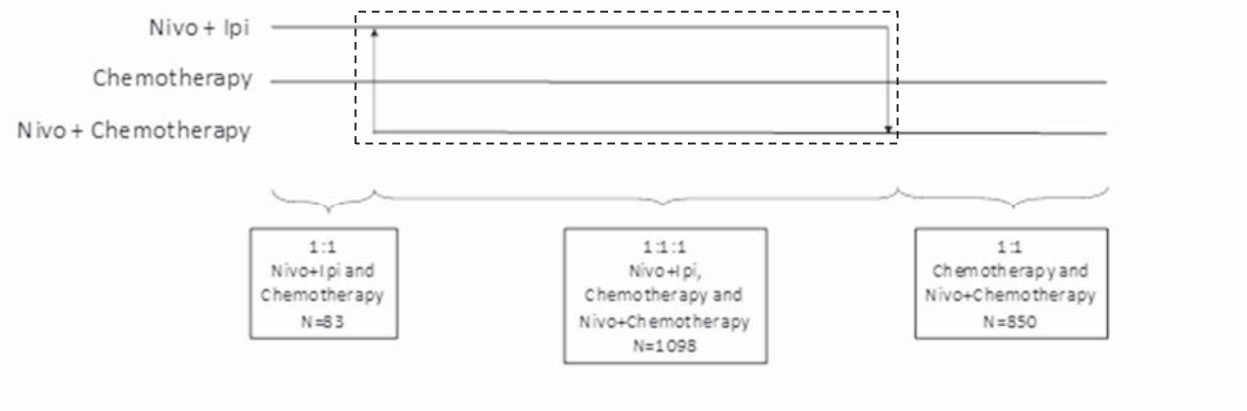
The concurrently randomized analysis is understood to be the valid comparison. To support this analysis, a sensitivity analysis on all randomized subjects will be conducted in consideration of the non-overlapping randomization, which has the potential to introduce bias. The analysis of this broader group considering the totality of the data is intended to be supportive with a balance of increased precision and less validity.

The following two populations will be used for these analyses comparing N+I to N+C:

All Concurrently Randomized Subjects to 3 Arms: all subjects concurrently randomized to nivo+chemo or chemo or nivo+ipi from 27-Mar-2017 (included) to 05-Jun-2018 (N~1098, see Figure 8.1-1).

All Randomized Subjects to 3 Arms: all subjects randomized to nivo+chemo or chemo or nivo+ipi. This entails all subjects randomized in the study (N~2031).

Figure 8.1-1: Randomization to the 3 Arms



All the analyses that will be conducted to support the contribution of components are considered exploratory intended for descriptive purpose. These analyses will not be adjusted for multiplicity (no p-value will be provided) and estimates will be accompanied by 95% CI.

The following analyses will be produced for the 2 populations described above and for those 2 populations restricted to subjects with PD-L1 CPS ≥ 5 and for PD-L1 CPS <5.

Unless stated otherwise, whenever a stratified analysis is specified in this section, the following stratifications factors (recorded at randomization as per IRT) will be used see Table 7.5-1

Table 8.1-1: Stratification Factors Used for Stratified Analyses Supporting CoC

- Region (Asia vs. US vs. RoW)
- ECOG performance status (0 vs. 1)
- PD-L1 expression level (≥ 1% vs. < 1% or indeterminate)

8.1.1 OS and PFS

For OS and PFS as assessed by BICR the HR between nivo+ipi and nivo+chemo with its associated two-sided 95% CIs will be estimated via a stratified Cox model with treatment arm as the only covariate in the model. For the smaller subset some small stratum are possible, so in addition the unstratified Cox model will be implemented as a sensitivity analysis. The HR of nivo+chemo vs. chemo and nivo+ipi vs. chemo will also be estimated using the same population. Overall survival and PFS as assessed by BICR (using both primary and secondary definitions, as defined in Sections 4.1.2.1 and 4.1.2.2, respectively) for each of the three treatment arms will be estimated and overlayed using the KM product-limit method. Median survival time along with 95% CI will be constructed based on a log-log transformed CI for the survival function.

The analyses above will be repeated for the following populations:

- All concurrently randomized subjects to the 3 arms
- All concurrently randomized subjects to the 3 arms with PD-L1 CPS ≥ 5

- All concurrently randomized subjects with PD-L1 CPS < 5
- All randomized subjects to the 3 arms
- All randomized subjects to the 3 arms with PD-L1 CPS ≥ 5
- All randomized subjects to the 3 arms with PD-L1 CPS < 5

For OS, considering that the survival function may potentially have different shapes for the N+I arm and the N+C arm, the restricted mean survival time (RMST) will be evaluated for each of these two arms. The difference of RMST with its associated two-sided 95% CIs will be estimated. The RMST will be estimated by the area under the KM curve up to several time points (12 months, 24 months, 36 months, and the minimum of the longest survival time in each treatment arm, regardless of censoring). Since follow-up differs for the arms, these analyses will be done solely on all subjects concurrently randomized subjects to the 3 arms, all concurrently randomized subjects to the 3 arms with PD-L1 CPS ≥ 5 and for all concurrently randomized subjects to the 3 arms with PD-L1 CPS <5.

8.1.2 ORR and DOR

For ORR and DOR as assessed by BICR, ORR will be computed in each of the 3 treatment groups along with the exact 95% CI using Clopper-Pearson method. An estimate of the difference in ORRs and corresponding 95% CI between each pair of treatment arms will be calculated using CMH methodology and adjusted by the stratification factors as specified in Section 7.1.1. The stratified odds ratios (Mantel-Haenszel estimator) between each pair of treatment arms will be provided along with the 95% CI. DOR for each of 3 treatment groups will be estimated and overlayed using KM product limit method for subjects who achieve PR or CR. Median values along with two-sided 95% CI will be calculated.

These analyses will be produced for

- All concurrently randomized subjects to the 3 arms with measurable disease (BICR)
- All concurrently randomized subjects to the 3 arms with measurable disease (BICR) and PD-L1 CPS ≥ 5
- All concurrently randomized subjects to the 3 arms with measurable disease (BICR) and with PD-L1 CPS < 5.
- All randomized subjects to the 3 arms with measurable disease (BICR)
- All randomized subjects to the 3 arms with measurable disease (BICR) and PD-L1 CPS ≥ 5
- All randomized subjects to the 3 arms with measurable disease (BICR) and PD-L1 CPS < 5.

9 CONVENTIONS

The following conventions may be used for imputing partial dates for analyses requiring dates:

- For missing and partial adverse event onset dates, imputation will be performed using the Adverse Event Domain Requirements Specification[22].

- Missing and partial Non-Study Medication Domain dates will be imputed using the derivation algorithm described in Section 4.3.3 of BMS Non-Study Medication Domain Requirements Specification.[23]

For death dates, the following conventions will be used for imputing partial dates:

- If only the day of the month is missing, the 1st of the month will be used to replace the missing day. The imputed date will be compared to the last known date alive and the maximum will be considered as the death date.
- If the month or the year is missing, the death date will be imputed as the last known date alive.
- If the date is completely missing but the reason for death is present the death date will be imputed as the last known date alive

For date of progression, the following conventions will be used for imputing partial dates:

- If only the day of the month is missing, the 1st of the month will be used to replace the missing day.
- If the day and month are missing or a date is completely missing, it will be considered as missing.
- In case of the date of death is present and complete, the imputed progression date will be compared to the date of death. The minimum of the imputed progression date and date of death will be considered as the date of progression.

For other partial/missing dates, the following conventions may be used:

- If only the day of the month is missing, the 15th of the month will be used to replace the missing day.
- If both the day and the month are missing, “July 1” will be used to replace the missing information.
- If a date is completely missing, it will be considered as missing.

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

1 month = 30.4375 days and 1 year = 365.25 days.

Duration (eg, DoR, etc) will be calculated as follows:

Duration = (Last date - first date + 1)

All statistical analyses will be carried out using SAS (Statistical Analysis System software, SAS Institute, North Carolina, USA) unless otherwise noted.

10 CONTENT OF REPORTS

At the time of first analysis time point (12 months minimum follow-up), the final analysis of PFS and interim analysis of OS in PD-L1 CPS≥5 randomized subjects in the nivo+chemo vs chemo arms will be performed. Below are potential scenarios for the planning of contents report:

1) If both of the primary endpoints are not meeting the pre-specified statistical significance criteria, the study will continue and there is no unblinding of either endpoint. As such, no report is planned.
2) If PFS meets the pre-specified statistical criteria, a clinical study report for nivo+chemo vs. chemo will be written with all the analyses listed in this SAP except for these analyses related to OS.
3) If both primary endpoints meet the pre-specified statistical significance or only OS meets the statistical significance, a clinical study report for nivo+chemo vs. chemo will be written with all analyses listed in this SAP.
4) There will be no analysis performed for the nivo+ipi vs. chemo comparison.

At the second analysis time point (24 months minimum follow-up), a report for nivo+ipi vs. chemo will be written with all analyses listed in this SAP. A report for nivo+chemo vs chemo will be written with the primary endpoint(s) that have not yet reported at the first time point if any.

Refer also to the Data Presentation Plan for mock-ups of all tables and listings.

11 DOCUMENT HISTORY

Table 11-1: Document History

Version Number	Author(s)	Description
1 25-Apr-2017	[redacted]	Original SAP based on revised Protocol 02
2 02-Dec-2019	[redacted]	Updated to align with IO CORE SAP Updated to reflect revised protocol 09.

Table 11-1: Document History

Version Number	Author(s)	Description
3 29-May-2020	[redacted]	Based on revised protocol 09. Minor updates from SAP version 2.0 to add/modify sensitivity analyses and other minor edits. Details are summarized below: • Section 7.5.2.2, 7.5.2.5 and 7.5.3.2 Replace the Fleming-Harrington test F(0,1) by the max combo in case of NPH. • added potential sensitivity analysis in case of imbalance in PD-L1 CPS≥5 group in Sections 7.5.2.2, 7.5.2.5 and 7.5.3.2 • Update blinding and unblinding Section 2.3 regarding unblinding of the PFS results but not OS results in case that PFS is positive but OS is negative at the interim. • Section 4.1.2, added some text to state that PFS regardless of subsequent therapy is considered as primary analysis by EMA. The PFS account for subsequent therapy is the primary endpoint of the study and will be the basis of the primary test. • For swimmer plot for ADA and Nab in relation to efficacy endpoint, added BOR per BICR • Section 7.8.3.2, corrected typo, 6 KM curves for each combination of biomarker status and treatment • Corrected 24-Mar to 27-Mar in definition of All randomized subjects for nivo+chemo vs chemo. It has no impact on datasets, change was made for consistency. • Added AE tables for the all treated subjects with CPS>=5. Added IMAE outputs for all treated subjects with CPS>=5. • Added surgery to the list of concurrent prohibited anti-cancer therapy to reflect protocol. • Changed “Planned Chemotherapy” to “Chemotherapy” for subgroup analyses as we will use the Chemotherapy as treated. • Correction to Time Window for EQ-5D-3L, FACT-Ga, for the survival for follow-up. • Removed Siewert-Stein classification from the list of subsets for PFS, OS and ORR as this is collected only for GEJ. • Added Baseline albumin (< LLN, ≥LLN) to the list of subsets for PFS, OS and ORR • Censoring efficacy endpoints for subsequent therapy. Subjects are censored for radiotherapy with reason curative and added reason palliative. • Added UNKNOWN category for MSI • added appendix 4 with analyses of data from China

Table 11-1: Document History

Version Number	Author(s)	Description
4.0 04-Aug-2020	[redacted]	Added Appendix 5 to update the expected number of subjects, expected number of PFS and OS events and updated power based on the actual CPS>=5 prevalence. The prevalence was estimated using the locked database of July 10, 2020. The estimation was done in a blinded manner using the pooled estimate for the 3 arms.
5.0 14-May-2021	[redacted]	This update was implemented after the analysis of PFS and OS IA for nivo+chemo vs chemo for which the CSR was written. The updates do not affect any of the planned analyses for nivo+ipi vs chemo. The updates include: • Update to Section 7.4.1, to allow reporting of the overall RDI for nivolumab in the nivo+ipi treatment arm, the formula was changed. • Specific section (Section 7.10) related to COVID 19 data reporting. • Specific section (Section 8) with exploratory analyses to support the justification of the ipilimumab contribution in the nivo+ipi regimen.

APPENDIX 1 TIME-TO ONSET AND TIME-TO RESOLUTION DEFINITION AND CONVENTIONS FOR SELECT ADVERSE EVENTS, IMMUNE-MEDIATED ADVERSE EVENTS AND EVENTS OF SPECIAL INTEREST

Time-to onset definition

Time-to onset of AE (any grade) for a specific category is defined as the time between the day of the first dose of study treatment and the onset date of the earliest AE (of any grade) in this category.

The time-to onset of AE (grade 3-5) for a specific category is defined similarly with an onset date corresponding to a grade 3-5 AE.

Time-to onset of drug-related AE (any grade or grade 3-5) for a specific category is defined similarly but restricted to drug-related AE.

Time-to onset for a specific subcategory is defined similarly but restricted to event of this subcategory.

Time-to resolution definition

In order to derive the time-to resolution, overlapping or contiguous AEs within a specific category or subcategory will be collapsed into what will be termed "clustered" AEs. For example, if a subject (without pre-treatment AE) experienced an AE from 1st to 5th January, another AE (with different PT but within same category) from 6th to 11th January and same AE from 10th to 12th January, these will be collapsed into one clustered AE from 1st to 12th January. Table 1 is summarizing key derivation steps for each type of clustered AEs.

Time-to resolution of AE (any grade) for a specific category is defined as the longest time from onset to complete resolution or improvement to the grade at baseline among all clustered AEs experienced by the subject in this category per adverse event criteria category. Events which worsened into grade 5 events (death) or have a resolution date equal to the date of death are considered unresolved. If a clustered AE is considered as unresolved, the resolution date will be censored to the last known alive date. Improvement to the grade at baseline implies that all different events in the clustered adverse event should at least have improved to the corresponding (ie, with same preferred term) baseline grade. This measure is defined only for subjects who experienced at least one AE in the specific category.

The time-to resolution of AE (grade 3-5) for a specific category is defined similarly with an onset date corresponding to a grade 3-5 AE.

Time-to resolution of drug-related AE (any grade or grade 3-5) for a specific category is defined similarly but restricted to drug-related AE.

The time-to resolution of AE (any grade or grade 3-5, drug-related or all) where immune modulating medication was initiated is defined similarly. For data presentation not restricted to IMAE, the additional condition that the subject started an immune modulating medication during the longest AE resolution period will be applied.

Time-to resolution for a specific subcategory is defined similarly but restricted to event of this subcategory.

Table 1 Derivation of Clustered AE

Type of clustered AE	Derivation
Any grade	Collapse any on-treatment AE from the same category
Drug-related of any grade	Collapse any on-treatment drug-related AE from the same category
Grade 3-5	Collapse any on-treatment AE from the same category. Resolution will be based on the onset date of the earliest grade 3-5 records (if no grade 3-5 record, clustered AE is excluded)
Drug-related of Grade 3-5	Collapse any on-treatment drug-related AE from the same category Resolution will be based on the onset date of the earliest grade 3-5 record (if no Grade 3-5 record, clustered AE is excluded)

The algorithm for collapsing adverse event records is using the following conventions:

For each subject and specified category, the corresponding adverse event records will be collapsed when:

1) Multiple adverse event records have the same onset date.
2) The onset date of an event record is either the same day or 1 day later than the resolution date of a preceding event record (contiguous events).
3) The onset date of an event record is after the onset date and prior to or on the resolution date of a preceding event record (overlapping events).

APPENDIX 2 MISSING AND PARTIAL RADIOTHERAPY AND SURGERY DATES IMPUTATION ALGORITHMS

Procedures – Imputation Rules.

If reported procedure start date is a full valid date then set start date equal to the date part of procedure start date.

In case of partial date use imputation rules described below:

- If only day is missing then
  - If month and year of procedure match month and year of first dose date then impute as date of first dose;
  - If month and year of procedure don't match month and year of first dose date then impute as first day of that month and year.
- If both day and month are missing, then impute as maximum between 01JAN of the year and date of the first dose;
- If date is completely missing or invalid then leave missing.

Note: Imputation is not applicable to data where start date is not collected (for example "PRIOR RADIOTHERAPY" CRF). Set start date to missing in this case.

If reported end date is a full valid date then set end date equal to the date part of the reported end date.

In case of partial date use imputation rules described below:

- If reported end date is partial then set end date equal to the last possible reported end date based on the partial entered reported end date.
- If reported end date is missing, continuing, unknown or invalid then set end date equal to the most recent database extraction date.

If end date was imputed then compare end date to the death date or last known alive date if subject is not dead. If posterior then end date should be imputed to death date (or last known alive date if subject not dead).

Note: Imputation of partial dates only applies to data entered on "RADIOTHERAPY" CRF page. For other CRF pages in case of partial dates set end date to missing.

Surgeries – Imputation Rules.

If reported surgery date is a full valid date then set start date equal to the date part of surgery date.

In case of partial date, use one of the two imputation rules described below:

A. For data collected on "PRIOR SURGERY RELATED TO CANCER" CRF page:

- If only day is missing then impute as the first day of the month;
- If both day and month are missing then then impute as 01JAN of the year;

- If date is completely missing or invalid then leave missing.

B. For data collected on other CRF pages (deemed to be on-treatment/subsequent surgeries):

- If only day is missing then
  - If month and year of surgery match month and year of first dose date then impute the missing date as the date of first dose;
  - If month and year of surgery don't match month and year of first dose date then impute as first day of that month and year;
- If both day and month are missing then impute as maximum between 01JAN of the year and date of the first dose;
- If date is completely missing or invalid then leave missing.

APPENDIX 3 IMMUNOGENICITY ANALYSIS: BACKGROUND AND RATIONALE

The following summary is from the FDA Guidance for Industry Immunogenicity Assessment for Therapeutic Protein Products and White Paper on Assessment and Reporting of the Clinical Immunogenicity of Therapeutic Proteins and Peptides – Harmonized Terminology and Tactical Recommendations by Shankar et al. The program-level definitions of sample- and subject-level ADA status are based on recommendation from the BMS Immunogenicity Council.

Immune responses to therapeutic protein products may pose problems for both patient safety and product efficacy. Immunologically based adverse events, such as anaphylaxis and infusion reactions, have caused termination of the development of therapeutic protein products or limited the use of otherwise effective therapies. Unwanted immune responses to therapeutic proteins may also neutralize the biological activity of therapeutic proteins and may result in adverse events not only by inhibiting the efficacy of the therapeutic protein product, but by cross-reacting to an endogenous protein counterpart, if present. Because most of the adverse effects resulting from elicitation of an immune response to a therapeutic protein product appear to be mediated by humoral mechanisms, circulating antibody has been the chief criterion for defining an immune response to this class of products.

ADA is defined as biologic drug-reactive antibody, including pre-existing host antibodies that are cross-reactive with the administered biologic drug (baseline ADA). Titer is a quasiquantitative expression of the level of ADA in a sample. By employing a serial dilution-based test method, titer is defined as the reciprocal of the highest dilution of the sample (eg, dilution of 1/100 = titer of 100). The ADA is also tested, via a cell-based biologic assay or a non cell-based competitive ligand-binding assay for a subpopulation of ADA known as neutralizing antibodies (NAb), which inhibits or reduces the pharmacological activity of the biologic drug molecule regardless of its in vivo clinical relevance. Non-neutralizing ADA (non-NAb) is ADA that binds to the biologic drug molecule but does not inhibits its pharmacological activity.

ADA should be tested using sensitive and valid methods and employing an appropriate strategy for elucidating immunogenicity. Detection of ADA is typically performed in three tiers (screening, confirmatory, and titer) using statistically determined cutpoints and samples testing positive in the ADA assay are analyzed for neutralizing activity, especially in late-stage clinical studies. “Detection” of ADA implies that drug-specific ADA was confirmed. The ‘drug tolerance” of an assay (highest drug concentration that does not interfere in the ADA detection method) is not an absolute value and differs between individuals due to the varying avidities of ADA immune responses. An ADA sampling strategy of collecting samples at times when the least drug concentration is anticipated (trough concentrations) can increase the likelihood of accurate ADA detection.

It is useful to present ADA results from clinical studies as (a) characteristics of the ADA immune response, (b) relationship of ADA with pharmacokinetics (PK) and, when relevant, pharmacodynamics (PD) biomarkers, and (c) relationship of ADA with clinical safety and efficacy.

Clinical consequences of ADA can range from no apparent clinical effect to lack of efficacy (primary treatment failure), loss of efficacy (secondary treatment failure) or heightened effect due to altered exposure to the biologic drug, adverse drug reactions (administration-related systemic or site reactions), and severe adverse drug reactions (anaphylaxis and unique clinical problems associated with cross-reactivity and neutralization of endogenous molecules). Thus it becomes important to examine any associations between ADA or any of its attributes with the various clinical sequelae. The presence of ADA may or may not preclude the administration of drug to ADA-positive subjects because the outcome is dependent upon the magnitude of the impact of ADA on PK and PD. Hence, the relationship of ADA with PK/PD is an important additional consideration, but does not necessarily result in a clinically impactful consequence per se.

Immunogenicity Endpoints

A fundamental metric that informs clinical immunogenicity interpretation is the incidence of ADA in a study or across comparable studies. ADA incidence is defined as the proportion of the study population found to have seroconverted or boosted their pre-existing ADA during the study period.

Terms and Definitions

Validated ADA test methods enable characterization of samples into ADA-positive vs. ADA-negative. To classify the ADA status of a subject using data from an in vitro test method, each sample from the subject is categorized based on the following definitions:

Sample ADA Status:

- Baseline ADA-positive sample: ADA is detected in the last sample before initiation of treatment
- Baseline ADA-negative sample: ADA is not detected in the last sample before initiation of treatment
- ADA-positive sample: After initiation of treatment, (1) an ADA detected (positive seroconversion) sample in a subject for whom ADA is not detected at baseline, or (2) an ADA detected sample with ADA titer to be at least 4-fold or greater (≥) than baseline positive titer
- ADA-negative sample: After initiation of treatment, ADA not positive sample relative to baseline

Next, using the sample ADA status, subject ADA status is defined as follows:

Subject ADA Status:

- Baseline ADA-positive subject: A subject with baseline ADA-positive sample
- ADA-positive subject: A subject with at least one ADA positive-sample relative to baseline at any time after initiation of treatment

1) Persistent *Positive (PP):* ADA-positive sample at 2 or more consecutive time points, where the first and last ADA-positive samples are at least 16 weeks apart
2) *Not PP-Last Sample Positive:* Not persistent positive with ADA-positive sample at the last sampling time point

3) Other Positive: Not persistent positive but some ADA-positive samples with the last sample being negative
4) Neutralizing Positive: At least one ADA-positive sample with neutralizing antibodies detected

- **ADA-negative subject:** A subject with no ADA-positive sample after the initiation of treatment.

(Note: 16 weeks was chosen based on a long half-life of IgG4.)

Population for Analyses

Analysis of immunogenicity data will be based on ADA evaluable subjects defined as all treated subjects with baseline and at least 1 post-baseline immunogenicity assessment. Analysis dataset and data listing will include all available ADA samples. However, subject-level ADA status will be defined based on only adequate samples (eg, excluding 1-hour post-infusion samples when clearly indicated).

APPENDIX 4 ANALYSES OF DATA FROM CHINA

Analysis methods for China

Most analyses detailed in SAP will be repeated for the patients randomized in China using the analysis sets described in this appendix. The analysis methods for the China subpopulation will be the same as for the global population unless otherwise noted. No formal hypothesis testing will be performed to evaluate consistency of China subpopulation. Instead, descriptive statistics will be provided to assess the consistency. All statistical analyses for China will be considered exploratory and only performed when at least one of the primary endpoints is statistically significant.

No adjustment for multiplicity will be made.

Given that all patients in China subpopulation are from geographic region Asia, region (Asia vs. US vs. RoW) will be not be used as stratification factors in stratified analyses conducted for the China.

Definition of analysis sets

China subpopulation

Chinese subpopulation is defined as China subgroup. It should contain subjects that are Chinese by race and enrolled from China.

Asian subpopulation

Asian subpopulation is defined as Asian subgroup. It should contain subjects that are Asian by race and enrolled from Asian countries.

APPENDIX 5 SAMPLE SIZE UPDATE

This appendix is completed after the database lock for the efficacy interim analysis (dated July 10, 2020) and before the DMC efficacy review meeting (planned on August 6, 2020). The interim analysis consists of the final analysis of PFS and interim analysis of OS in subjects with PD-L1 CPS ≥ 5 for the comparison of nivolumab plus chemotherapy vs chemotherapy.

During the development of the SAP v3.0 the actual prevalence of PD-L1 CPS ≥ 5 in study CA209649 was unknown. For that reason, it was not possible to estimate the expected number of PFS and OS events. The expected final number of OS events is needed to derive the adjusted significance level for the interim analysis of OS using the Lan-DeMets α spending function with O'Brien and Fleming type of boundary. The significance level will be adjusted according to the fraction of actual OS events observed at the interim against the estimated final number of events. At final analysis of OS, the significance level will be calculated using the number of events in the database at time of database lock and considering the α-level already spent at interim analysis using the EAST Interim Monitoring Tool.

There were 2031 subjects randomized to the study. Based on the different periods of randomization (see Table 5.1-1 Randomization Allocation and Sample Size), it is estimated that approximately 1582 subjects were randomized concurrently to nivolumab+chemotherapy or chemotherapy.

The prevalence of PD-L1 CPS ≥ 5 using the data from July 10, 2020 (blinded and pooled over the 3 arms) is 60% among all randomized subjects. Based on this prevalence, it is estimated that the number of subjects with PD-L1 CPS ≥ 5 who were randomized to nivolumab + chemotherapy and chemotherapy will be 949.

OS Events Estimation Using Actual CPS prevalence

The final number of OS events is obtained via simulations in EAST 6.4.1. The OS model is described in Section 5.2.1 and consists of a 2-piece hazard ratio; a delayed effect with a HR of 1 versus chemotherapy for the first 6 months followed by a constant HR of 0.65. The distribution for the OS of the chemotherapy arm is a 4-piece hazard (see Table 5.2.1-3). Based on these simulations the final number of events at final analysis among 949 subjects is estimated to be 800.

Update of Power Calculation for PFS and OS

As indicated in section 8.1.2 of the protocol, "the power statement of the primary endpoints of PFS and OS will be updated using the actual prevalence and documented in the final SAP prior to first database lock for efficacy analysis." Based on the observed prevalence of PD-L1 CPS ≥5 of 60% among all randomized subjects, Table 5.2.1-2 and Table 5.2.1-4 are therefore updated with Table 1 and Table 2. Specifically, the number of subjects with PD-L1 CPS≥5, the expected number of events at interim and final, and the power at interim and final were updated and are highlighted in the tables.

Similarly for the nivolumab+ipilimumab comparions, based on the observed prevalence of PD-L1 CPS ≥5 of 60% among all randomized subjects and α=0.035 hierarchically passed from the

primary nivolumab+chemotherapy versus chemotherapy results at the primary database lock on 10Jul2020, Table 5.2.1-2 is therefore updated with Table 3.

Table 1 Updated Summary of Sample Size Parameters and Schedule of Analyses for PFS (Nivolumab + Chemotherapy vs. Chemotherapy)

	3 months Delay	6 months Delay
# subjects with PD-L1 CPS ≥ 5[a]	949	949
Hypothesized delayed period	3	6
Hypothesized HR after delayed period	0.56	0.56
Hypothesized median in control arm	5.5 months	5.5 months
Significance level (2-sided)	0.02	0.02
Enrollment Period (from start of 1:1:1)	25.4 months	25.4 months
Minimum follow-up / Expected number of events[a,b,c]	12 months /841	12 months /857
Time of Analysis	37.4 months	37.4 months
power[b]	99.9%	84.0%
average HR[b]	0.66	0.80

[a] Based on 60% prevalence of PD-L1 CPS ≥ 5

[b] Results based on simulations.

[c] Number of events for the nivolumab plus chemotherapy vs. chemotherapy comparison in total.

Table 2 Updated Summary of Sample Size Parameters and Schedule of Analyses for OS (Nivolumab + Chemotherapy vs. Chemotherapy)

# with PD-L1 CPS ≥ 5[a]	949
Hypothesized delayed period	6 months
Hypothesized HR after delayed period	0.65
Hypothesized median in control arm	11.1 months
Significance level (2-sided)	0.03
Enrollment Period (from start of 1:1:1)	25.4 months
INTERIM ANALYSIS #1 for OS	
Minimum follow-up/Expected number of events[a,b,c]	12 months/680
Time of analysis[d]	37.4 months

Table 2 Updated Summary of Sample Size Parameters and Schedule of Analyses for OS (Nivolumab + Chemotherapy vs. Chemotherapy)

Significance level[e]	0.0164
Power	87.3%
FINAL ANALYSIS	
Minimum follow-up/Expected number of events[a,b,c]	24 months/800
Time of analysis[d]	49.4 months
Significance level	0.0252
Power[b]	97.9%
Average HR	0.74

[a] Based on a 60% prevalence of PD-L1 CPS ≥ 5.

[b] Results based on simulations

[c] Number of events for the nivolumab plus chemotherapy vs. chemotherapy comparison in total.

[d] After first patient randomized to nivolumab + chemotherapy or chemotherapy in 1:1:1 randomization.

[e] Significance levels will be calculated based on the actual number of deaths at each interim analysis

Table 3: Updated Summary of Sample Size Parameters and Schedule of Analyses for OS (Nivolumab + Ipilimumab vs. Chemotherapy)

	Alpha=0.035
# with PD-L1 CPS ≥ 5[a]	489
Hypothesized distribution	See Table 5.2.2-1
Hypothesized median in control arm	11.1 months
Significance level (two-sided)	0.035
Enrollment Period (from start of 1:1)	18.7 months
Time of analysis /Expected number of events[a,b,c]	54.4/411
Power[b]	93%
Average HR[b]	0.70

[a] Based on 60% prevalence of PD-L1 CPS ≥ 5

[b] Results based on simulations.

[c] Number of events for the nivolumab plus ipilimumab vs. chemotherapy comparison in total.

12 REFERENCES

1 Dolan P. Modeling valuations for EuroQol health states. Medical Care 1997;35: 1095-1108

2 The EuroQol Group: EuroQol: A new facility for the measurement of health-related quality of life—The EuroQol Group. Health Policy 16:199-208, 1990.

3 Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 2007;5:70.

4 Tabernero J, Bang Y-J, Fuchs CS et al. KEYNOTE-062: Phase III study of pembrolizumab (MK-3475) alone or in combination with chemotherapy versus chemotherapy alone as first-line therapy for advanced gastric or gastroesophageal junction (GEJ) adenocarcinoma. ASCO 2019 abstract LBA4007.

5 Charles S. Fuchs, et al. Pembrolizumab (pembro) vs paclitaxel (PTX) for previously treated advanced gastric or gastroesophageal junction (G/GEJ) cancer: Phase 3 KEYNOTE-061 trial. ASCO 2018 abstract. 4062

6 Fuchs C , Shitara K, Di Bartolomeo M et al RAINFALL: A randomized, double-blind, placebo-controlled phase III study of cisplatin (Cis) plus capecitabine (Cape) or 5FU with or without ramucirumab (RAM) as first-line therapy in patients with metastatic gastric or gastroesophageal junction (G-GEJ) adenocarcinoma.ASCO GI Cancers Symposium 2018, abstract # 5,

7 Ohtsu A, Shah MA, Cutsem EV, et al. Bevacizumab in Combination With Chemotherapy As First-Line Therapy in Advanced Gastric Cancer: A Randomized, Double-Blind, Placebo Controlled Phase III Study. J Clin Oncol 2011;29:3968-3976.

8 Lordick F, Kang YK, Chung HC, et al. Capecitabine and cisplatin with or without cetuximab for patients with previously untreated advanced gastric cancer (EXPAND): a randomised, open-label phase 3 trial. Lancet Oncol 2013;14:490.

9 Lan KKG, Lachin JM, Implementation of Group Sequential Logrank Tests in a Maximum Duration Trial, Biomatrics 46, 759-770, September 1990.

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

11 Klein, J. P. and Moeschberger, M. L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag

12 Global Biometric Sciences, SAS Analysis Dataset Specification, Unique Adverse Events. Version 2.0, April 22, 2015

13 Greenwood, M. The errors of sampling of the survivorship tables, Reports on Public Health and Statistical Subjects, 33, Appendix 1, HMSO, London, 1926

14 Kalbfleisch, J. D. and Prentice, R. L. (1980), The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons.

[15] Willi Maurer & Frank Bretz (2013) Multiple Testing in Group Sequential Trials Using Graphical Approaches, Statistics in Biopharmaceutical Research, 5:4, 311-320

[16] Gail, M. and Simon, R. "Testing for qualitative interactions between treatment effects and subject subsets". Biometrics, 1985; 41:361-372.

[17] Analysis of survival by tumor response.J R Anderson, K C Cain and R D Gelber. JCO 1983 vol. 1 no. 11 710-719

[18] Clopper, C.; Pearson, E. S. (1934). "The use of confidence or fiducial limits illustrated in the case of the binomial". *Biometrika* **26**: 404–413.

[19] DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials, 1986; 7: 177-188

[20] Newcombe, R.G. Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in Medicine, 17, 857-872 (1998).

[21] Analysis of survival by tumor response. J R Anderson, K C Cain and R D Gelber. JCO 1983 vol. 1 no. 11 710-719

[22] Adverse Event Domain Requirements Specification. Bristol-Myers Squibb Co. PRI. Version 2.4.0. July 25, 2016.

[23] Non-Study Medication Domain Requirements Specification. Bristol-Myers Squibb, Co. PRI. Version 2.9 July 25, 2016