Official Title of Study:

A Randomized, Multicenter, Double Blind, Phase III Study of Adjuvant Nivolumab or Placebo in Subjects with Resected Esophageal, or Gastroesophageal Junction Cancer

(CheckMate 577: CHECKpoint pathway and nivoluMab clinical Trial Evaluation 577)

NCT Number: NCT02743494

Document Date (Date in which document was last revised): September, 30 2020

STATISTICAL ANALYSIS PLAN
FOR CLINICAL STUDY REPORT

A Randomized, Multicenter, Double Blind, Phase III Study of Adjuvant Nivolumab or Placebo in Subjects with Resected Esophageal, or Gastroesophageal Junction Cancer

Protocol CA209-577

VERSION # 4.0

DATE: 30-Sept-2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

[redacted]

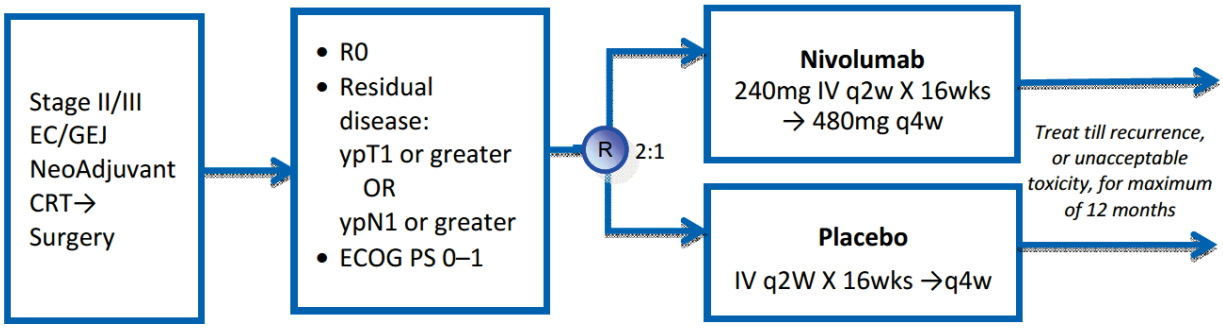
[redacted]

2 STUDY DESCRIPTION

2.1 Study Design

The study design schematic is presented in Figure 1.

Figure 1: Study Design



This is a phase 3, randomized, double-blind, placebo controlled study of adjuvant nivolumab in subjects with resected EC, or GEJ cancer who have received CRT followed by surgery.

After CRT followed by surgery, subjects will sign the informed consent form (ICF). Approximately 760 subjects whose tumors do not achieve pCR (non-pCR) will be randomized in a blinded fashion in a 2:1 ratio to two arms between nivolumab (BMS-936558) or placebo monotherapy administered IV over 30 minutes at 240mg every 2 weeks for 16 weeks (8 doses) followed by nivolumab 480mg as a 30 minute infusion every 4 weeks beginning at Week 17 (2 weeks after the 8th dose). The treatment will be given until disease recurrence, unacceptable toxicity, or subject withdrawal of consent with a maximum of 1-year total duration of study medication.

This study will consist of three phases: screening, treatment, and follow-up.

Screening phase:

- Begins by establishing the subject's initial eligibility and signing of the ICF. Subject must receive pre-operative CRT followed by curative surgery.
- Subjects must have been surgically rendered free of disease with negative margins on resected specimens, i.e., R0. Subjects whose tumors do not achieve pCR, i.e., pathological status ≥ ypN1 or ≥ ypT1 are eligible.
- The pathology reports of detectable lesion(s) confirming malignancy must be reviewed, dated, and signed by the investigator prior to randomization
- Subjects must have PD -L1 IHC testing, with results, performed by the central lab during the Screening period.

- All subjects must have disease-free status documented by a complete physical examination and imaging studies within 4 weeks prior to randomization. Imaging studies must include CT scan of chest, and abdomen.

Treatment phase:

- Following confirmation of the subject's eligibility, the randomization entry to the IWRS can be made. The subject is randomly assigned in a 2:1 ratio to the nivolumab or to the placebo arm.
- Administration of nivolumab or placebo is to begin within 3 calendar days of randomization.
- Treated subjects will be evaluated for recurrence every 12 weeks ± 1 week.
- This phase ends when the subject is discontinued early from study therapy (i.e., disease recurrence, unacceptable toxicity, or subject withdrawal of consent) or at a maximum of 1 year of treatment.

Follow up phase:

- Begins after 1 year of treatment or when the decision is made to discontinue a subject from study therapy.
- After completion of the first two follow -up visits (FU 1 at day 30 days ± 1 week and FU 2 at 84 days ± 1 week from FU1), subjects will be followed every 3 months or more frequently as needed for survival.
- Subjects who discontinue treatment for reasons other than distant recurrence will continue to have surveillance assessments (until distant recurrence) every 12 weeks ± 1 week during the first year after randomization, every 12 weeks ± 2 week during the second year, after that follow local standard in the range of 6-12 months between year 3 and year 5 with the last assessment at year 5.

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 IWRS to obtain the subject number. Every subject that signs the informed consent form must be assigned a subject number in IWRS.

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

Subjects meeting all eligibility criteria will be randomized in a 2:1 ratio to nivolumab or placebo **stratified by the following factors**:

1) PD-L1 status (≥ 1% vs. < 1% or indeterminate)
2) Pathologic lymph node status (positive ≥ ypN1vs. negative ypN0)
3) Histology (squamous vs. adenocarcinoma).

2.3 Blinding and Unblinding

The Sponsor, subjects, investigator and site staff will be blinded to the study therapy administered. Each investigative site must assign an unblinded pharmacist/designee, and an unblinded site monitor will be assigned to provide oversight of drug supply and other unblinded study documentation.

Blinding of treatment assignment is critical to the integrity of this clinical study. However, in the event of a medical emergency or pregnancy in an individual subject in which knowledge of the investigational product is critical to the subject's management, the blind for that subject may be broken by the investigator. The subject's safety takes priority over any other considerations in determining if a treatment assignment should be unblinded. For this study, the method of unblinding for emergency purposes is through the IWRS.

At the first and second interim looks of this study, if the superiority is demonstrated in DFS but not in OS, the subjects, investigator and site staff will remain blinded to the treatment information and the study will continue for OS. A restricted BMS team will be unblinded to the DFS data, but not unblinded the aggregated OS results by treatment arm (unless requested by the health authorities. The restricted BMS team will then determine whether to unblind the BMS project team in order to prepare regulatory filings.

2.4 Protocol Amendments

Global amendments incorporated in the protocol (Version 3.0, dated 06-Jun-2019) with relevant changes are described in Table 2.4-1.

Table 2.4-1: Summary of Protocol Amendment

Document	Date of Issue	Summary of Change
Revised Protocol 03	30-May-2019	• Added exclusion criteria, on-study, and post-study requirements regarding live/attenuated vaccines. • Updated language regarding hepatitis B or C virus exclusion criteria. • Updated language for WOCBP in Section 3.3.3 and Appendix. • Moved OS from co-primary endpoint to secondary endpoint. • [redacted] • Added language regarding monitoring for infusion-related reactions. • Added language regarding dose interruptions, delays, and discontinuation. • Added myocarditis to Grade 3 non-skin drug-related adverse events (AEs). • Removed AE assessment from survival follow-up visits. • Clarified language regarding diagnosis of recurrence. • Added option for review of reports by a blinded independent central review at a later date. • Replaced original sample-size determinations based on co-primary endpoints.

Table 2.4-1: Summary of Protocol Amendment

Document	Date of Issue	Summary of Change
		• Added new section to provide data for the assumptions regarding DFS and OS in the control arm. • Added new section to provide data for the assumed treatment effect. • Added new section to update sample size and power estimates based on new assumptions. • Updated to provide new triggers and timing for the interim and final analyses. • Added explanation of how timing of DFS/OS analyses will be adjusted to maintain a strong control of type I error. • Moved discussion of addressing family-wise error rate across DFS and OS analyses at interim and final analyses to a new section. • Moved error-rate discussion to new section and removed mention of OS as a primary endpoint. • Added two separate sections to address each the DFS and OS analyses. Made multiple updates to bring protocol in line with current program standards.
Administrative Letter 03	01-Aug-2017	Medical Monitor information updated
Revised Protocol 02	04-May-2017	Incorporates Amendment 06 and Administrative Letter 02
Amendment 06	04-May-2017	The main purpose of the amendment was to modify the inclusion criteria to increase the time between complete resection and randomization from 4-14 weeks to 4-16 weeks. Other changes incorporated included: • [redacted] • Revised the estimated enrollment and study duration, time to achieve 455 disease-free survival (DFS) events, and time to achieve 330 deaths and 440 deaths based on the current subject accrual rate • Revised the maximum dose delay window to 42 days during Cycles 1-8 and 70 days during Cycles 9-17 • Revised the study design/schematic to remove the reference to 'distant' recurrence • Revised the screening window from 28 days to 49 days • Revised the study drug dosing window. For Cycles 1-8, subjects may have study drug administered up to 2 days before or 3 days after the scheduled dosing date. For Cycles 9-17, subjects may be dosed within a +/- 3 day window. • [redacted]

Table 2.4-1: Summary of Protocol Amendment

Document	Date of Issue	Summary of Change
		• [redacted] • Revised the Flow Chart/Time and Events Schedule to resolve minor inconsistencies and to provide clarifications • Fixed typos and resolved minor inconsistencies
Administrative Letter 02	30-Nov-2016	To correct a formatting issue with the indenting of the bullets in protocol Section 4.5.2 (Dose Delay Criteria).
Revised Protocol 01	24-Aug-2016	Incorporates Amendment 05 and Administrative Letter 01
Amendment 05	24-Aug-2016	The main purpose the amendment was to: • Modify the nivolumab Dose Delay Criteria (Section 4.5.2), Criteria to Resume Treatment (Section 4.5.4), and Discontinuation of Subjects from Treatment (Section 4.5.5) criteria to align with the US Package Insert and EU Summary of Product Characteristics Other changes incorporated included: • Change the BMS Medical Monitor • Revised the study schematic • Clarify that subjects will receive their randomized treatment (nivolumab or placebo) for the duration of the On-Treatment Period • Changed the term BMS and BMS Medical Monitor to Sponsor or designee • Clarified that a blinded independent central review may occur during or at the end of the trial • Increased the time from complete resection to randomization to 4-14 weeks • Specified the order of priority of the imaging modalities for this trial • Specified that adverse events will be documented for a minimum of 100 days after last dose of study drug • Specified that during the Follow-Up phase survival visits may be may be accomplished by in-person visit or phone contact • Revised certain Inclusion/Exclusion criterion • Updated the duration of contraception use for WOCBP and males subjects with female partners that are WOCBP • Removed the methods of contraception from protocol Section 3.3 • Removed reference to unblinded site staff and an unblinded site monitor • Added information regarding resuming dosing following resolution of an AE or immunosuppression tapering

Table 2.4-1: Summary of Protocol Amendment

Document	Date of Issue	Summary of Change
		• Increased the number of tumor slides from the surgically resected specimen from 10 to 20 slides • Specified that 5 slides would be required for the optional tumor samples • Added albumin to the list of analytes required at the Screening Visit • Allowed for Total T3/T4 to be reported by the lab if free T3/T4 are not available based on site capabilities • Updated the On-Treatment Procedural Outline notes to reflect that assessments should be performed prior to dosing at the required Cycles • Removed reference to the plasma samples in the On-Treatment Procedural Outline table • Clarified that urinalysis is required at the Follow-Up visits if clinically indicated • [redacted] • Specified the collection timepoints for the Outcomes Research Assessments during the On-Treatment Period • Added language that allows for additional pregnancy testing to be performed during the Follow-Up Period • Separated the Safety Assessment section of the protocol into sub-sections based on the study phase • Included information regarding pulmonary adverse events and treatment • [redacted] • Added information regarding immune-mediated adverse events • Added information regarding AE and SUSAR reporting • Added information that a female partner of a male subject must sign an informed consent form to disclose information regarding a pregnancy • Updated terminology used for the statistical censoring scheme • Updated the abbreviations list • Revised the Appendix 2 (Safety Management Algorithms) • Added Appendix 3 (Women of Childbearing Potential and Methods of Contraception) Other minor changes incorporated into this amendment include changes in document names, study materials that will be provided to sites, removal of duplicate statements, revisions to section numbering, and formatting changes
Administrative Letter 01	10-Feb-2016	• Fixed the protocol title • Removed the reference to neck as an anatomical imaging area for the CT/MRI scan Updated a section number
Original Protocol	06-Jan-2016	Not applicable

2.5 Role of DMC

A DMC will be established to provide oversight of safety and efficacy considerations in protocol CA209577. 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. 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 be also responsible for reviewing and making recommendations at the two interim looks of efficacy. (see Section 1).

3 OBJECTIVES

3.1 Primary

- To compare DFS of nivolumab versus placebo in subjects with resected EC or GEJ cancer.

3.2 Secondary

- To compare OS of nivolumab versus placebo in subjects with resected EC or GEJ cancer.
- To evaluate 1, 2, and 3 year survival rates of nivolumab versus placebo in subjects with resected EC or GEJ cancer.

[redacted]

4 ENDPOINTS

4.1 Efficacy Endpoints

4.1.1 Disease Free Survival

DFS is the primary endpoint.

Disease-Free Survival is time between randomization date and first date of recurrence or death, whichever occurs first. Recurrence is defined as the appearance of one or more new lesions, which can be local, regional, or distant in location from the primary resected site (by imaging or pathology). DFS will be programmatically determined based on the disease recurrence date provided by the investigator. For subjects who remain alive and without recurrence, DFS will be censored on the date of last evaluable disease assessment. The on-study disease assessment considers CT scan, MRI and biopsy

Detailed censoring rules for the primary definition of DFS are presented in Table 4.1.1-1; graphical display is presented in Figure 2.

Table 4.1.1-1: Censoring Scheme Used in the Primary Definition of DFS

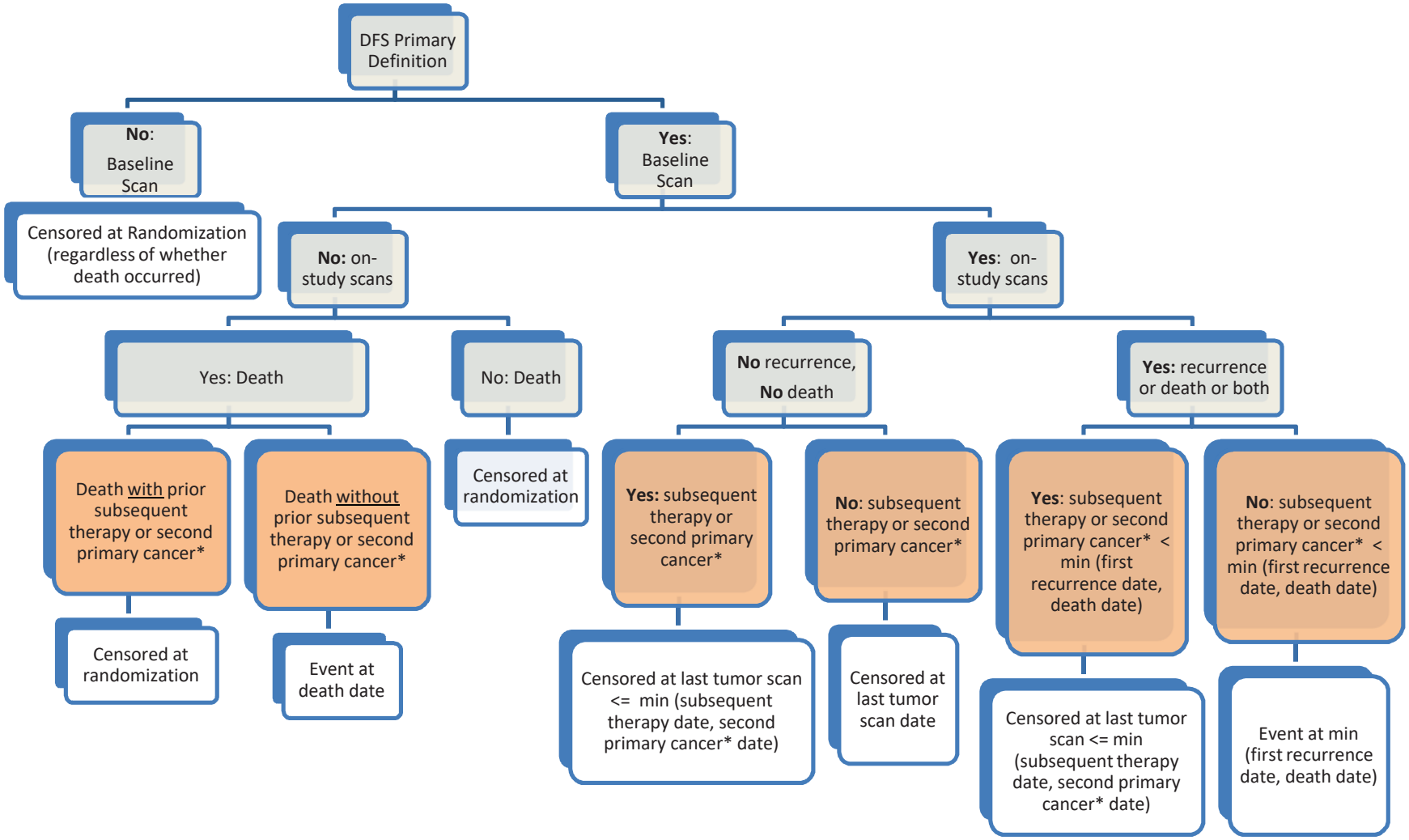
Situation	Date of Event or Censoring	Outcome
Recurrence (a) (c) (d) (e)	Date of first recurrence	Event
Death (b) (c) (d) without recurrence	Date of death	Event
No baseline disease assessment	Date of randomization	Censored
No on-study disease assessments and no death (d)	Date of randomization	Censored
Death with prior subsequent therapies or second non-esophageal and non-GEJ primary cancer (d)	Date of randomization	Censored
No recurrence and no death (d) (e)	Date of last evaluable disease assessment	Censored
New anticancer therapy, tumor-directed radiotherapy, or tumor-directed surgery received or the second non-esophageal and non-GEJ primary cancer date without recurrence or death reported prior to or on the same day of disease assessment (d) (e)	Date of last evaluable disease assessment prior to or on the same date of initiation of subsequent therapy or the second non-esophageal and non-GEJ primary cancer date whichever is earlier	Censored

Table 4.1.1-1: Censoring Scheme Used in the Primary Definition of DFS

(a) recurrence = appearance of one or more new lesions, which can be local, regional, or distant in location from the primary resected site
(b) All deaths will be included as DFS event - regardless of cause or of how long it has been since the last known disease evaluation.
(c) Without receiving preceding new anticancer therapy, tumor-directed radiotherapy, tumor-directed surgery, or second non-esophageal and non-GEJ primary cancer.
(d) Without baseline disease.
(e) With on-study disease assessment.

Per protocol inclusion criteria, subjects must have complete resection (R0), have been surgically rendered free of disease. Should a subjects not be disease-free by imaging at study entrance, this will be considered as a relevant protocol deviation per section 7.2 of this SAP. A subject with disease at baseline is considered as having a DFS event at randomization.

Figure 2: Graphic Display of DFS Primary Definition



* Second non-esophageal and non-GEJ primary cancer; A subject with disease at baseline is considered as having a DFS event at randomization.

DFS per BICR

For this study, all radiologic imaging will be transmitted to a centralized imaging core lab for storage by an independent review committee (BICR). These images will be read upon request, in which case DFS by BICR will also be derived.

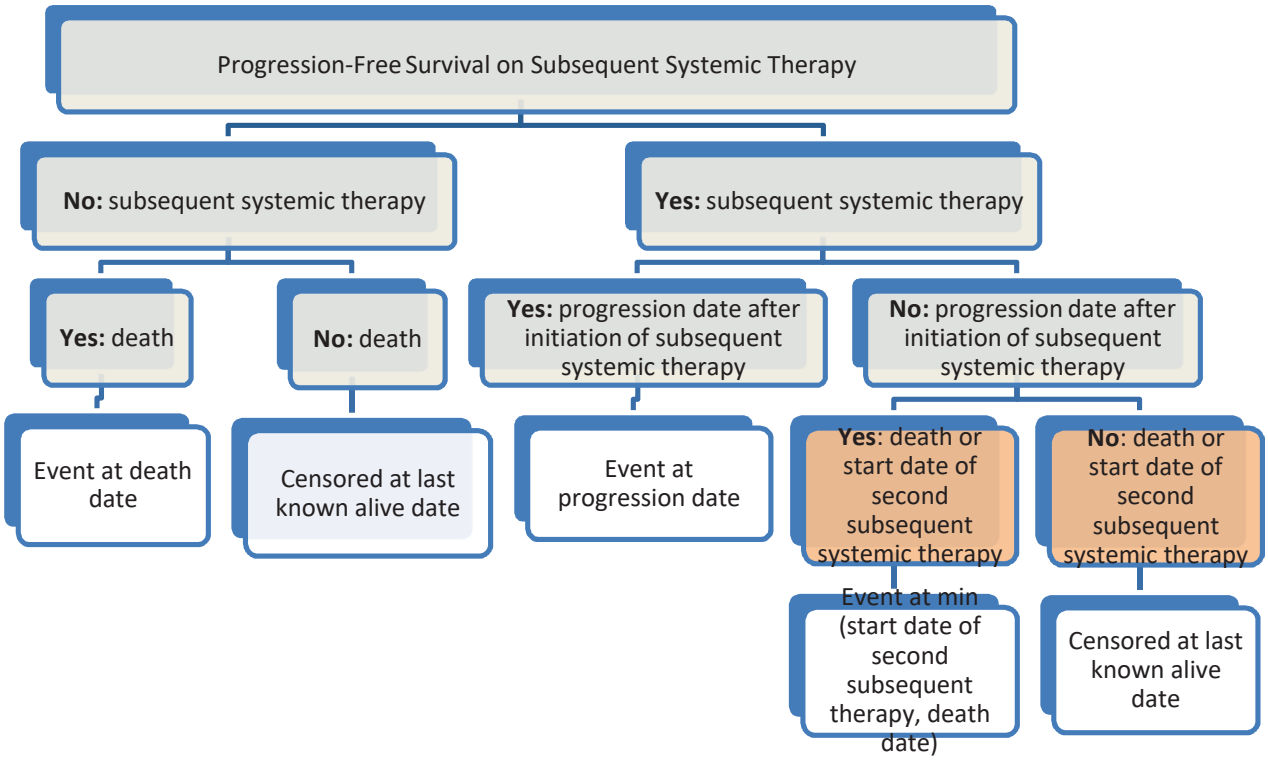
4.1.2 Overall Survival

Overall survival (OS) is defined as the time from 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.

4.1.3 Overall Survival Rates

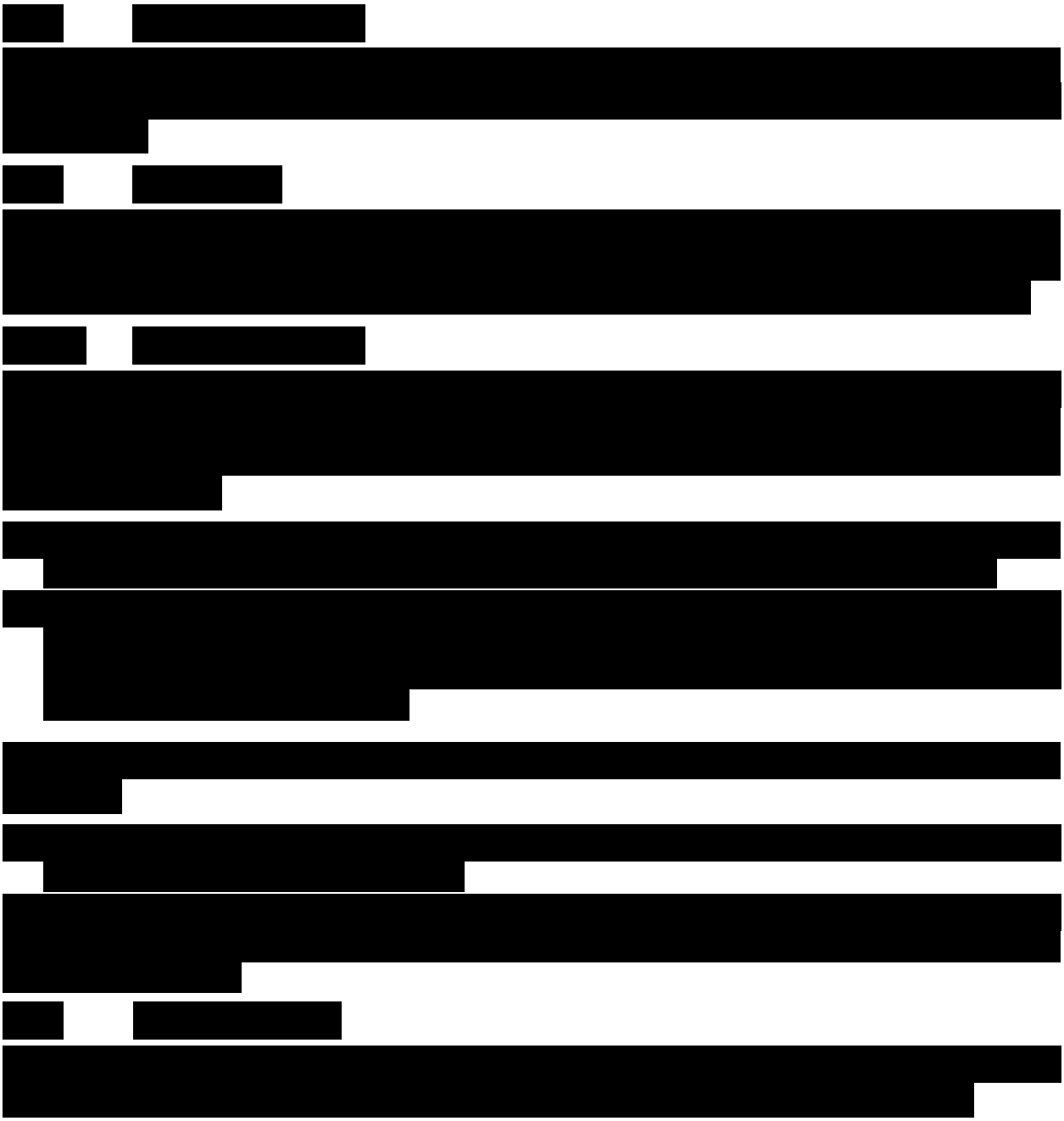
Overall survival rate at 1, 2, and 3 years is defined as the probability that a subject is alive at 1, 2, and 3 years using KM method, respectively, following randomization.

Figure 3: **Graphic Display of Progression-Free Survival on Subsequent Systemic Therapy**



4.2 Safety 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 (i.e. development of anti-drug antibody) will be analyzed.



[redacted]

All questionnaires completed at baseline and on-study will be assigned to a time-point according [redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

5 SAMPLE SIZE AND POWER

The sample size determination takes into consideration the comparison of the primary endpoint of DFS and the first secondary endpoint of OS between the 2 treatment arms.

5.1 Assumptions of the Control Arm

The assumptions of DFS and OS in the control arm of this study are made using the chemoradiotherapy followed by surgery (CRT + S) arm in the CROSS trial with long-term follow up (CROSS LT).[4]

Assumption for DFS

The DFS in the control arm is assumed to follow a piece-wise exponential distribution with a median of 21 months, as described below:

- The median PFS of the CRT+S arm in the CROSS LT trial was 37.7 months, which was defined from the start of CRT to disease progression.
- Considering the median time from the start of CRT+S to the completion of surgery is approximately 6 months, the median time from the completion of surgery to disease progression (to approximate DFS) is then estimated to be approximately 31 months in the setting of the CROSS LT trial.
- The study population in CA209577 are subjects who have residual pathologic disease, ie, nonpathologic complete response (non-pCR). This population represents a high-risk subgroup in the CROSS LT trial that included 29% subjects who achieved complete pathological response (pCR) after resection and 71% non-pCR subjects. This subgroup of non-pCR subjects have worse clinical outcome compared with pCR subjects.[5] Therefore, the assumed median DFS in CA209577 should be less than 31 months.
- The assumption of 21-months median DFS in the control arm is made in consultation with external experts. Additionally, a single-center study of 518 esophageal adenocarcinoma patients suggested that 85.1% and 94.4% of relapses occurred within 24 and 36 months of surgery, respectively.[6] This also supports the median DFS being less than 24 months in the study population.
- Chemoradiotherapy + surgery has demonstrated long-term survival benefit, as indicated by a long-lasting plateau toward the end of the survival curve in multiple clinical trials.[7, 8] As such, the distribution of the DFS curve in the control arm of this study is assumed to follow a piece-wise exponential distribution with landmark DFS hazard rates and corresponding landmark DFS rates in Table 5.1-1. The shape of the DFS curve in the control arm reflects the shape of the PFS curve in the CRT+S arm of the CROSS LT trial.

Assumption for OS

The OS in the control arm of this study is also assumed to follow a piece-wise exponential distribution with a median of 31 months. The assumption of the median OS is 10 months longer than the assumption of the median DFS in the control arm as described below:

- The median PFS and median OS of the CRT+S arm in the CROSS LT trial were 37.7 months and 48.6 months respectively, which give 11 months difference between median PFS and median OS.
- The differences between median PFS and median OS are all less than 12 months in the meta-analysis of 14 adjuvant gastric cancer studies. [9]

The shape of OS curve in the control arm of this study reflects the observed OS curve in the CRT+S arm of the CROSS LT trial. The assumed landmark OS hazard rates and corresponding landmark OS rates in this study are provided in Table 5.1-2.

Table 5.1-1: Assumed DFS Hazard Rates, Landmark Rates and Hazard Ratios

By Time	DFS Hazard Rate/DFS Rate (%) in the Control Arm	DFS Hazard Rate/DFS Rate (%) in the Nivolumab Arm	Hazard Ratio
3 months	0.035/90.0	0.035/90.0	1
1 year	0.035/65.7	0.023/73.0	0.667
2.5 years	0.03/38.3	0.02/50.9	0.667
5 years	0.02/21.0	0.013/34.1	0.667
10 years	0.003/17.5	0.002/29.5	0.8

Table 5.1-2: Assumed OS Hazard Rates, Landmark Rates and Hazard Ratios

By Time	OS Hazard Rate/OS Rate (%) in the Control Arm	OS Hazard Rate/OS Rate (%) in the Nivolumab Arm	Hazard Ratio
4 months	0.026/90.1	0.026/90.1	1
1 year	0.026/73.2	0.018/78.2	0.685
2.5 years	0.0205/50.6	0.014/60.7	0.685
6 years	0.018/23.8	0.012/36.2	0.685
12 years	0.002/20.5	0.002/32.2	0.8

5.2 Assumptions of the Treatment Effect

Assumption for DFS

An average HR of 0.72 is assumed for DFS between nivolumab and placebo in CA209577 at the final DFS analysis, which accounts for the piece-wise HRs summarized in Table 5.1-1. This translates to a median DFS of 31 months in the nivolumab arm.

In many trials studying immunotherapies, the PFS and OS curves in the immunotherapy-treated arms tend to have a delayed separation from the control arms across tumor indications;[10] therefore, the treatment effect for DFS is assumed to have a delayed separation with an HR of 1 for the first 3 months and a larger treatment effect after delayed separation with an HR of 0.667. After 5 years, the treatment effect for DFS is assumed to decrease slightly with an HR of 0.8. This takes into consideration that patients who have not had disease recurrence within 5 years of surgery are generally cured from the disease and the treatment effect is more reflective of the survival benefit.

Assumption for OS

The average HR of 0.73 is assumed for OS between nivolumab and placebo in CA209577 at the final OS analysis, which accounts for the piece-wise HRs summarized in Table 5.1-2. This translates to median OS of 44 months in the nivolumab arm.

The assumption of the OS treatment effect between the 2 arms follows the same thought process as the assumption of the DFS treatment effect. The treatment effect for OS is assumed to have a delayed separation with HR of 1 for the first 4 months, and a larger treatment effect after delayed separation with an HR of 0.685. After 6 years, the treatment effect for OS is assumed to decrease slightly with an HR of 0.8. This takes into consideration the combined effect of potential for more subsequent immunotherapies in the control arm and other effective subsequent therapies in both arms for long-term survivors.

5.3 Sample Size and Power

According to the assumptions for DFS described in Sections 5.1 and 5.2, the study will require approximately 760 subjects to be randomized at a 2:1 ratio to nivolumab and placebo and observations of at least 440 DFS events in order to achieve approximately 91% power to detect an average HR of 0.72 at a 2-sided alpha of 0.05.

The sample-size determination accounts for 1 DFS interim analysis. Details of the DFS analysis timing are provided in Section 5.4.

It is estimated that these 760 subjects will be accrued and randomized within approximately 36 months from the first-patient first-visit (FPFV).

Overall survival will be tested following the overall hierarchical testing procedure[11] upon demonstration of superiority in DFS at either interim or final analyses for all randomized subjects.

With the sample size of 760, it is required to observe at least 460 OS events at the final OS analysis in order to achieve approximately 90% power to detect an average HR of 0.73 at a 2-sided alpha of 0.05. The power of the OS final analysis accounts for 2 OS interim analyses that occur at the same time as the DFS interim and DFS final analyses, respectively. Details of the OS analysis timing are provided in Section 5.4.

5.4 Timing of the DFS and OS Interim and Final Analyses

This study includes two interim looks and one final look. The first interim look of this study will include the DFS interim analysis (DFS IA) and the first OS interim analysis (OS IA1). The second interim look of this study will include DFS final analysis (DFS FA) and the second OS interim analysis (OS IA2). The final look of this study will be OS final analysis (OS FA).

Timing of the Interim and Final DFS Analyses

Considering the potential plateau of the DFS curve after 3 years,[4] the study is designed with 1 interim DFS analysis when at least 85% of all 440 DFS events (374 DFS events) are observed. Based on the current assumptions, it is projected to occur approximately 12 months after the last

patient being randomized. The final DFS analysis will be triggered when at least 440 DFS events have been observed. The final DFS analysis is projected to occur approximately 22 months after the last patient being randomized. The stopping boundaries at the interim and final DFS analyses will be derived based on the exact number of DFS events observed using Lan-DeMets alpha spending function with O'Brien-Fleming boundaries.

Timing of the Interim and Final OS Analyses

Based on the current design, the first interim OS analysis is planned at the time of the DFS interim analysis, and the second interim OS analysis is planned at the time of the DFS final analysis. It is expected to have approximately 65% and 80% of OS events at the time of the interim DFS and final DFS analysis respectively. This formal comparison of OS will allow for early stopping of the study for superiority. The stopping boundaries at the interim and final OS analyses will be derived based on the exact number of deaths observed using Lan-DeMets alpha spending function with O'Brien-Fleming boundaries.

The final OS analysis will be triggered when at least 460 OS events have been observed and is projected to occur approximately 41 months after the last patient being randomized.

The timing and power of the interim and final DFS and OS analyses are summarized in Table 5.1-3.

The statistical software East version 6.3 (Cytel, Inc 2010) was used for study design calculations.

Table 5.1-3: Timing and Power of Interim and Final DFS and OS Analyses

	DFS	OS
Interim Analysis (IA) for DFS/IA 1 for OS		
Projected time from 1st patient randomized	48 months	48 months
Number of events	374 (85% of all DFS events)	299 (65% of all OS events)[a]
Significance level	0.03	0.01
Probability of crossing boundary	81%	45%
Final Analysis (FA) for DFS/IA 2 for OS		
Projected time from 1st patient randomized	58 months	58 months
Number of events	440	368 (80% of all events)[b]
Significance level	0.042	0.022
Cumulative Probability of crossing boundary	91%	70%

Table 5.1-3: Timing and Power of Interim and Final DFS and OS Analyses

	DFS	OS
FA for OS		
Projected time from 1st patient randomized	NA	77 months
Number of events		460
Significance level		0.042
Cumulative Probability of crossing boundary		90%

[a] IA 1 for OS is at the time of DFS interim analysis. The actual number of OS events may be different.

[b] IA 2 for OS is at the time of DFS final analysis. The actual number of OS events may be different.

5.5 Potential Adjustment of Analysis Timing for DFS and OS

As discussed in the Section 5.1 of this SAP, there was no historical trial with the exact same population studied in this trial. As such, the assumptions made for the DFS and OS in the control arm might be very different from the actual rates. Since the events rates will impact the data maturity and timing of DFS and OS analyses, BMS will conduct periodic monitoring of the DFS and OS events rates in a pooled, blinded fashion under the assumed HR of DFS and OS once a majority of the subjects are randomized. Should such monitoring suggests a big variability of the assumptions made for DFS and OS in the control arm, the impact on the data maturity and timing of analyses for DFS and OS will be evaluated and the trigger for the timing of the DFS and OS analyses may be adjusted accordingly.

If the blinded monitoring suggested a modification of the trigger of the interim looks, this document will be used to document such changes prior to the first interim look of this study.

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

- Baseline period:
  - Baseline evaluations or pre-treatment 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, vital signs, and biomarkers) 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, vital signs, and biomarkers) will be defined as evaluations with a date (and time if collected) on or prior to the date of first dose of study treatment.
- If there are multiple valid observations in the baseline period, then the latest non missing observation will be used as the baseline in the analyses. If multiple observations exist on the latest collection date (and time if collected), the record with the latest data entry date and time will be used. If multiple observations exist on the latest collection date (and time if collected) and data entry date and time, then the first observation is used as baseline, unless otherwise specified.
  - For PD-L1, non-missing is identified as those with quantifiable test result. After applying the rule above, if there are no records with a 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 which are not reported or not available result (all other records).
  - For Anti-Drug Antibody (ADA), the baseline record of nivolumab immunoglobulin (IMG) evaluation must be less than the date and time of the first nivolumab dose date and time.

- 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.

6.2 Treatment Regimens

Treatment group "as randomized" corresponds to the treatment group assigned by the Interactive Response Technology (IRT) system.

- Arm A: Experimental Arm (monotherapy) nivolumab
- Arm B: Control arm Placebo

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

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

- Enrolled subjects: All subjects who signed the informed consent form and obtained a subject number.
- Randomized subjects: All subjects who were randomized through the IRT.
- Treated subjects: All randomized subjects who received at least one dose of any study treatment.

Unless otherwise specified, the safety analyses will include all treated subjects.

Unless otherwise specified, the efficacy analyses will include all randomized subjects.

- Per protocol subjects: All randomized subjects who had no relevant deviation and those received at least one dose of study drug.
- [redacted]

7 STATISTICAL ANALYSES

7.1 General Methods

Unless otherwise noted, discrete variables will be tabulated by the frequency and proportion of subjects falling into each category, grouped by treatment. Percentages given in these tables will be rounded to the first decimal and, therefore, may not always sum to 100%. Percentages less than 0.1 will be indicated as '< 0.1'. Continuous variables will be summarized by treatment group using the mean, standard deviation, median, minimum, and maximum values.

Time-to-event variables (e.g. 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[12] (using log-log transformation for constructing the confidence intervals[13]).

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 event results will be graded for severity using the NCI Common Terminology Criteria for Adverse Events (CTCAE) and the version of the criteria specified in the protocol 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[14] 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/ROW Submissions)

The select Adverse Events (select AEs) consist of a list of preferred terms grouped by specific category (e.g. 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 (e.g. 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 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 (e.g. 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, 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, 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 first analyzed using International System of Units (SI).

Analyses 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 according to the protocol schedule. Samples will be evaluated for development of Anti-Drug Antibody (ADA) by a validated electrochemiluminescent (ECL) immunoassay.

7.2 Study Conduct

The Relevant Protocol Deviations will be summarized and listed based 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 ClinSIGHT listings.

At Entrance:

- Subjects who, at the time of the initial diagnosis, do not have histologically confirmed esophageal or gastroesophageal junction cancer; OR the stage at initial diagnosis is not Stage II or Stage III.
- Subjects who did not receive concurrent neo-adjuvant CRT prior to complete resection (i.e. they received only neo-adjuvant chemotherapy or only radiation or none prior to surgery).
- Subjects who received treatment directed against the resected GEJ or esophageal cancer (e.g., chemotherapy, targeted agents, radiation, or biologic therapy) that is administered after the complete resection.
- Subjects do not have residual disease (i.e. subjects who have ypN0 AND ypT0)
- Subjects who had complete resection > 18 weeks before randomization.
- Subject with baseline ECOG performance status > 1.
- Subjects who are not disease-free by imaging

On-study:

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

A subject listing will also be produced.

Enrollment by country and site, and enrollment by month will be summarized and listed for all enrolled subjects.

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

7.3 Study Population

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

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. This analysis will be performed on the all enrolled subjects population only.

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. Reason for discontinuation will be derived from subject status CRF page. This analysis will be performed only on the all treated subjects population.

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 study along with the reason for going off study. A by-subject listing for all enrolled subjects will also be provided, showing whether the subject was randomized along with the reason for not being randomized.

7.3.2 Demographics and Other Baseline Disease Characteristics

The following demographic and baseline disease characteristics will be summarized and listed by treatment group as randomized:

- Age (continuous)
- Age categorization (< 65, ≥ 65 and < 75, ≥ 75 and < 85, ≥ 85, ≥ 75, ≥ 65)
- Sex (Male vs. Female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Region (US/Canada, Europe, Asia, Rest of the World [RoW])
- Baseline weight (descriptive statistics)
- Smoking status (current/former, never smoker, unknown)

- Baseline ECOG PS
- Disease at initial diagnosis (EC, GEJ cancer)
- Disease stage at initial diagnosis (Stage I, Stage II, Stage III, Stage IV)
- Disease at study entry (EC, GEJ cancer)
  - EC: body location (Lower third, middle third, upper third)
  - GEJ cancer: Siewert-Stein (Type I vs. Type II vs. Type III)
- Histology (adenocarcinoma, squamous cell carcinoma, other) (stratification factor: squamous vs. adenocarcinoma)
- Histological grade (G1, G2, G3, G4, GX)
- Pathologic TN classification at study entry
  - Tumor (ypT0, ypT1, ypT2, ypT3, ypT4, unknown)
  - Nodes (ypN0, ypN1, ypN2, ypN3, unknown) (stratification factor: ypN0 vs. ≥ ypN1)

- Time from complete resection to randomization (< 4 weeks, 4 - 10 weeks, < 10 - 16 weeks, > 16 weeks).
- Time from randomization to first dose date (≤ 3, 4 - 5, 6 - 7, 8 - 14, 15 - 21, > 21, Not Reported or Not Treated)

- HER-2 status at study entry (negative, positive, unknown)
- Microsatellite instability (MSI-H, MSI-L, MSI-S, unknown)
- EBV status (positive, negative, unknown)

- Baseline LDH (≤ ULN, > ULN; ≤ 2*ULN, > 2*ULN)
- Baseline hemoglobin (< 10g/dL, ≥ 10g/dL)

- Baseline PD-L1+ status based on a 1% cut off (≥ 1% vs. < 1% or indeterminate)
- Baseline PD-L1+ status based on a 5% cut off (≥ 5% vs. < 5% or indeterminate)
- Baseline PD-L1+ status based on a 10% cut off (≥ 10% vs. < 10% or indeterminate)

For the purposes of displaying baseline characteristics, stratification factors will be retrieved from the CRF.

7.3.3 Medical History

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

7.3.4 Prior Therapy Agents

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

Prior anti-cancer therapy:

- Setting of prior systemic therapy regimen received (adjuvant, metastatic disease, neo-adjuvant).
- Prior radiotherapy (yes or no).
- Time from beginning of neoadjuvant CRT to complete resection (< 4weeks, 4-6 weeks, < 6-8, > 8weeks)
- Prior surgery related to cancer (yes or no)
  - type of surgery (esophagectomy, proximal gastrectomy, total gastrectomy, distal gastrectomy, d0, d1, d2, d3, three field lymph node dissection, other)
- Prior systemic therapy classified by therapeutic class and generic name.

Other Prior therapy:

- Prior/current non-study medication classified by anatomic and therapeutic classes.

Agents and medication will be reported using the generic name. A listing by subject will also be provided.

7.3.5 Physical Examinations

Subjects with abnormal baseline physical examination will be listed by subject.

7.3.6 Baseline Physical Measurements

Baseline physical measurements will be listed by subject.

7.3.7 Discrepancies between IWRS and CRF Stratification Factors

Summary tables (cross-tabulations) by treatment group for stratification factor will be provided to show any discrepancies between what was reported through IWRS vs. CRF data (baseline):

1) PD-L1 status (≥1% vs. < 1% or indeterminate) (IWRS vs. clinical database)
2) Pathologic lymph node status (positive ≥ ypN1vs. negative ypN0) (IWRS vs. CRF data)
3) Histology (squamous vs. adenocarcinoma) (IWRS vs. CRF data).

7.4 Extent of Exposure

Listings will include all available exposure data. Analyses will be performed by treatment group "as treated" in all treated subjects, unless otherwise specified.

7.4.1 Administration of Study Therapy

For this study, dosing schedule per protocol is as follows:

- 240mg every 2 weeks for 16 weeks (8 doses) [240mg dosing]
  - Note that a subject has to receive all 8 240mg doses before he/she starts on the 480mg doses. (Even if, due to dose delays, a patient would receive his/her 8th 240mg dose later than Week 16.)
- 480mg every 4 weeks beginning at Week 17 (2 weeks after the 8th dose) [480mg dosing]

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

- Number of doses received
- Cumulative dose
- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%
- Duration of treatment

Duration of study therapy will be summarized (descriptive statistics) by treatment group.
Duration of study therapy will be calculated 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.

A by-subject listing of dosing of study medication (record of study medication, infusion details, and dose changes) will be also provided.

Table 7.4-1: Administration of study therapy: definition of parameters

	Nivolumab
Dosing schedule per protocol	

Table 7.4-1: Administration of study therapy: definition of parameters

	Nivolumab
Dose	240mg dosing: ratio of Total Volume Infused with Total Volume Prepared x 240 in mg 480mg dosing: ratio of Total Volume Infused with Total Volume Prepared x 480 in mg Volume infused/prepared in mL at each dosing date is collected on the CRF.
Cumulative Dose	The sum of all doses (mg) administered to a subject during the treatment period
Relative dose intensity (%)	$100 \times \frac{cumulative\ dose\ (mg)}{(Last\ dose\ date\ of\ nivo - first\ dose\ date\ of\ nivo + X) \times \frac{120\ (mg)}{7\ (days)}}$ X is 14 days if the last dose of nivolumab the patient plans to receive is 240mg; X is 28 days if the last dose of nivolumab the patient plans to receive received is 480mg.
Duration of study therapy	Last dose date - Start dose date + 1

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose Delays

Per protocol, a dose will be considered as actually delayed if the delay is exceeding 3 days. The maximum delay allowed between doses (for both nivolumab and placebo) is 42 days. Dose delay is defined as follows:

- 240mg dosing: duration of preceding cycle in days - 14
- 480mg dosing: duration of preceding cycle in days - 28

Dose delays will be divided into following categories: on-time, 4 - 7 days, 8 - 14 days, 15 - 42 days, 43 - 70 days, > 70 days. Reason for dose delay will be retrieved from CRF dosing pages.

The following parameters will be summarized by treatment arm:

- Number of dose delayed per subject, Length of Delay and Reason for Dose Delay.

7.4.2.2 Infusion Interruptions and Rate Changes

Per protocol, there will be no dose escalations or reductions of study therapy allowed. Subjects may be dosed no less than 12 days (240mg dosing) or 26 days (480mg dosing) from the previous dose.

Each nivolumab or placebo infusion can be interrupted and/or the IV infusion rate can be reduced. This information will be retrieved from CRF dosing pages.

The following parameters will be summarized by 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.

[redacted]

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 (EU/ROW Submissions)
- management of IMAEs (any grade, grade 3-5) by IMAE category (US Submission)

will 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 Cancer Therapy

Number and percentage of subjects receiving subsequent therapies will be summarized. 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
- Radiotherapy for treatment of tumors
- 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

7.5.1 Strong Control of Type I Error

For the analysis purposes of the primary endpoint of DFS and the first secondary endpoint of OS, the overall hierarchical approach[11] will be used to strongly control family-wise error rate of 0.05 across 2 endpoints and repeated DFS and OS analyses at interim and final analyses.

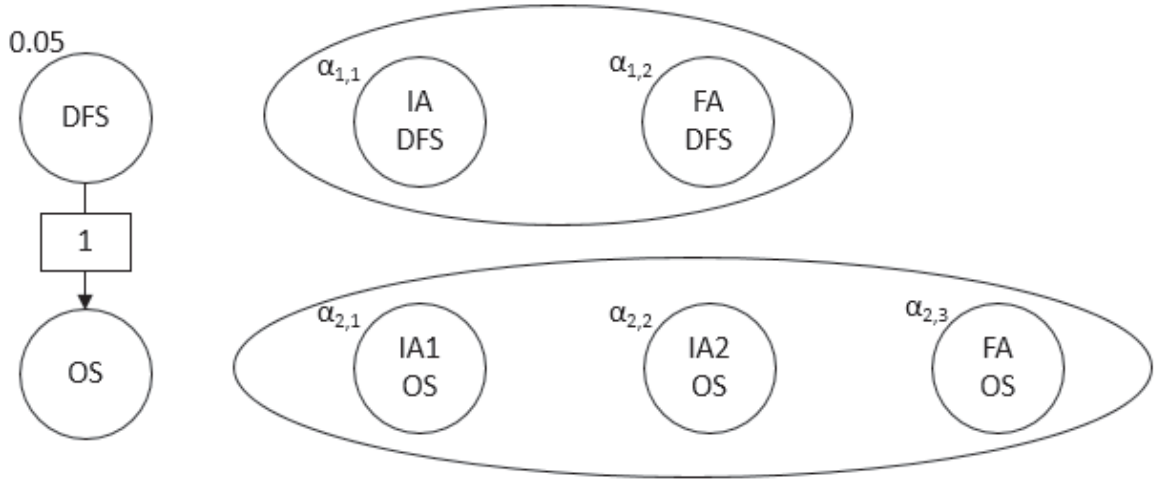
Let i = 1, 2 denote index indicating the primary endpoint DFS and the first secondary endpoints OS respectively. Accordingly, H_1 and H_2 denote two-sided null hypotheses of no effect in primary endpoint DFS and the first secondary endpoint OS. Let t = 1, 2, 3 denote index indicating the analysis timing. Correspondingly, $p_{i,t}$ denote the nominal p-value of testing H_i at the analysis time t, and the corresponding decision boundaries can be expressed as nominal significance levels $\alpha_{i,t}$ such that H_i is rejected if $p_{i,t}$ is smaller than $\alpha_{i,t}$ at time t.

The nominal significance levels for DFS at the interim and final DFS analyses $\alpha_{1,1}$ and $\alpha_{1,2}$ will be computed using EAST software based on a generalization of the Lan-DeMets error spending function approach using an O'Brien-Fleming stopping boundary given overall two-sided alpha = 0.05. The stopping boundary will depend on the actual number of DFS events observed at the time of the DFS interim and final analysis.

Test of OS will be performed following the overall hierarchical testing procedure[11] upon demonstration of superiority in DFS analysis at DFS IA or DFS FA for all randomized subjects. Once OS receives the full alpha from DFS, the Lan-Demets alpha spending function with O'Brien-

Fleming Boundaries will be used to determine the significance level $\alpha_{2,1}$ and $\alpha_{2,2}$ and $\alpha_{2,3}$ for OS at each analysis. The alpha will be adjusted based on actual number of OS events observed.

Figure 4: Testing Strategy for the Primary Endpoint of DFS and the Secondary Endpoint of OS



Specifically, at the interim analysis time of DFS, H_1 is tested with nominal significance level $\alpha_{1,1}$, leading to the following two possibilities of rejecting H_1:

- If $p_{1,1} < \alpha_{1,1}$, H_1 (DFS) will be rejected, and H_2 (OS) will be tested with nominal significance level $\alpha_{2,1}$, leading to the following two possibilities of rejecting H_2 (OS):
    - If $p_{2,1} < \alpha_{2,1}$, H_2 (OS) will be rejected.
    - If $p_{2,1} \geq \alpha_{2,1}$, H_2 (OS) will not be rejected, and the trial continues to the IA2 OS analysis. At the IA2 OS analysis, H_2 (OS) is tested with nominal significance level $\alpha_{2,2}$, leading to the following two possibilities of rejecting H_2 (OS):
        - If $p_{2,2} < \alpha_{2,2}$, H_2 (OS) will be rejected.
        - If $p_{2,2} \geq \alpha_{2,2}$, H_2 (OS) will not be rejected, and the trial continues to the FA OS analysis. At the FA OS analysis, H_2 (OS) is tested with nominal significance level $\alpha_{2,3}$, leading to the following two possibilities of rejecting H_2 (OS):
            - If $p_{2,3} < \alpha_{2,3}$, H_2 (OS) will be rejected.
            - If $p_{2,3} \geq \alpha_{2,3}$, H_2 (OS) will not be rejected.
- If $p_{1,1} \geq \alpha_{1,1}$, H_2 (OS) will not be tested at the interim analysis of DFS and the trial continues to the FA DFS analysis. At the FA DFS analysis, H_1 (DFS) is tested with nominal significance level $\alpha_{1,2}$, leading to the following two possibilities of rejecting H_1 (DFS):

- If $p_{1,2} < \alpha_{1,2}$, H_1 (DFS) will be rejected, and H_2 (OS) will be tested with nominal significance level $\alpha_{2,2,}$ leading to the following two possibilities of rejecting H_2 (OS).
  - If $p_{2,2} < \alpha_{2,2}$, H_2 (OS) will be rejected, and the trial will stop.
  - If $p_{2,2} \geq \alpha_{2,2}$, H_2 (OS) will not be rejected, and the trial continues to the final OS analysis. At the FA OS analysis, H_2 (OS) is tested with nominal significance level $\alpha_{2,3}$, leading to the following two possibilities of rejecting H_2 (OS):
    - If $p_{2,3} < \alpha_{2,3}$, H_2 (OS) will be rejected.
    - If $p_{2,3} \geq \alpha_{2,3}$, H_2 (OS) will not be rejected.
- If $p_{1,2} \geq \alpha_{1,2}$, H_2 (OS) will not be tested, and the trial will be considered as not meeting the primary objective.

If the analysis is performed with the event number exactly the same as those specified in Table 5.1-3, the alpha allocation for DFS and OS is as specified in Table 7.5.1-1. For example if the timing of the interim DFS analysis is performed exactly at 85% of all 440 DFS events (374 DFS events), the boundary in terms of statistical significance for declaring superiority would be $\alpha_{1,1} = 0.03$. The boundary for declaring superiority in terms of statistical significance for the final DFS analysis after 440 DFS events would be 0.042. The significance levels for the final DFS analysis will be adjusted according to the actual alpha spent at the interim analysis and the actual number of DFS events at the interim and final analyses.

The significance levels for the final OS analysis will be adjusted according to the actual alpha spent at interim analysis 1 and interim analysis 2 and the actual number of OS events at OS interim analysis 1, interim analysis 2, and final analysis.

Table 7.5.1-1: Alpha Allocation if the Analysis is Performed with the Event Number Exactly the Same as Specified.

Analysis Timing	DFS		OS	
	Event Number (%of all events)	Alpha	Event Number (%)	Alpha
IA DFS & IA1 OS	374 (85%)	$\alpha_{1,1} = 0.03$	299 (65%)	$\alpha_{2,1} = 0.01$
FA DFS & IA2 OS	440	$\alpha_{1,2} = 0.042$	368 (80%)	$\alpha_{2,2} = 0.022$
FA OS	NA	NA	460	$\alpha_{2,3} = 0.042$

7.5.2 Primary Efficacy Endpoint

DFS will be compared between treatment arms using a two-sided log rank test, stratified by the three randomization stratification factors as recorded in IWRS.

The HR for DFS with its corresponding two-sided 100 x (1 - adjusted α)% CI will be estimated via a stratified Cox model with treatment arm as the only covariate in the model. Adjustment on the CI will be based on the reallocated alpha level.

DFS for each treatment arm will be estimated and plotted using the KM product-limit method. Median survival time will be computed using the KM estimate and a 95% CI for the median will be computed based on a log-log transformation of the survivor function.

Source of DFS Event

The source of DFS event will be summarized by treatment group:

- Recurrence (local recurrence, regional recurrence, distant recurrence)
- Death

Status of Censored Subjects

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

1) Censored on randomization date
   a) no baseline disease assessment
   b) no on-study disease assessment and no death
   c) no on-study disease assessment but death with prior subsequent therapy or second primary cancer
2) Censored on date of last evaluable disease assessment on-study
   a) Received subsequent anti-cancer therapy
   b) Second non-esophageal and non-GEJ primary cancer
   c) Still on treatment
   d) In follow-up
   e) Off study
      i) lost to follow-up
      ii) subject withdrew consent
      iii) other

DFS rates at fixed time points (e.g. 6 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.

7.5.2.1 Sensitivity Analyses

The following four sensitivity analyses are planned for DFS to evaluate the robustness of the primary analysis of DFS.

- *Stratified analysis using stratification factors as obtained from the baseline CRF pages* (instead of IWRS). This analysis will be performed only if the stratification factors at randomization (as per IWRS) and baseline are not concordant for at least 10% of all randomized subjects considering all three stratification factors together.
- *Analysis using a 2-sided, un-stratified log-rank tes*t and an un-stratified Cox proportional hazards model with treatment as the single covariate
- *Analysis for subjects with no relevant deviation and those receiving at least one dose of study drug (per protocol subjects)*. This analysis will be conducted only if there are more than 10% subjects with relevant protocol deviations.

Each of the above analysis will use the same significance level as the corresponding primary analysis. Estimate of the HR, its two sided 100(1-α)% CI and p-value will be presented.

- A multivariate stratified (by factors as specified in Section 2.2) Cox model will be fitted to assess the treatment effect when adjusted for potential prognostic factors. The following potential prognostic factors will be included in the model. A backward selection approach, starting from the full model, will be used as primary method with a significance level set to 0.10 of each of the covariates.
  - Age categorization (< 65, ≥ 65)
  - Gender (Male vs. Female)
  - Baseline ECOG PS (0 vs. 1)
  - Region (Asia vs. RoW [including US/Canada, Europe])
  - Disease at study entry (EC vs. GEJ)
  - Time from beginning of neoadjuvant CRT to complete resection (< 6 weeks vs. ≥ 6 weeks)
  - Time from complete resection to randomization (< 10 weeks vs. ≥ 10 weeks)

If there is more than 10% imbalance between two arms within the subgroup of baseline demographics and disease characteristics, the factor may also be considered in the adjusted multivariate model.

For the multivariate analysis, HR and 95% CI will be provided for treatment variable and all covariates. Descriptive p-values will be provided.

- The key analyses of the primary endpoint DFS are the log-rank test and the Cox proportional hazards regression model, which are most efficient under the proportional hazard assumption. However delayed effect of immunotherapy interventions has been observed in many studies. The proportional hazards assumption will be examined by fitting a Cox model including a time-dependent variable defined by treatment by time interaction.

If the two-sided Wald Chi-square p-value of the treatment by time interaction is less than 0.1 or a visual examination of the K-M plot of the DFS indicates a delayed separation of the curves, the following analyses will be conducted in order to report the treatment effect when taking into consideration of the delayed effect.

− DFS 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.

− The DFS hazard ratios will be estimated in 2 time periods. The periods will be defined by a cutoff point. The cutoff point will be calculated in two ways: 1) using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. The cutoff 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 will also be presented. 2) use the pre-specified cutoff point for delay separation used in the sample size determination, ie.3 months.

Visual interpretation of the curves may lead to additional analyses post database lock with more cut off points.

Sensitivity analyses for DFS to investigate alternative censoring schemes

- *Analysis accounting for assessment on/after subsequent therapy:* DFS will be defined similarly to the primary definition except that events (recurrence or death) and disease assessments that occurred on or after subsequent anti-cancer therapy will be considered (no time point truncation). Although the statistical significance will be determined by the stratified log-rank test using the primary definition of DFS, This analysis will be used to support EMEA filing.
- *DFS accounting for two or more consecutively missing disease assessments prior to DFS event*: This analysis will be performed only if at least 10% of DFS events have two or more consecutively missing prior to 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 DFS event.
- Analysis in which recurrence-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.

Each of the above analysis will use the same significance level as the corresponding primary analysis. Estimate of the HR, its two sided 100(1-α)% CI and p-value will be presented.

7.5.2.2 Consistency of Treatment Effect in Subsets

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

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

- Age category (< 65, ≥ 65 and < 75, ≥ 75; < 65 and ≥ 65)
- Gender (male, female)
- Race (White, Black or African American, Asian, Other)
- Region (Asia, RoW [including US/Canada, Europe])
- ECOG PS (0 vs. 1)
- Disease at study entry (EC, GEJ cancer)
  - EC: (Lower third, middle third, upper third)
  - GEJ cancer: Siewert-Stein (Type I vs. Type II vs. Type II)
- Disease stage at initial diagnosis (Stage I-II, Stage III-IV)
- Histology (stratification factor; squamous vs adenocarcinoma)
- Histological grade (G1/G2, G3/G4, GX)
- Pathologic lymph node status (stratification factor; ypN0 vs. ≥ ypN1)
- Pathologic tumor status (ypT0, ypT1/ypT2, ypT3/ypT4, unknown)

- Time from beginning of neoadjuvant CRT to complete resection (< 6 weeks vs. ≥ 6 weeks)
- Time from complete resection to randomization (< 10 weeks vs. ≥ 10 weeks).

- HER-2 status at study entry (negative, positive, unknown)
- Microsatellite instability (MSI-H, MSI-L, MSS, unknown)
- EBV status (positive, negative, unknown)
- Baseline PD-L1+ status based on a 1% cut off (≥ 1% vs. < 1% or indeterminate)
- Baseline PD-L1+ status based on a 5% cut off (≥ 5% vs. < 5% or indeterminate)
- Baseline PD-L1+ status based on a 10% cut off (≥ 10% vs. < 10% or indeterminate)

For the purposes of the subset analyses, stratification factors will be retrieved from the CRF.

7.5.2.3 Subjects follow-up for DFS

The currentness of follow-up for DFS is defined as the time between last tumor assessment date (regardless of initiation of subsequent therapy or second primary cancer) and cut-off date. Subjects who have a DFS event (regardless of initiation of subsequent therapy or second primary cancer)

and subjects with last tumor assessment date on or after data cut-off will be considered as current for this analysis.

The currentness of follow-up for DFS will be summarized by treatment group, and will be categorized in categories.

7.5.3 Secondary Efficacy Endpoints

7.5.3.1 Primary Analysis of OS

One of the secondary objectives of the study is to compare the overall survival between treatment groups in all randomized subjects.

Overall survival will be compared between the treatment groups at the two interim looks and final analyses, using the log-rank test, stratified by the three randomization stratification factors as recorded in IWRS. An O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses. 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 (e.g. 6 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[15] for variance derivation and on log-log transformation applied on the survivor function[16].

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 Sensitivity Analyses

- The sensitivity analysis specified in section 7.5.2.1 will also be performed for OS when applicable. The sensitivity analyses specific for DFS will not be performed for OS.

7.5.3.3 Consistency of Treatment Effect on OS in Subsets

The influence of baseline demographic and disease characteristics on the treatment effect among all randomized 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 subgroups listed in Section 7.5.2.2 of this SAP.

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

For the purposes of the subset analyses, stratification factors will be retrieved from the CRF.

7.5.3.4 Subjects follow-up for OS

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 subjects.

The currentness of follow-up for survival, defined as the time between last OS contact (i.e., 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 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.

7.5.3.5 Survival Rate Analysis

Survival rate analysis will be carried out only for those time points which are mature enough by the time of the given database-lock - i.e. minimum follow-up must be approximately longer than timepoint to generate the rate.

Survival rates at 6, 12, 18, 24, 36, 48 months and at 5 years will be provided using KM product-limit method. (Note that survival rates at 1, 2, and 3 years are positioned as secondary endpoints.) For each survival rate per treatment arm, two-sided 95% CIs using will be computed using the Greenwood's formula for variance derivation and on log-log transformation applied on the survivor function S(t).

Note that although not positioned as secondary endpoint, DFS rates will also be provided at the same timepoints as survival rates.

[redacted]

[redacted]

7.5.5 Regular Safety Interim Analyses

The Data Monitoring Committee (DMC) will have access to periodic un-blinded interim reports of efficacy and safety to allow a risk/benefit assessment. Details are included in the DMC charter. The chair of the DMC and the sponsor can call an unscheduled review of the safety data.

7.5.6 First Interim Look of the Study

The first interim look of this study will include the DFS interim analysis (DFS IA) and the first OS interim analysis (OS IA1). The first interim look of this study will be triggered when at least 85% of all 440 DFS events (374 DFS events) are observed. Based on the current assumptions, it is projected to occur approximately 12 months after the last patient being randomized.

An independent statistician external to BMS will perform the analysis.

At the time of the first interim look of the study, the DMC may recommend continuing or stopping the trial based on the pre-specified statistical significance for the DFS and OS. If the trial continues beyond the first interim look of the study, ie. DFS is positive but OS is negative, or neither DFS nor OS is positive, the nominal critical point for DFS at the final analysis and OS at the subsequent analyses will be determined using the recalculated information fraction at the time of the interim analysis, as described above. The final DFS and OS hazard ratio and corresponding confidence interval will be reported whereby the confidence interval will be adjusted accordingly (i.e. using the recalculated nominal α level at the final DFS analysis and subsequent OS analysis).

If the trial is stopped for superiority of DFS and OS at the first interim, the analyses results from the first interim stratified log-rank test will be considered the final primary analysis result.

7.5.7 Second Interim Look of the Study

Depending on the outcome of the first interim look of the study, the second interim look of this study will include DFS final analysis (DFS FA) and the second OS interim analysis (OS IA2), or just OS IA2. If both DFS and OS are positive in the first interim look of the study, then there will not be the second interim look of the study. The second interim look of this study will be triggered when at least 440 DFS events have been observed.

An independent statistician external to BMS will perform the analysis.

At the time of the second interim look of the study, the DMC may recommend continuing or stopping the trial based on the pre-specified statistical significance for the DFS and OS. If the trial continues beyond the second interim look of the study when OS is negative in both first and the second interim look of the study, the nominal critical point for the final OS analysis will be determined using the recalculated information fraction at the time of the interim analysis, as described above. The OS hazard ratio and corresponding confidence interval will be reported

whereby the confidence interval will be adjusted accordingly (i.e. using the recalculated nominal α level at the final OS analysis).

If the trial is stopped for superiority of DFS and OS at the second interim, the p-value from the second interim stratified log-rank test will be considered the final primary analysis result.

7.6 Safety

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

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

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 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.

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.
- 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.
- 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.

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.

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 (EU/ROW Submissions)

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.
- 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.
- 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 (US Submission)

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.
- Overall summary of endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- 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.

- Overall summary of endocrine IMAEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- 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
- 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.
- Summaries of time-to onset and time-to resolution of non-endocrine IMAEs where immune modulating medication was initiated presented by Category.
- Summaries of time-to onset and time-to resolution of endocrine IMAEs presented by Category.

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 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

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, i.e., after duplicates have been eliminated and overlapping and contiguous occurrences of the same event (i.e. 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.

7.6.10.1 Hematology

The following will be summarized by treatment group as worst CTC grade on-treatment per subject and 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).

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 worst CTC grade on-treatment per subject and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: ALT, AST, alkaline phosphatase (ALP), total bilirubin and creatinine.

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 worst CTC grade on-treatment per subject and 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), magnesium (high and low), and Glucose Serum (fasting hyperglycemia and hypoglycemia regardless of fasting status).

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 collected on the CRF will be provided in a listing.

7.6.12 Physical Measurements

Physical measurements will be listed by subject.

[redacted]

[redacted]

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, RoW [including US/Canada, Europe])

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

[redacted]

[redacted]

[redacted]

[redacted]

8 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[17]
- For missing and partial adverse event resolution dates, imputation will be performed as follows (these conventions may change):
  - If only the day of the month is missing, the last day of the month will be used to replace the missing day. If the imputed date is after the death date or the last known alive date, then the latest known alive date or death date is considered as the resolution date.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- Missing and partial non-study medication domain dates will be imputed using the derivation algorithm described in 4.1.3 of BMS Non-Study Medication Domain Requirements Specification[18].
- Missing and partial radiotherapy and surgery dates will be imputed using algorithm described in APPENDIX 2.
- 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 alive date 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 alive date.
  - 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 recurrence after start of study therapy, 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. In case of the date of death is present and complete, the imputed recurrence date will be compared to the date of death. The minimum of the imputed recurrence date and date of death will be considered as the date of recurrence.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- For date of recurrence to prior therapies, 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.
- For other partial/missing dates, the following conventions were 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 (e.g. time-to onset, time-to resolution) will be calculated as follows:

Duration = (Last date - first date + 1)

Last known alive date will be defined based on all appropriate dates collected on the CRF.

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

9 DOCUMENT HISTORY

Table 9-1: Document History

Version Number	Author(s)	Description
1.0 Apr 25, 2017	[redacted]	Original Issue based on revised protocol 01 dated Aug 24, 2016

Table 9-1: Document History

2.0 Oct 24, 2019	[redacted]	Based on revised protocol 03 dated June 06, 2019. The following changes are made in the SAP amendment to reflect the protocol amendment of having DFS as the primary endpoint, and OS as the secondary endpoint. [redacted]. • Objectives and endpoints to keep DFS as the sole primary endpoint and change OS to the secondary endpoint • Sample size and power • Schedule of analyses • The method to maintain blinding of OS if not positive at two interim looks of the studyThe method for alpha spending between IA and FA for DFS and OS PFS2 definition and analyses The SAP amendment also incorporates the IO core SAP.
3.0 May 4, 2020	[redacted]	Minor updates from SAP version 2.0 to add sensitivity analyses in Section 7.5.2.1 and other minor edits. Details are summarized below. • • Remove dose reduction in Section 7.6.4 as it is not applicable in this study. • Include in 7.5.2.2 Consistency of Treatment Effect in Subsets, Pathologic tumor status (ypT0, ypT1/ypT2, ypT3/ypT4, unknown) • Remove cumulative plot of baseline PD-L1 expression in section 7.8 as it is not very informtive. • Remove the log-rank test and keep the Cox model for the subgroup analysis by PD-L1 status in section 7.8. • Add "c) no on-study disease assessment but death with prior subsequent therapy or second primary cancer" for censored on randomization date in section 7.5.2
4.0 Sept 30, 2020	[redacted]	Minor format update on hyperlinks, references, and page numbers. Correction of the relative dose intensity formula in Table 7.4-1 Administration of study therapy: definition of parameters.

Table 9-1: Document History

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 9-2 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 (i.e. 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 9-2: 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.

[redacted]

[redacted]

[redacted]

[redacted]