Official Title of Study:

An Open Label, Randomized, Two Arm Phase III Study of Nivolumab in Combination with Ipilimumab versus Extreme Study Regimen (cetuximab + cisplatin/carboplatin + fluorouracil) as First Line Therapy in Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck (SCCHN)

(CheckMate 651: CHECKpoint pathway and nivoluMAb clinical Trial Evaluation 651)

PROTOCOL(S) CA209-651

NCT Number: NCT02741570

Document Date (Date in which document was last revised): January 07, 2020

STATISTICAL ANALYSIS PLAN
FOR CLINICAL STUDY REPORT

An Open Label, Randomized, Two Arm Phase III Study of Nivolumab in Combination with Ipilimumab versus Extreme Study Regimen (cetuximab + cisplatin/carboplatin + fluorouracil) as First Line Therapy in Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck (SCCHN)

PROTOCOL CA209651

VERSION # 2.0

DATE: 07-JAN-2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 BACKGROUND AND RATIONALE

Head and neck cancers are among the most common cancers worldwide, accounting for more than 550,000 cases and around 300,000 deaths each year,[1] and these trends are increasing. In the United States, there were ~45,780 new cases of oral cavity and pharynx cancer and ~8,650 people were expected to die of this disease in 2015.[2] Approximately 90% of all head and neck cancers are squamous cell carcinomas. Most SCCHNs arise from the epithelial lining of the oral cavity, oropharynx, larynx and hypopharynx. The most important risk factors identified in SCCHNs, include tobacco and alcohol use, and, in a subgroup of SCCHNs (particularly oropharynx tumors), human papillomavirus (HPV) as a strong independent prognostic factor, with HPV positive-infected tumors associated with more favorable clinical outcomes.[3]

The treatment approach and prognosis for patients with SCCHN is mostly determined by the tumor stage at presentation. About one-third of patients present with early stage disease, whereas the majority present with advanced disease with lymph node metastasis. With standard of care treatment, the 5-year survival for localized oral cavity and pharynx cancer is 83%, but survival drops to 37.7% for metastatic disease[2] Approximately half of the treated population returns with recurrent or refractory disease, and for these patients, the 1-year survival rate ranges from 5% to 33% with a median overall survival (OS) of 6 to 9 months. Recurrent disease that is not amenable to curative-intent radiation or surgery have the same treatment approach as metastatic disease, and participation in a clinical trial is a recommended treatment option for these patients.[4] A Phase 3 randomized trial in first-line recurrent or metastatic (1L R/M) SCCHN showed that cetuximab plus platinum-fluorouracil chemotherapy (EXTREME regimen) improved survival over platinum-fluorouracil alone as first-line treatment with a median OS of 10.1 months versus 7.4 months (Hazard Ratio [HR] = 0.80, 95% CI 0.64-0.99, p = 0.04). As a result, the EXTREME regimen has been adapted as a standard treatment approach in this population. While the addition of cetuximab to platinum-fluorouracil chemotherapy improved response rates and disease control rates, the duration of response did not differ significantly between treatment regimens. Since the introduction of the EXTREME regimen in 2008,[4] no marked improvement in survival has been achieved, with patients still frequently developing recurrences and distant metastasis. After progression, the limited treatment options available are associated with substantial morbidity and mortality. Thus, new treatment approaches need to be explored as the first-line treatment in this population.[5]

Checkmate 141[6] showed that nivolumab monotherapy, as a second line therapy (2L) in R/M SCCHN, increased the length of time patients lived when compared with standard, single agent chemotherapy in in all study participants (HR = 0.68, 95% CI 0.54 to 0.86; p = 0.01).

Preclinical data indicate that the combination of PD-1 and CTLA-4 receptor blockade may improve antitumor activity. In vitro combinations of nivolumab plus ipilimumab increase IFN-γ production 2- to 7-fold over either agent alone in a mixed lymphocyte reaction. Increased antitumor activity of the combination was also observed in 3 of 5 syngeneic murine cancer models. In a murine melanoma vaccine model, blockade with either CTLA-4 or PD-1 antibodies increased the proportion of

CTLA-4 and PD-1-expressing CD4/CD8 tumor infiltrating T effector cells, and dual blockade increased tumor infiltration of T effector cells and decreased intratumoral T regulatory cells, as compared to either agent alone[7].

The combination of ipilimumab with nivolumab showed increased activity versus nivolumab alone in multiple tumor types, including melanoma, multiple types of lung cancer, gastric cancer, and renal cancer.

Given the similarity in patient profiles of NSCLC and SCCHN, this study will evaluate the efficacy and safety profile of the combination of ipilimumab with nivolumab as first-line therapy in recurrent or metastatic squamous cell carcinoma of the head and neck (SCCHN).

Research Hypothesis:

- Subjects with PD-L1 Combined Positive Score (CPS) ≥ 20 who receive nivolumab in combination with ipilimumab for first-line treatment of recurrent or metastatic SCCHN will have longer OS than comparable subjects who receive the EXTREME regimen
- Subjects who receive nivolumab in combination with ipilimumab for first-line treatment of recurrent or metastatic SCCHN will have longer OS than comparable subjects who receive the EXTREME regimen

Schedule of Analyses:

A formal interim analysis for superiority of OS will be performed after approximately 593 OS events (out of 741 required, 80%) have occurred on all randomized subjects or all randomized subjects have been followed up for at least 12 months, whichever is later. An interim analysis for superiority of OS for subjects with PD-L1 CPS ≥ 20 will be performed simultaneously. An independent statistician external to BMS will perform these analyses.

Final analysis of OS will be conducted when approximately 741 events have occurred on all randomized subjects across the two treatment arms.

2 STUDY DESCRIPTION

2.1 Study Design

Protocol CA209651 is a randomized open-label, Phase 3 trial in subjects ≥ 18 years old with untreated metastatic or recurrent SCCHN that is not amenable to curative therapy, evaluating nivolumab + ipilimumab versus the EXTREME regimen (cetuximab + cisplatin/carboplatin + fluorouracil) as a first-line treatment.

Subjects will undergo screening evaluations to determine eligibility prior to randomization. Approximately 930 subjects will be randomized to the two treatment arms in a 1:1 ratio and stratified by the following factors:

- Tumor cell PD-L1 status (expressing [≥ 1%] vs non-expressing [< 1%] or non-evaluable). Up to 10% of randomized subjects per cohort can be included into the study as "non-evaluable." After this point, subjects with non-evaluable results will not be permitted to be randomized; the site would need to submit an additional tumor tissue for testing with either a result of "expressing" or non-expressing" in order to randomize the subject.

- HPV p-16 status (oropharyngeal HPV p-16 positive vs oropharyngeal HPV p-16 negative or non-oropharyngeal).
- Prior chemotherapy (adjuvant/neoadjuvant/multimodal treatment) status (Yes/No)

The dose and treatment schedules for each arm are as follows:

- Arm A: Nivolumab + Ipilimumab Arm:
  - Nivolumab 3 mg/kg IV every 2 weeks + Ipilimumab 1 mg/kg IV every 6 weeks until progression, unacceptable toxicity, or a maximum of 24 months from first study treatment.
- Arm B: EXTREME regimen Arm:
  - Cetuximab 400 mg/m^2 IV for the initial dose only, then 250 mg/m^2 weekly + cisplatin (100 mg/m^2) or carboplatin (AUC of 5 mg per milliliter per minute) on Day 1 and fluorouracil (1000 mg/m^2 per day for 4 days) every 3 weeks for maximum of 6 cycles followed by maintenance cetuximab at 250 mg/m^2 weekly (or every 2 weeks, per local prescribing information) until disease progression or unacceptable toxicity; the choice of cisplatin or carboplatin is at the discretion of the investigator.

Tumor progression or response will be assessed by investigator using RECIST 1.1 criteria. Treatment with study medication will continue until RECIST 1.1 defined progression, unacceptable toxicity, a maximum of 24 months from the first study treatment, or withdrawal of consent.

Dose reductions will be not be allowed for nivolumab or ipilimumab.

Treatment beyond initial investigator-assessed progression (either clinical or radiographical) is permitted for nivolumab and ipilimumab if the subject has an investigator-assessed clinical benefit and is tolerating study drug.

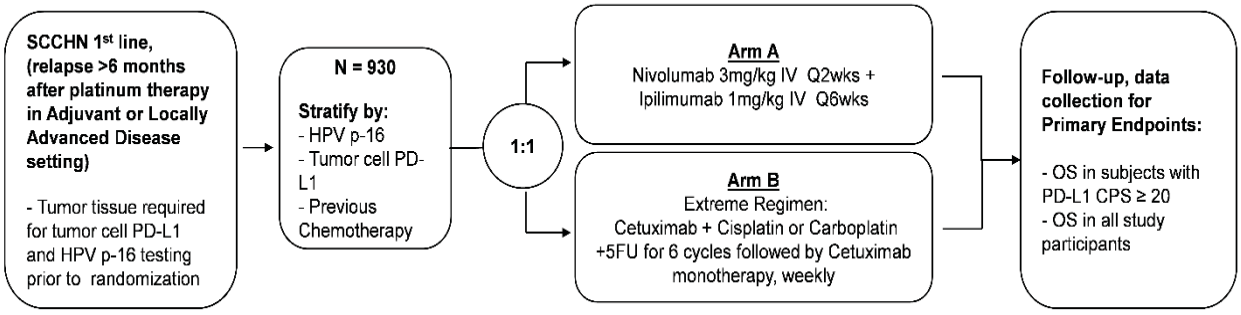
In addition to investigator assessment of response and progression, there will be a blinded independent central review (BICR) of tumor scans and study sites will need to submit tumor scans for central radiology review.

A DMC will be established and meet regularly during the study to ensure that subject safety is carefully monitored and to provide oversight regarding safety and efficacy considerations in protocol CA209651.

The maximum duration of the study from start of randomization to final analysis of OS is projected to be 51 months, assuming 26 months accrual duration. Additional survival follow-up may continue for up to 5 years from the time of this analysis. The study will end once survival follow-up has concluded.

The study design schematic is presented in Figure 2.1-1.

Figure 2.1-1: Study Design Schematic



2.2 Treatment Assignment

This is an open label, randomized, Phase 3 trial in subjects ≥ 18 years old with untreated metastatic or recurrent SCCHN that is not amenable to curative therapy.

Subjects in each arm will be stratified by:

- Tumor cell PD-L1 status (expressing [≥ 1%] vs non-expressing [< 1%] or non-evaluable)
- HPV p-16 status (oropharyngeal HPV p-16 positive vs oropharyngeal HPV p-16 negative or non-oropharyngeal)
- Prior chemotherapy (adjuvant/neoadjuvant/multimodal treatment) status (Yes/No)

Subjects will be randomized in 1:1 and treated with one of open-label treatments Arm A or Arm B.

2.3 Blinding and Unblinding

This is an open label study. However, the sponsor is strictly prohibited from looking at unblinded aggregate efficacy data while the study is still ongoing. The sponsor is forbidden from looking at data by arm while the study is ongoing.

2.4 Protocol Amendments

Table 2.4-1: Protocol Amendments

Document	Date of Issue	Summary of Changes
Revised Protocol 05	19-Dec-2020	• Removed second interim analysis of OS. • Added a Myocarditis adverse event management algorithm.
Revised Protocol 04	20-Jun-2019	• The primary objectives were changed to compare: - OS for subjects who are receiving nivolumab + ipilimumab versus EXTREME regimen in subjects with PD-L1 CPS ≥ 20 (changed from tumor PD-L1 ≥ 1) and - OS for subjects who are receiving nivolumab + ipilimumab versus

Table 2.4-1: Protocol Amendments

Document	Date of Issue	Summary of Changes
		EXTREME regimen in all study subjects (irrespective of PD-L1 expression). • Description of the statistical analyses were changed based on the changes in objectives • Key secondary objective was changed to OS in subject with PD-L1 CPS ≥ 1. • Efficacy evaluation based on biomarker subgroups were added to exploratory objectives. • Added supporting data for change in objectives from recent years and updated literature references • Updated protocol to align with current standards for BMS clinical studies. • Minor formatting and typographical corrections.
Revised Protocol 03	01-May-2018	• The primary objectives were changed from PFS and OS in all randomized subjects to PFS and OS in subjects with PD-L1 expressing tumors. • PFS and OS in all study subjects moved to secondary endpoints. • The hierarchy for the analysis was updated -PFS in all study subjects would be tested hierarchically after PFS in subjects with PD-L1 expressing tumors. OS in all study subjects would be tested hierarchically after OS in subjects with PD-L1 expressing tumors. • The sample size was increased to allow for evaluation of the updated primary end-points. The sample size determination was updated accordingly. • Added brief description of the analysis of TMB data. • Change in medical monitor and study director for the study (Administrative Letter 04 (date of issue:20-Mar-2018)) • Added that nivolumab should be permanently discontinued in case of grade 3 drug-related myocarditis. • Minor formatting and typographical corrections.
Revised Protocol 02	26-Oct-2017	• The sample size was increased. • A maximum duration of nivolumab and ipilimumab treatment of 24 months from the start of treatment was added. • Duration of response has been changed from an exploratory objective to a secondary objective based on the expected improvement in-duration of response as seen in other nivolumab trials. Response to first therapy after disease progression was changed to an exploratory endpoint to limit the secondary objectives only to critical endpoints. • Exploratory biomarker objectives were added in order to evaluate the potential associations between efficacy outcomes (such as objective response rate, PFS, and OS) with select biomarkers, including tumor mutational burden, from tumor tissue and peripheral blood.
Administrative Letter 03	06-Oct-2016	Change in medical monitor.
Administrative	14-Sep-2016	Change to section 5.8, Additional Research section modified.

Table 2.4-1: Protocol Amendments

Document	Date of Issue	Summary of Changes
Letter 02		
Revised Protocol 01	22-Jun-2016	Incorporates Amendment 02
Amendment 02	22-Jun-2016	Week 7 biopsy optional instead of required, Section 5.8 "Additional Research" added, PK and IMG Follow up visit samples no longer required to be collected, Radiographic Imaging wording updated for clarification. Updated Algorithms for Renal, Pulmonary and Skin to match with updated Nivolumab IB v15 (includes Nivo IB 15 erratum update). Other minor edits, clarifications, corrections.
Original Protocol	23-Mar-2016	Not applicable

* Other protocol amendments are site specific and did not require protocol to be updated.

2.5 Data Monitoring Committee

A Data Monitoring Committee (DMC) will be utilized to provide general oversight and safety considerations for this study. The DMC will provide advice to the Sponsor regarding actions the committee deems necessary for the continuing protection of subjects enrolled in this study. The DMC will be charged with assessing such actions in light of an acceptable risk/benefit profile for nivolumab. The DMC will act in an advisory capacity to BMS and will monitor subject safety data for the study.

The DMC will be advisory to the clinical study leadership team. The clinical study leadership will have responsibility for overall conduct of the study including managing the communication of study data. The group will be responsible for promptly reviewing the DMC recommendations, for providing guidance regarding the continuation or termination of the study, and for determining whether amendments to the protocol or changes to the study conduct are required.

Details of the DMC responsibilities and procedures are specified in the DMC charter.

2.6 Blinded Independent Central Review (BICR)

A BICR will be employed for interpretation of radiographic progression events. At the time of investigator-assessed initial radiographic progression per RECIST 1.1 criteria in any given subject, the site must request the independent central review from the third party radiology vendor for confirmation of progression.

Tumor assessments for each subject should be submitted to the radiology vendor as they are performed on an ongoing basis. The blinded, independent radiologists will review all available tumor assessments for that given subject and determine if RECIST 1.1 criteria for progression have been met. The independent assessment of whether or not the given subject met criteria for progression will be provided to the site. Subjects whose disease progression is not confirmed centrally will be

required to continue treatment and tumor assessments according to the protocol-specified schedule. Subsequent tumor assessments must be submitted to the third party radiology vendor for subsequent review and may be discontinued when the investigator and independent radiologists both assess the subject to have met criteria for progression.

The BICR will also review tumor images in all randomized subjects to determine RECIST 1.1 best overall response for the analyses of ORR.

At time of analysis of ORR, tumor assessments will use BICR in all randomized subjects to determine progression and response for the analyses of PFS and ORR. Details of the BICR responsibilities and procedures will be specified in the BICR charter.

3 OBJECTIVES

3.1 Primary Objectives

- To compare the OS of subjects with PD-L1 CPS ≥ 20 who are receiving nivolumab combined with ipilimumab to those receiving EXTREME regimen.
- To compare the OS of all study subjects who are receiving nivolumab combined with ipilimumab to those receiving EXTREME regimen

3.2 Secondary Objectives

- To compare the OS of subjects with PD-L1 CPS ≥ 1 who are receiving nivolumab combined with ipilimumab to those receiving EXTREME regimen
- To evaluate progression-free survival (PFS) based on BICR in all study subjects and those with PD-L1 CPS ≥ 20 who are receiving nivolumab combined with ipilimumab, and those receiving EXTREME regimen.
- To evaluate objective response rate (ORR) based on BICR in all study subjects and those with PD-L1 CPS ≥ 20 who are receiving nivolumab combined with ipilimumab, and those receiving EXTREME regimen.
- To evaluate the duration of response (DOR) based on BICR in all study subjects and those with PD-L1 CPS ≥ 20 who are receiving nivolumab combined with ipilimumab, and those receiving EXTREME regimen.

3.3 Exploratory Objectives

- To evaluate OS of subjects with tumor cell PD-L1 < 1% and PD-L1 CPS < 20 who are receiving nivolumab combined with ipilimumab, and those receiving EXTREME regimen.
- To evaluate the OS of subjects with tumor inflammation score measured as gene expression profile (GEP) ≥ 10 who are receiving nivolumab combined with ipilimumab, and those receiving EXTREME regimen.
- To evaluate the OS of subjects with tumor mutation burden (TMB) ≥ 7 who are receiving nivolumab combined with ipilimumab, and those receiving EXTREME regimen.
- To evaluate clinical outcomes (OS, PFS, ORR, DOR) by select baseline and on-treatment biomarker expression levels in peripheral blood and tumor biopsy specimens.
- To assess safety and tolerability of nivolumab in combination with ipilimumab and of EXTREME regimen, in all study subjects

- To characterize pharmacokinetics and immunogenicity of nivolumab in combination with ipilimumab as first line therapy in subjects with recurrent or metastatic SCCHN
- To evaluate time to symptom deterioration (TTSD) in each arm as assessed using the Functional Assessment of Cancer Therapy - Head and Neck (FACT-H&N) 10-item Symptom Index (FHNSI-10), in subjects with PD-L1 CPS ≥ 20 and also in all study subjects
- To assess the subject's overall health status and health utility using the 3-level version of the EQ-5D (EQ-5D-3L) visual analog scale (VAS) and utility index, respectively, in subjects with PD-L1 CPS ≥ 20 and also in all study subjects
- To assess the subject's cancer-related symptoms and quality of life using components of the FACT-H&N questionnaire, in subjects with PD-L1 CPS ≥ 20 and also in all study subjects.

4 ENDPOINTS

4.1 Primary Endpoints

4.1.1 Overall Survival

- OS in randomized subjects with PD-L1 CPS ≥ 20
- OS in all randomized subjects

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.

OS rate at time T is defined as the probability that a subject is alive at time T following randomization.

Survival follow-up will be conducted every 3 months after subject's off-treatment date.

4.2 Secondary Endpoints

4.2.1 Overall Survival

Overall survival (OS) in randomized subjects with PD-L1 CPS ≥ 1 is defined the same way as for the primary endpoints.

4.2.2 Progression-Free Survival

Two definitions are used for analysis of Progression-Free Survival (PFS). The primary definition accounts for subsequent therapy by censoring at the last evaluable tumor assessment on or prior to the date of subsequent therapy. The secondary definition is irrespective of subsequent therapy and does not account for subsequent therapy.

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

PFS rate at time T is defined as the probability that a subject has not progressed and is alive at time T following randomization.

The first on-study tumor assessment is scheduled to be conducted at 6 weeks (± 1 week) following randomization. Subsequent tumor assessments are scheduled every 6 weeks (± 1 week) up to 12 months, then every 12 weeks until disease progression.

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

The primary definition of PFS (PFS truncated at subsequent therapy) is defined as the time between the date of randomization and the date of first documented tumor progression, based on BICR assessments (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first. Figure 4.2.2.1-1 show the graphic display of primary definition of PFS.

Subjects who die without a reported progression will be considered to have progressed on the date of their death. The following censoring rules will be applied for the primary definition of PFS:

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

Censoring rules for the primary definition of PFS (PFS truncated at subsequent therapy) are presented as follows and in Table 4.2.2.1-1.

Figure 4.2.2.1-1: PFS Primary Definition

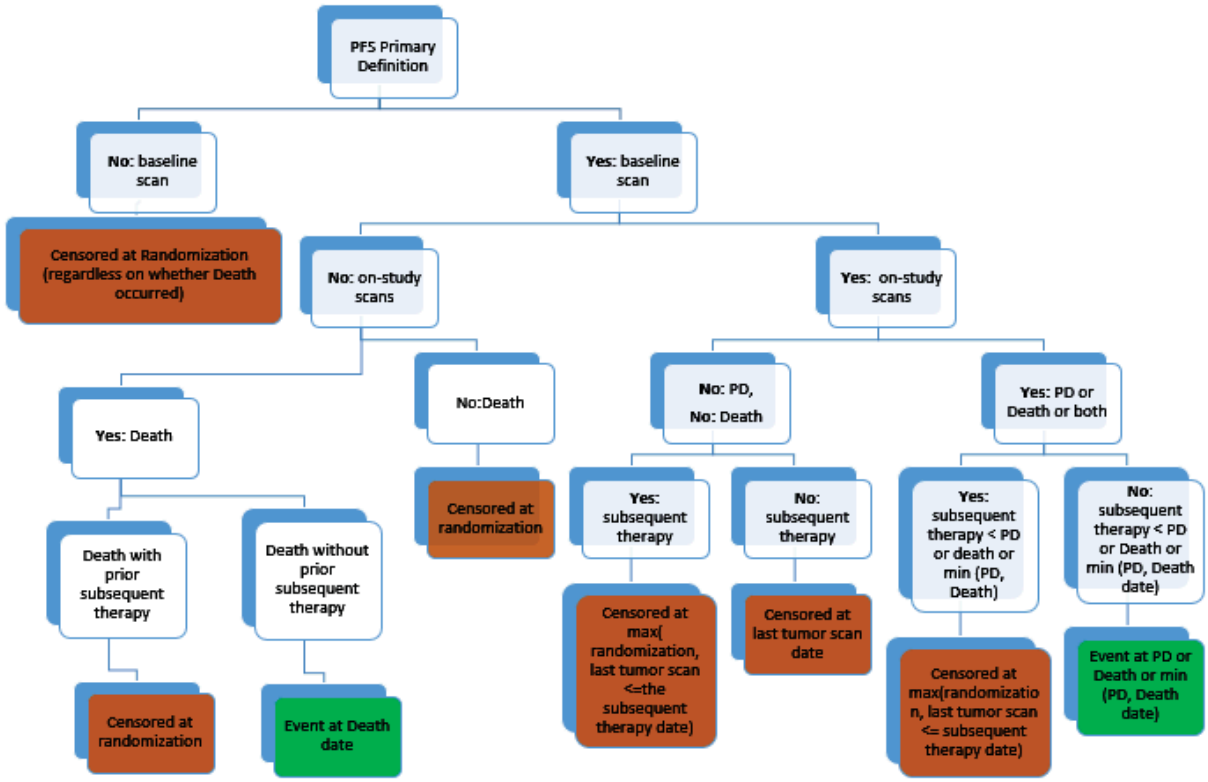


Table 4.2.2.1-1: Censoring Scheme Used in Primary Definition of PFS

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

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

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

The secondary definition of PFS (ITT definition) is defined as the time between the date of randomization and the date of first documented tumor progression, based on BICR assessments (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first. Figure 4.2.2.2-1 show the graphic display of secondary definition of PFS.

Subjects who die without a reported progression will be considered to have progressed on the date of their death. The following censoring rules will be applied for the secondary definition of PFS:

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

Censoring rules for the secondary definition of PFS (ITT definition) are presented as follows and in Table 4.2.2.2-1.

Figure 4.2.2.2-1: PFS Secondary Definition

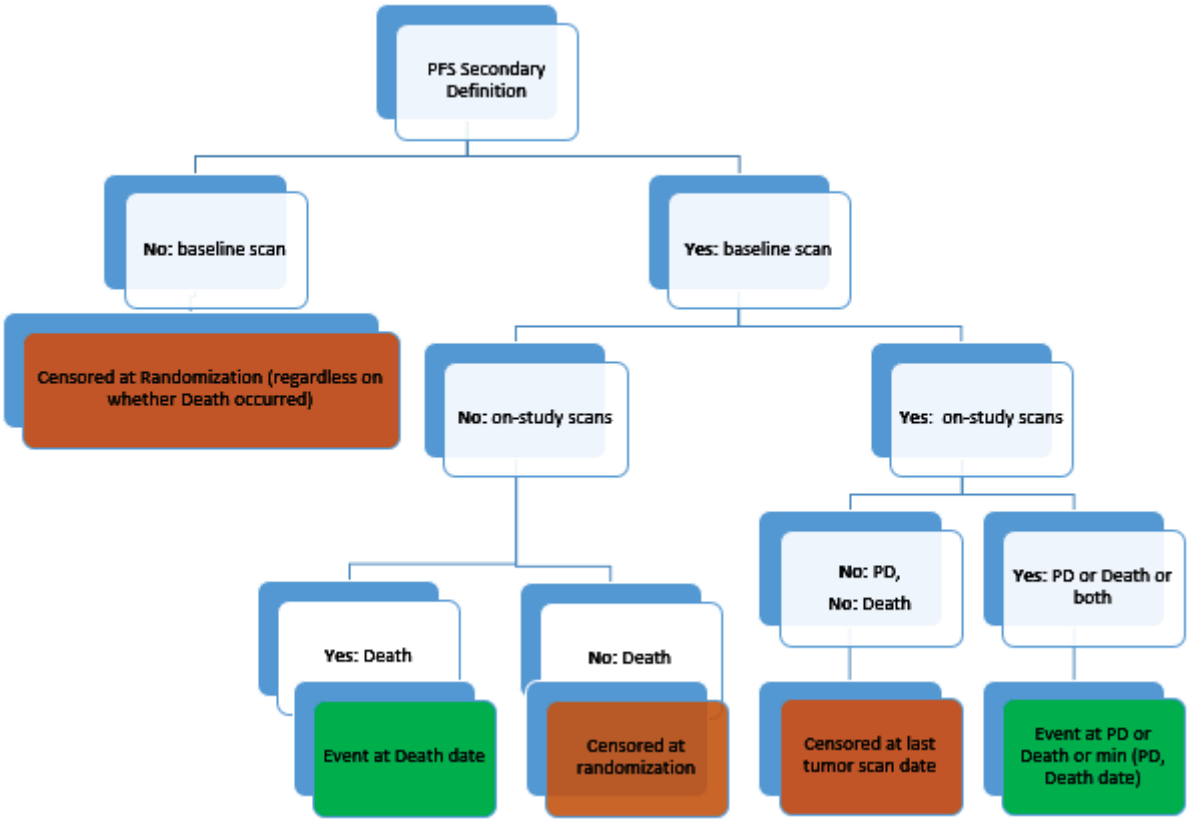


Table 4.2.2.2-1: Censoring Scheme for Secondary Definition of PFS

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

4.2.3 Objective Response Rate

Objective Response Rate (ORR) is defined as the number of randomized subjects who achieve a best response of complete response (CR) or partial response (PR) based on BICR assessments (using RECIST v1.1 criteria) divided by the number of all randomized subjects. Best Overall Response (BOR) is defined as the best response, as determined by the BICR, recorded between the date of randomization and the date of objectively documented progression per RECIST v1.1 criteria or the date of subsequent therapy (including tumor-directed radiotherapy and tumor-directed surgery), whichever occurs first. For subjects without documented progression or subsequent therapy, all available response designations will contribute to the BOR determination. Confirmation of response is required at least 4 weeks after the initial response.

4.2.4 Duration of Response

Duration of Response (DOR) is defined as the time between the date of first documented response (CR or PR) to the date of the first documented tumor progression as determined by the BICR (per RECIST v1.1 criteria), or death due to any cause, whichever occurs first. Subjects who start subsequent therapy without a prior reported progression will be censored at the last evaluable tumor assessments prior to initiation of the subsequent anticancer therapy. Subjects who die without a reported prior progression will be considered to have progressed on the date of their death. Subjects who neither progress nor die, DOR will be censored on the date of their last evaluable tumor assessment. DOR will be evaluated for responders (confirmed CR or PR) only.

As a supportive endpoint to DOR, Time to Response (TTR) is usually provided along with DOR and is defined as the time from randomization to the date of the first confirmed documented response (CR or PR), as assessed by the BICR. TTR will be evaluated for responders (confirmed CR or PR) only.

4.3 Exploratory Endpoints

Efficacy endpoints (OS, PFS, ORR, DOR) in selected subgroups (based on biomarkers such as PDL1 CPS, PDL1 TPS, GEP and TMB) are defined the same way as described in Section 4.1 and 4.2. Other exploratory endpoints are described as follows.

4.3.1 PFS2

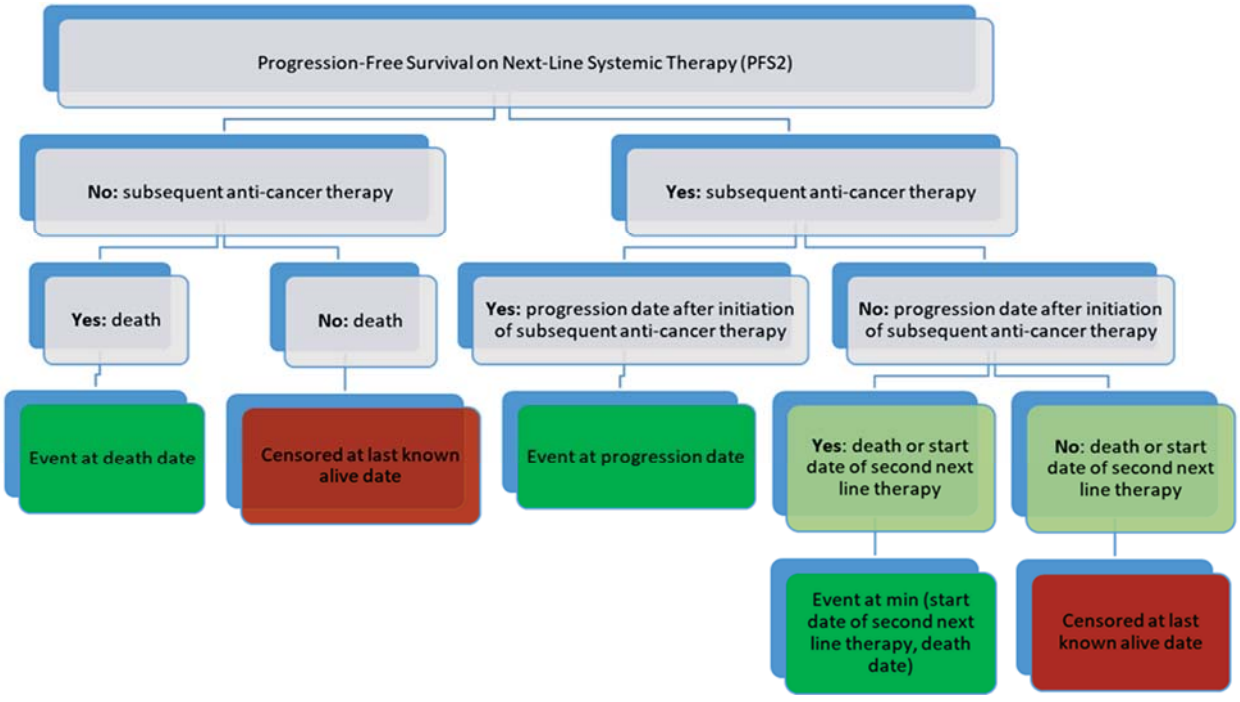
Although not specified in the protocol, PFS on next-line therapy (PFS2) is added as an exploratory endpoint to further evaluate the efficacy of the investigational therapy vs the Extreme Regimen.

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

The following censoring rules will be applied for PFS2:

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

Figure 4.2.2.2: PFS2 Definition



4.3.2 Safety and Tolerability

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 Submission, 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.

4.3.3 Pharmacokinetics

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

4.3.4 Immunogenicity

Serum samples collected for immunogenicity will be assessed for nivolumab and ipilimumab anti-drug-antibodies (ADA) and characterization of neutralizing antibodies. The immunogenicity data will be used to determine immune response rates and relationship of ADA with clinical safety and efficacy.

4.3.5 EuroQoL EQ-5D-3L

Subjects' reports of general health status will be assessed using the EuroQoL Group's EQ-5D-3L[8]. EQ-5D-3L essentially has 2 components: the descriptive system and the visual analogue scale (VAS).

The instrument's descriptive system consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 3 levels, reflecting "no health problems," "moderate health problems," and "extreme health problems." A dimension for which there are no problems is said to be at level 1, while a dimension for which there are extreme problems is said to be at level 3. Thus, the vectors 11111 and 33333 represent the best health state and the worst health state, respectively, described by the EQ-5D-3L. Altogether, the instrument describes $3^5 = 243$ health states. Empirically derived weights can be applied to an individual's responses to the EQ-5D-3L descriptive system to generate an index measuring the value to society of his or her current health. Such preference-weighting systems have been developed for the UK, US, Spain, Germany, and numerous other populations. For this study, EQ-5D-3L utility index values will be computed using a scoring algorithm based on the United Kingdom Time-Trade-Off[9] (UK TTO) value set.

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

A change from baseline of 0.08 for the EQ-5D-3L utility index score and of 7 for the EQ-5D-3L VAS are considered minimally important differences[10] for the EQ-5D-3L.

All questionnaires completed at baseline and on-study will be assigned to a time-point according to the windowing criteria in Table 4.3.5-1 and included in the analysis. In case a subject has two on-study assessments within the same window, the assessment closest to the time-point will be used. And, in the case of two assessments at a similar distance to the time-point, the latest one will be chosen. In the event where the subject has no assessment at all in a specific window, the observation will be treated as missing for that time-point.

Table 4.3.5-1: Time Windows for EQ-5D and FACT-HN (FHNSI-10)

Nominal Time-Point	Time Window
Week 1 (Baseline)	≤ Date of first dose of study therapy
On-Treatment: (until last dose date)	
Week 7 (Day 43)	Day 2 - ≤ Day 64
Week 13 (Day 85)	Day 65 - ≤ Day 106
Week 19 (Day 127)	Day 107 - ≤ Day 148
Every 6 weeks thereafter while on treatment	Nominal Day (+ 21 days / - 20 days, inclusive)
Follow-up Visits 1 and 2	
Follow-up 1	If assessment is post last dose and within 76 days of last dose (if date of discontinuation is within 35 day after last dose) If assessment is post last dose and within 45 days of date of discontinuation (if date of discontinuation is greater than 35 days after last dose)
Follow-up 2	If assessment is post 76 days of last dose and within 173 days of last dose (if date of discontinuation is within 35 day after last dose) If assessment is post 45 days of date of discontinuation and within 137 days of date of discontinuation (if date of discontinuation is greater than 35 days after last dose)
Survival Follow-up Visit	
Survival Follow-up Visit 1	If 173 < assessment date - last dose date ≤ 173 + 90 (if date of discontinuation is within 35 days after last dose) If 137 < assessment date - date of discontinuation ≤ 137 + 90 (if date of discontinuation is greater than 35 days after last dose
Survival Follow-up Visit i	If 173 + (i-1) * 90 < assessment date - last dose date ≤ 173 + i * 90 (if date of discontinuation is within 35 days after last dose) If 137 + (i-1) * 90 < assessment date - date of discontinuation ≤ 137 + i * 90 (if date of discontinuation is greater than 35 days after last dose

Table 4.3.5-1: Time Windows for EQ-5D and FACT-HN (FHNSI-10)

Nominal Time-Point	Time Window
.....	.....

4.3.6 FACT-H&N

The FACT-H&N questionnaire will be used to assess the effects of disease symptoms on functioning and well-being. As a generic cancer-related core, the FACT-H&N includes the 27-item FACT-General (FACT-G) to assess physical well-being (PWB; seven items), social/family well-being (SWB; seven items), emotional well-being (EWB; six items), and functional well-being (FWB; seven items). The FACT-H&N includes a 12-item disease-specific Head and Neck Cancer (HNC) subscale that assesses concerns related to vocalization, communication, breathing, eating, swallowing, appearance, and bother due to adverse events. Ten items included in the PWB, EWB, FWB, and HNC subscales contribute to the scoring of the FHNSI-10, which can be used to provide a targeted assessment of the effects of head and neck cancer symptoms. Each FACT-H&N item is rated on a five-point scale ranging from 0 (not at all) to 4 (very much). Scores for the PWB, FWB, SWB, and EWB subscales can be combined to produce a FACT-G total score, which provides an overall indicant of generic quality of life, while the FACT-G and HNC subscale scores can be combined to produce a total score for the FACT-H&N, which provides a composite measure of general and targeted quality of life. A variant of the total score that is often more sensitive to physical and functional outcomes, the Trial Outcome Index (TOI), can be derived by summing scores for the PWB, FWB, and HNC subscales. All scores are scaled so that higher values indicate better functioning as well as lower symptom burden.

The analysis window is the same as presented in Table 4.3.5-1.

4.3.7 Time to Disease-related Symptom Deterioration

Time to disease-related symptom deterioration (TTSD) is defined as the time from randomization to a clinically meaningful decline from baseline FHNSI-10 score (worsening from baseline ≥ 3 points[11]).

5 SAMPLE SIZE AND POWER

The primary objectives of this study are to compare OS between treatment groups, in all randomized subjects and randomized subjects with PD-L1 CPS ≥ 20. OS will be compared at the 0.025 alpha level for each of the above two populations.

The OS comparison in all randomized subjects will require up to 741 deaths. This number of events ensures that a two-sided, alpha = 0.025 group sequential test using the O'Brien and Fleming spending function and with an interim analysis after approximately 80% of events, will have overall 97% power for a HR of 1.2 for the first 7 months and 0.55 thereafter.

Delayed separation of KM curves is included in the alternative hypotheses for OS because it was observed in CA209141, a randomized phase 3 study of nivolumab versus investigator's choice

therapy in platinum-refractory, recurrent/metastatic SCCHN[12] and in KEYNOTE-048, a randomized phase 3 study of pembrolizumab or pembrolizumab plus chemotherapy versus cetuximab with chemotherapy in untreated recurrent/metastatic SCCHN[13].

The final analysis of OS in all randomized subjects is projected to occur 51 months after the first patient was randomized. An interim analysis in this population will be performed when approximately 593 deaths occur, or all randomized subjects have been followed up for at least 12 months, whichever is later. It is projected for 39 months after the first patient was randomized.

At the same time as the OS analysis is done for all randomized subjects, OS will also be compared between arms in the randomized subjects with PD-L1 CPS ≥ 20 population via a two-sided, alpha = 0.025 group sequential test procedure incorporating the O'Brien and Fleming alpha spending function. The group sequential test would have an interim analysis and the final analysis of OS, at which approximately 298 and 372 deaths, respectively, are expected. With this many events, the test procedure would have an overall 99% power if the HR was 1.06 for the first 7 months and 0.45 thereafter. Table 5-1 summarizes the statistical assumption and power calculation.

Table 5-1: Statistical Assumption and Power Calculation

Primary endpoints	OS for ITT (N = 930)	OS for PD-L1 CPS ≥ 20 (N = 465)
Median OS months (N+I/ EXTREME regimen)	13/10.8	14.9/10.8
Average hazard ratio	0.74	0.61
Alpha level (IA/FA)	0.01/0.022	0.01/0.022
Event Number (IA/FA)	593/741	298/372
Critical hazard ratio (IA/FA)	0.81/0.845	0.74/0.788
Power (IA/FA)	66%/97%	87%/99%
Estimated time of LPLV (IA/FA)	39/51 months	39/51 months

If comparison of OS in the randomized subjects with PD-L1 CPS ≥ 20 is significant, OS in randomized subjects with PD-L1 CPS ≥ 1 will be subsequently tested at the same time using group sequential testing procedure. The interim and final analyses of this endpoint are expected to observe approximately 505, and 632 deaths at time of analyses, and the corresponding power is 74%, and 98% respectively with an average hazard ratio of 0.7.

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

- Baseline period:
  - Baseline evaluations or events will be defined as evaluations or events that occur before the date and time of the first dose of study treatment. Evaluations (laboratory tests, pulse and vital signs) on the same date and time of the first dose of study treatment will be

considered as baseline evaluations. Events (AEs) on the same date and time of the first dose of study treatment will not be considered as pre-treatment events.

- In cases where the time (onset time of event or evaluation time and dosing time) is missing or not collected, the following definitions will apply:
  - Pre-treatment AEs will be defined as AEs with an onset date prior to but not including the day of the first dose of study treatment;
  - Baseline evaluations (laboratory tests, pulse oximetry and vital signs) will be defined as evaluations with a date on or prior to the day of first dose of study treatment.
- If there are multiple valid assessments on or prior to the first dose of study treatment:
  - For laboratory tests, the latest non missing labs value on or before first dose date (and time if collected) will be used as the baseline in the analyses. For 'LIPASE' and 'GLUCOSE', for treated subjects only, the last predose assessment with non-missing toxicity grade will be considered as baseline. If multiple assessments exist with the same collection date (and time if collected) and entry date and time, then the first observation is used as baseline.
  - For Eastern Cooperative Oncology Group (ECOG) performance status (PS), the latest ECOG PS value prior to or on the first dose date (and time if collected) will be used as the baseline in the analyses. If multiple records fall on the last date then the record with the highest value of ECOG PS will be considered as baseline.
  - For PD-L1, among the records prior to or on first dose date (and time if collected), identify first those with quantifiable test result. If there are no records with quantifiable test result, then select those with indeterminant result ("INDETERMINATE"). If there are no records with indeterminant test result, then select those with unavailable result ("NOT EVALUABLE). If there are no records with unavailable test result, then select those with not reported or not available result (all other records). The latest record will be used as the baseline in the analyses. If there is more than one record for the latest date, then choose the one with the greatest specimen ID.
  - For Anti-Drug Antibody (ADA), the record related to the most recent assessment among those records where date (and time if collected) of nivolumab/ipilimumab immunoglobulin (IMG) assessment is less than or equal to the date (and time if collected) of the first nivolumab/ipilimumab dose date.

- 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 100 days 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, pulse oximetry 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 100 days after the last dose of study treatment.

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

6.2 Treatment Regimens

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

- Arm A: Nivolumab 3 mg/kg IV every 2 weeks + Ipilimumab 1 mg/kg IV every 6 weeks until progression, unacceptable toxicity, or a maximum of 24 months from first study treatment.
- Arm B: Cetuximab 400 mg/m^2 IV for the initial dose only, then 250 mg/m^2 weekly + cisplatin (100 mg/m^2) or carboplatin (AUC of 5 mg per milliliter per minute) on Day 1 and fluorouracil (1000 mg/m^2 per day for 4 days) every 3 weeks for maximum of 6 cycles followed by maintenance cetuximab at 250 mg/m^2 weekly (or every 2 weeks, per local prescribing information) until disease progression or unacceptable toxicity; the choice of cisplatin or carboplatin is at the discretion of the investigator.

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

- All Enrolled Subjects: All subjects who signed an informed consent form and were registered into the IVRS.
- All Randomized Subjects: All enrolled subjects who were randomized to either treatment arm.
- All Randomized PD-L1 CPS ≥ 20 Subjects: All randomized subjects with PD-L1 CPS ≥ 20
- All Randomized PD-L1 CPS ≥ 1 Subjects: All randomized subjects with PD-L1 CPS ≥ 1.
- All Randomized Tumor PD-L1 expression < 1% Subjects: All randomized subjects with Tumor PD-L1 expression < 1%
- All Randomized PD-L1 CPS < 20 Subjects: All randomized subjects with PD-L1 CPS <20
- All Randomized GEP ≥ 10 Subjects: All randomized subjects with GEP ≥ 10.
- All Randomized TMB ≥ 7 Subjects: All randomized subjects with TMB ≥ 7.
- All Treated Subjects: All subjects who received at least one dose of nivolumab, ipilimumab, cetuximab, cisplatin, carboplatin or 5-FU.
- Treated Subjects with PD-L1 CPS ≥ 20: All subjects who received at least one dose of nivolumab, ipilimumab, cetuximab, cisplatin, carboplatin or 5-FU and have PD-L1 CPS ≥ 20 at baseline.
- All PK Subjects: All subjects with available serum time-concentration data
- Immunogenicity Evaluable Subjects:

  - Nivolumab ADA Evaluable Subjects: all treated subjects with baseline and at least 1 post-baseline pre-infusion nivolumab immunogenicity assessment.
  - Ipilimumab ADA Evaluable Subjects: all treated subjects with baseline and at least 1 post-baseline pre-infusion ipilimumab immunogenicity assessment.
- All PD-L1 CPS Tested subjects: All subjects, randomized or not, who had a tumor biopsy specimen available for tumor cell expression testing. This includes both randomized and screen failure subjects.

Key analyses of study conduct, demographics, and efficacy will be done on the all randomized population and repeated on the randomized PD-L1 CPS $\geq$ 20 population. Subjects in these analyses will be grouped as assigned at randomization.

Key analyses of exposure and safety will be done on the all treated population and repeated on the treated PD-L1 CPS $\geq$ 20 population. Subjects in these analyses will be grouped by treatment received rather than treatment assigned at randomization. Treatment arm received will be equal to treatment arm as randomized, unless the subject received the non-assigned regimen for all doses.

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[14] (using log-log transformation for constructing the confidence intervals[15]).

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

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

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

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

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

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

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

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

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

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

Analyses that take into account the multiple occurrences of a given adverse event will be conducted (see Section 7.6.9). To prepare these analyses, the CRF data will be processed according to standard BMS algorithms[16] 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 + 101 days)/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 100 days after last dose of study treatment.

7.1.1.1 Select Adverse Events (EU Submission)

The select Adverse Events (select AEs) consist of a list of preferred terms grouped by specific category (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 of select adverse event, time-to onset and time-to resolution are described in APPENDIX 1.

7.1.1.2 Other Events of Special Interest

Other events of special interest (OEOSI) consist of a list of preferred terms grouped by specific category. The list of MedDRA preferred terms used to identify OEOSI is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock by categories will be provided.

7.1.1.3 Immune-Mediated Adverse Events (US Submission)

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

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

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

7.1.2 Laboratory Tests

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

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

Clinical laboratory data will be 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 nivolumab and ipilimumab Anti-Drug Antibody (ADA) by a validated electrochemiluminescent (ECL) immunoassay and characterized for neutralizing antibodies by validated cell-based assays.

7.2 Study Conduct

The following programmable deviations will be considered as relevant protocol deviations.

Eligibility:

1) Subjects whose primary anatomical sites other than: oral cavity, oropharynx, hypopharynx and larynx
2) Subjects who received prior treatment with systemic anti-cancer therapy within 6 months of randomization
3) Subjects who progressed within 6 months of last dose of most recent prior systemic therapy
4) Subjects without measurable disease at baseline
5) Subjects with ECOG PS > 1 at study entry (last measurement on or prior to randomization)
6) Subjects who received prior treatment with an anti-PD-1, anti-PD-L1, anti-PD-L2, anti-CTLA-4 antibody
7) Prior treatment with cetuximab or EGFR inhibitors

On-study:

8) Subjects receiving concurrent antineoplastic therapy (i.e., chemotherapy, hormonal therapy, immunotherapy, surgery for tumor resection*, radiation therapy directed at target lesions) while on study therapy (Surgery has to be for tumor resection. Tumor biopsy, for example, would not count)
9) Subjects whose "as treated" arm different than their as randomized arm (subjects who received the wrong treatment, excluding the never treated)

The Relevant Protocol Deviations will be summarized based on the "all randomized" and "PD-L1 CPS ≥ 20" populations, by treatment group and overall. A listing of all relevant protocol deviations will be provided.

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

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, Other)
- Ethnicity (Hispanic/Latino and Not Hispanic/Latino)
- Region (North America, EU, Asia, Rest of World)
- Country by geographic region
- Baseline ECOG performance status
- Tobacco use (Never, Former, Current, Unknown)
- Alcohol use
- Disease stage at study entry (locally recurrent, locally recurrent and metastatic, metastatic)
- Disease stage at initial diagnosis (stage I, II, III, IVA, IVB, IVC)
- TNM Classification at initial diagnosis
- Time from initial disease diagnosis to randomization (< 1 year, ≥ 1 year)
- Sites of primary tumor per investigator
- Subsites of diseases per investigator
- Sites of diseases (all lesions) per BICR
- Number of disease sites per subject (all lesions) per BICR
- Number of target lesions, non-target lesions and disease sites at baseline as per BICR
- Tumor burden: sum of the diameters of target lesions at baseline per BICR

- Tumor burden: sum of the diameters of target lesions at baseline per investigator
- HPV p-16 status per CRF
- Tumor cell PD-L1 proportion score (TPS) per CRF
- PD-L1 CPS
- CNS metastasis

Summary table (cross-tabulation) by treatment group for stratification factor will be provided to show any discrepancies between what was reported through IRT vs. CRF at baseline. This summary will be performed based on all randomized subjects.

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 systemic cancer therapy will be summarized by treatment group and overall and listed by subject. Prior radiotherapy and prior surgery related to cancer will be listed by subject.

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.

7.4 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

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

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

Table 7.4.1-1: Study Therapy Parameter Definitions- Nivolumab and Ipilimumab

	Nivolumab	Ipilimumab
Dosing schedule per protocol	3 mg/kg every 2 weeks	1 mg/kg every 6 weeks

Table 7.4.1-1: Study Therapy Parameter Definitions- Nivolumab and Ipilimumab

	Nivolumab	Ipilimumab
Dose	*Dose (mg/kg)* is defined as Total Dose administered (mg)/Most recent weight (kg). Dose administered in mg at each dosing date and weight are collected on the CRF.	*Dose (mg/kg)* is defined as Total Dose administered (mg)/Most recent weight (kg). Dose administered in mg at each dosing date and weight are collected on the CRF.
Cumulative Dose	*Cum dose (mg/kg) is* sum of the doses (mg/kg) administered to a subject.	*Cum dose (mg/kg) is* sum of the doses (mg/kg) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/kg)/[(Last Nivolumab dose date - Nivolumab start dose date + 14) x 3/14] x 100	Cum dose (mg/kg)/[(Last Nivolumab dose date - Nivolumab start dose date + 42) x 1/42] x 100
Duration of treatment (overall)	Last Nivolumab dose date - Nivolumab start dose date +1	Last Ipilimumab dose date - Ipilimumab start dose date +1

Table 7.4.1-2: Study Therapy Parameter Definitions-Cetuximab/5-FU

	Cetuximab	5-FU
Dosing schedule per protocol	400 mg/ m² for the 1st dose and 250 mg/ m² weekly or biweekly per local prescribing information	1000mg/ m² (from Day 1 to Day4) every 3 weeks
Dose	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.
Cumulative Dose	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.
Relative dose intensity (%)	[Cum dose (mg/m²) - initial dose of cetuximab(mg/m²)]/[(Last Cetuximab dose date - Cetuximab Start dose date)x250/7] x 100	Cum dose (mg/m²)/[(First dose date of 5-FU in Last Cycle - 5-FU Start dose date + 21) x 4000/21] x 100
Duration of treatment (overall)	Last dose date - Start dose date +1	Last dose date - Start dose date +1

Table 7.4.1-3: Study Therapy Parameter Definitions- Cisplatin/Carboplatin

	Cisplatin	Carboplatin
Dosing schedule per protocol	100mg/ m² on day 1 every 3 weeks	AUC 5 on day 1 of every 3 week

Table 7.4.1-3: Study Therapy Parameter Definitions- Cisplatin/Carboplatin

	Cisplatin	Carboplatin
Dose	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.	*Dose (AUC)* is defined as Total Dose administered (mg)/(creatinine clearance +25). Dose administered in mg at each dosing date is collected on the CRF and creatinine clearance derived from the CRF data.
Cumulative Dose	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.	*Cum dose (AUC) is* sum of the doses (AUC) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/m²)/[(Last Cisplatin dose date - Cisplatin Start dose date + 21) x 100/21] x 100	Cum dose (AUC)/[(Last dose date of Carbo - Start dose date of Carbo + 21) x 5/21] x 100
Duration of treatment (overall)	Last dose date - Start dose date +1	Last dose date - Start dose date +1

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose Delays

Each of study drug infusions may be delayed. A dose will be considered as actually delayed if the delay is exceeding 3 days (i.e. greater than or equal to 4 days from scheduled dosing date) for nivolumab and ipilimumab. All study drugs must be delayed until treatment can resume. Reason for dose delay will be retrieved from CRF dosing pages.

The following parameters will be summarized by treatment group:

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

7.4.2.2 Infusion Interruptions and Rate Changes

Each of study drug infusions 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.

7.4.2.3 Dose Escalations

Dose escalations (within subject) are not permitted for either nivolumab or ipilimumab.

7.4.2.4 Dose Reductions

Dose reductions (within subject) are not permitted for either nivolumab or ipilimumab. Therefore, no analyses on dose reductions will be presented for these two agents.

Up to two dose level reductions are allowed for cetuximab, cisplatin, carboplatin and 5-FU. The dose levels for each of these drugs are defined in the protocol as follows:

Table 7.4.2.4-1: Dose Levels for Cetuximab, Cisplatin, Carboplatin and 5-FU

	Cetuximab (mg/m2)	**Cisplatin (mg/m2)**	**Carboplatin (AUC)**	**5-FU (mg/m2/day)**
Starting Dose	250*	100	AUC 5.0	1000
Dose Level -1	200	75	AUC 4.0	750
Dose Level -2	150	56	AUC 3.0	560
Dose Level -3	Stop drug	Stop drug	Stop drug	Stop drug

*The dose level of cetuximab is 400 mg/m2 for the first dose and 250 mg/m2 afterwards.

For any cycle, it will be defined as a dose reduction if the observed dose level (based on calculated administered dose) is below the dose level of the previously administered dose. Dose ranges for dose levels of platinum doublet chemotherapy are defined in Table 7.4.2.4-2.

Table 7.4.2.4-2: Calculated Dose Ranges and Related Dose Levels

Dose Level	Dose Range			
	Cetuximab (mg/m2)	**Cisplatin** (mg/m2)	**Carboplatin** (AUC)	**5-FU** (mg/m2/day)
Level 0	≥225	≥87.5	≥4.5	≥875
Level -1	<225 and ≥ 175	<87.5 and ≥ 65.5	<4.5 and ≥3.5	<875 and ≥ 655
Level -2	<175	<65.5	<3.5	<655

The reason for dose reduction as reported by the investigator will be tabulated for all instances of dose reduction based on the Dose Change CRF page. A category 'Unknown' will be defined for all reductions with no reason reported by the investigator.

The following will be summarized for chemotherapeutic agent arm only:

- Number and percentage of subjects with at least one dose reduction and reason of the dose reduction, number and percentage of subjects with a dose reduction to dose level -1, number and percentage of subjects with a dose reduction to dose level -2.

7.4.2.5 Dose Omissions

Dose omissions are not permitted.

7.4.2.6 Partial Discontinuation of Ipilimumab in the Arm A

Subjects treated with nivolumab and ipilimumab may discontinue ipilimumab and continue to receive nivolumab (ie, partial discontinuation).

The following will be summarized for subjects receiving the immunotherapy in the arm A

- Number and percentage of subjects who had partial discontinuation of ipilimumab.
- Reason for partial discontinuation.

Reason for partial discontinuation will be retrieved from dosing CRF pages.

7.4.3 Concomitant Medications

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

The following summary table will be provided:

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

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

The following summary table will be provided:

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

A by-subject listing will accompany the table.

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 Submission)
- 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 is 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 cancer therapies will be summarized for all randomized subjects. Categories include:

- Subsequent systemic therapy
- Subsequent surgery for treatment of tumors
- Subsequent radiotherapy for treatment of tumors

A by-subject listing of subsequent cancer therapy will also be produced for all randomized subjects.

7.5 Efficacy

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

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

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

- Tumor cell PD-L1 status (expressing [≥ 1%] vs non-expressing [< 1%] or non-evaluable).
- HPV p-16 status (oropharyngeal HPV p-16 positive vs oropharyngeal HPV p-16 negative or non-oropharyngeal).
- Prior chemotherapy (adjuvant/neoadjuvant/multimodal treatment) status (Yes/No).

For assessing the secondary objectives of this study except to compare the OS of subjects with PD-L1 CPS ≥ 1 between treatment groups, no testing procedure will be used.

Confidence intervals (CI) for primary endpoint analyses will be based on nominal significance level adjusted for primary endpoints and interim analyses to preserve overall type one error rate.

Alpha (α) for the CI will be the same as nominal significance level for hypothesis testing. CIs for other endpoints will be at the two-sided 95% level. All p-values reported will be two-sided. P-values will be rounded to the fourth decimal place. Point estimates and confidence bounds for efficacy variables will be rounded to the second decimal place.

A by-subject listing of efficacy results will be presented including treatment group, treatment duration, BICR progression date, overall survival, death date, etc.

7.5.1 Analysis of Objective Response Rate

One of the objectives of the study is to estimate the ORR per BICR in the treatment groups among all randomized subjects.

The number and percentage of subjects in each category of BOR per BICR (complete response [CR], partial response [PR], stable disease [SD], progressive disease [PD], or unable to determine [UTD]) will be presented, by treatment group. Estimates of response rate, along with its exact two-sided 95% CI by Clopper and Pearson[17] will be presented, by treatment group.

Similar analyses will be repeated based on the investigator's assessment of ORR. A cross tabulation of BICR best response versus the investigator best response will be presented, by treatment group and by response categories. Concordance Rate of Responders will be computed as the frequency with which investigator and BICR agree on classification of a subject as responder vs. non responder/UTD as a proportion of the total number of randomized subjects assessed by both the investigator and BICR.

The following subject-level graphics will also be provided:

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

A by-subject listing of best overall response will be presented including treatment group, best overall response per BICR and dates of CR/PR/progression.

A by-subject listing of per time point tumor response per BICR will be presented.

7.5.2 Time to Tumor Response and Duration of Response

Duration of response (DOR) and time to response (TTR) will also be evaluated for subjects who achieved confirmed PR or CR. The DOR for each treatment group will be estimated using the Kaplan-Meier (KM) product limit method and will be displayed graphically. A table will be produced presenting number of events, number of subjects involved, medians, and 95% CIs for the medians. Median values of DOR, along with two-sided 95% CI in each treatment group will be computed based on a log-log transformation method.

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

- Ongoing follow-up (current [last scan within adequate window vs cutoff date], not current)

- Off-study (lost to follow-up, withdraw consent, never treated)
- Received subsequent anticancer therapy.

TTR, which does not involve censoring, will be summarized by treatment group in all responders using descriptive statistics.

A by-subject listing will be presented including treatment group, best response, time to response, duration of response, whether the subject was censored for duration of response, and, if so, the reason.

7.5.3 Analysis of DCR

Similar to ORR analysis, BICR-determined DCR in arm A and B will be estimated and its corresponding 95% exact two-sided CIs will be calculated using the Clopper Pearson method. This analysis will also be performed for DCR as assessed by the investigators.

7.5.4 Analysis of Progression-Free Survival

The PFS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median PFS in each treatment group will be computed via the log-log transformation method. PFS rates at fixed time points (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[18] formula for variance derivation and on log-log transformation applied on the survivor function[19].

The estimate of the PFS hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. A two-sided 95% CI for the hazard ratio will also be presented. Further, the estimate of the PFS hazard ratio based on weighted logrank test may be provided.

Analyses of PFS will be conducted based on the primary definition and secondary definition of PFS.

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

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

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

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

7.5.5 Supportive Analyses of Progression-Free Survival

The following sensitivity analyses will be conducted using both the primary and the secondary definition of PFS in all randomized subjects:

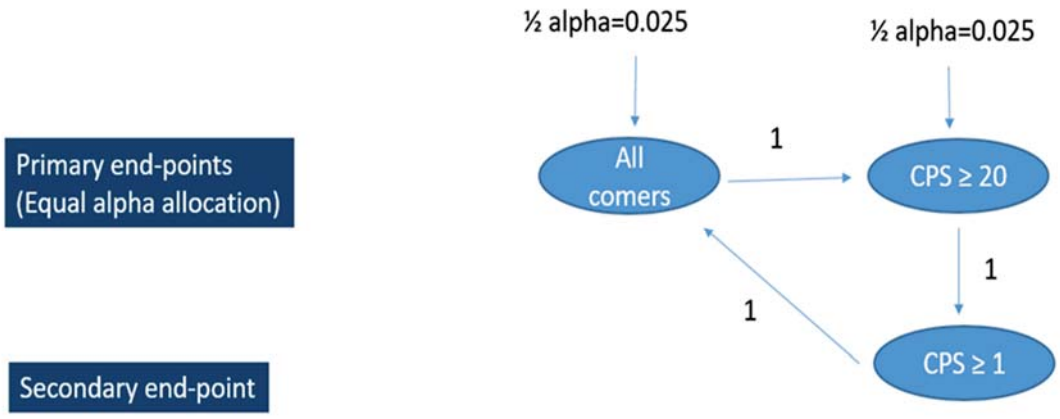
- A cross tabulation of PFS assessment by BICR versus PFS assessment by investigator will be presented, by treatment group. Concordance Rate of event will be computed as the frequency with which investigator and BICR agree on classification of a subject as event vs censored as a proportion of the total number of randomized subjects assessed by both the investigator and BICR.
- A by-subject listing of PFS assessment per BICR and investigator will be presented.

7.5.6 Analysis of Overall Survival

The primary objectives of the study are to compare the overall survival between treatment groups in randomized subjects with PD-L1 CPS ≥ 20 and in all randomized subjects.

Overall survival will be compared between the treatment groups at the interim and final analyses, using a stratified log-rank test. The stratification factors will be tumor cell PD-L1 status, HPV p-16 status and prior chemotherapy status. Graphic procedure[20] (Figure 7.5.6-1:) and an O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses. Details of this sequentially rejective procedure will be provided in Appendix 4.

Figure 7.5.6-1: Testing Hierarchy for the Endpoints



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 for variance derivation and on log-log transformation applied on the survivor function.

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.

7.5.7 Supportive Analyses of Overall Survival

The following OS sensitivity analyses will be conducted in all randomized subjects:

1) Delayed effect of immunotherapy interventions may cause a late separation in the OS KM curves and non-proportional hazards. OS will be compared between treatment groups via two-sided stratified weighted log-rank test among subjects. The two-sided stratified weighted log-rank p-value will be reported using G (rho = 0, gamma = 1) weights, in the terminology of Fleming and Harrington[13].

   The Fleming Harrington test can be unstable, so it is possible, though uncommon, that the p-value for this trial will not be estimable.

   Further, Max-combo testing and estimating procedure might be applied.

2) A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization, which, by definition, will be balanced across treatment groups, will still be included in the model as stratification factors. However, all additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include, but not limited to:
   a. ECOG PS (0, ≥1)
   b. SEX (Male, Female)
   c. AGE

   The level of the covariate normally associated with the worst prognosis will be coded as the reference level. The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 100*(1- α)% CIs (adjusted for interim).

3) OS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided 100*(1- α)% CI (adjusted for interim). This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the randomized subjects.

4) OS using an un-stratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided 100*(1- α)% CIs (adjusted for interim).

5) OS for subjects with no relevant protocol deviations. This analysis will be conducted only if there are more than 10% subjects with relevant protocol deviations. The hazard ratio

associated with treatment will be presented along with the associated two-sided 100*(1-α)% CIs (adjusted for interim).

6) OS will be compared between treatment groups using a two-sided stratified log-rank test in the All Treated Subjects population, using arm as treated. This analysis will be performed only if the proportion of randomized but never treated subjects exceeds 10%.
7) To examine the assumption of proportional hazards in the Cox regression model, in addition to treatment, a time-dependent variable defined by treatment by time interaction will be added into the model. A two-sided Wald Chi-square p-value of less than 0.1 may indicate a potential nonconstant treatment effect. In that case, additional exploratory analyses may be performed.

7.5.8 Subset Analyses of Overall Survival

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

- Age categorization (<65 vs.>=65 vs. ≥65-<75 vs >75)
- Sex (Male vs. Female)
- Race (White, Black, Asian, Others)
- Region (North America vs. EU vs. Asia vs. ROW)
- ECOG performance status(0, ≥1)
- Disease stage at initial diagnosis (Stage I, Stage II, Stage III, Stage IVA, Stage IVB, Stage IVC)
- Disease stage at study entry (locally recurrent, locally recurrent and metastatic, metastatic)
- Time from initial disease diagnosis to randomization (< 1, 1-<2, 2-<3,3-<4, 4<-5, >=5 year)
- Site of primary tumor (Oral cavity, Oropharynx, Hypopharynx, Larynx)
- Tobacco use (Never, Former, Current, Unknown)
- Alcohol use
- Prior radiotherapy (yes/no)
- Prior surgery (yes/no)
- Prior systemic cancer therapy (yes/no)
- Prior platinum-based chemotherapy (yes/no)
- HPV status per IRT

A forest plot of the OS hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above.

An analysis will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.9 Current Status of PFS and OS Follow-up

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

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

In addition, time to treatment discontinuation will be summarized and presented by treatment group using a Kaplan-Meier curve whereby the last dose date will be the event date for those subjects who are off study therapy. Median duration of study therapy and associated 95% CI will be provided. Subjects who are still on study therapy will be censored on their last dose date.

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

7.5.10 Interim Analyses of Overall Survival

An independent statistician external to BMS will perform the analyses. In addition to the formal planned interim analysis for OS, 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.

A formal interim analysis for superiority of OS will be performed after approximately 593 OS events (80% of 741 required) have occurred on all randomized subjects, or all randomized subjects have been followed up for at least 12 months, whichever is later. Both primary endpoints will be tested at interim using the group sequential test procedure with initial overall alpha = 0.025 for each. The overall alpha will be updated following graphical procedure[20] (Figure 7.5.6-1:). The stopping boundaries for the interim and final analyses will be determined by the O'Brien and Fleming alpha spending function, based on the actual number of events observed.

If interim comparison of OS in either all randomized subjects or randomized subjects with PD-L1 CPS $\geq$ 20 is significant, the DMC will inform the sponsor, as described in the DMC charter. If comparison of OS in the randomized subjects with PD-L1 CPS $\geq$ 20 is significant, OS in randomized subjects with PD-L1 CPS $\geq$ 1 will be subsequently tested using group sequential testing procedure. The overall alpha for this secondary endpoint will be determined following the graphical procedure which will be illustrated in Appendix 4.

7.5.11 PFS2

PFS2 will be analyzed similarly to PFS:

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

- The estimate of the PFS2 hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. A two-sided 95% CI for the hazard ratio will also be presented
- PFS2 will be graphically displayed along with the median and 95% CI.

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

7.6 Safety

All figures and summary tables for safety will be produced on the "all treated PD-L1 CPS ≥ 20," and "all treated" populations. Safety listings will be done on the all treated populations but will have variables for PD-L1 CPS added to them.

Analyses in this section will be tabulated by treatment group as treated, unless otherwise specified.

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/reduction will be summarized by treatment group:

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

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

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

7.6.5 Adverse Events

Adverse events will be summarized by treatment group.

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

- Overall summary of any AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT.
- 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 Submission)

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

- Summary of subjects who were re-challenged with nivolumab by IMAE category, with extended follow-up
- Summary of subjects who were re-challenged with nivolumab or ipilimumab by IMAE category, with extended follow-up

For these, re-challenge is considered to have occurred when last nivolumab and/or ipilimumab 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.

In addition, the rate (exposure adjusted) and its 95% CI evaluated for different time intervals will be displayed graphically for each treatment group. This analysis will be limited to the rate of all AEs and all drug-related AEs. The analyses will be conducted using the 30-day safety window.

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. Laboratory tests (in addition to the tests specified below) with CTC criteria collected in the specific studies may also be included in the summaries.

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 and Pulse Oximetry

Vital signs and pulse oximetry (i.e. % oxygen saturation) collected on the CRF will be provided in separate listings.

7.6.12 Physical Measurements

Physical measurements will be listed by subject.

7.6.13 Non-Protocol Medical Procedures

Non-protocol medical procedures will be listed by subject.

7.6.14 Immunogenicity Analysis

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

Incidence of ADA

Number (%) of subjects will be reported for the following parameters based on Evaluable Subjects.

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

A listing of all ADA assessments will be provided.

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

The effect of ADA on safety and efficacy

- Summary of Select Adverse Events of Hypersensitivity/Infusion Reaction by ADA Status (Positive, Negative)
- Swimmer plot of occurrence of ADA and NAb Occurrence in Relation to PFS, BOR per BICR and OS

7.6.15 Pregnancy

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

7.6.16 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 (North America vs. EU vs. Asia vs. ROW)
- Prior chemotherapy status (Yes/No)

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

7.7 Pharmacokinetics

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

7.8 Biomarkers

Analyses for Biomarkers such as PD-L1 CPS, PD-L1 TPS, Gene Expression Profile (GEP) and Tumor Mutation Burden (TMB) are described below.

7.8.1 Distribution of PD-L1 CPS

Descriptive statistics of PD-L1 CPS:

- Listing of all PD-L1 IHC data, all PD-L1 tested subjects.
- Summary of tumor specimen acquisition and characteristics, all randomized subjects.
- Summary statistics of PD-L1 CPS in all randomized subjects with quantifiable PD-L1 CPS.
- Summary of BOR and ORR by baseline PD-L1 CPS in all randomized subjects.
- Cumulative distribution plot of baseline PD-L1 CPS versus population percentile in all randomized subjects with quantifiable PD-L1 CPS.
- Box plots of PD-L1 CPS versus Response Status (BICR assessment) in all randomized subjects with quantifiable PD-L1 CPS.
- Waterfall plot of Individual PD-L1 CPS in all randomized subjects with quantifiable PD-L1 CPS.
- Histogram of the distribution of PD-L1 CPS.

7.8.2 Association Between PD-L1 CPS and Efficacy

Analyses of association between PD-L1 CPS at baseline and efficacy measures will be performed for scheduled formal efficacy analyses in the study. The analysis population will be among all randomized subjects unless otherwise specified.

For each PD-L1 CPS subgroup (further specified in the DPP) at baseline:

- OS/PFS curves for each treatment group will be estimated using the Kaplan-Meier product limit method. Two-sided, 95% confidence intervals for median OS/PFS will be constructed based on a log-log transformed CI for the survivor function S(t).
- Forest plot of Hazard Ratios with 95% CIs
- Frequency and percentage of BOR per BICR will be summarized for each treatment group.
- ORR will be computed by treatment group along with exact 95% CIs using the Clopper Pearson method.

The following analysis will be performed to evaluate the association between PD-L1 CPS level and BICR-determined PFS (per primary definition) or OS.

- An exploratory Cox proportional hazards model will be fitted for PFS per BICR or OS with PD-L1, treatment arm and PD-L1 treatment arm interaction, among All PD-L1 Evaluable Subjects. An appropriate transformation of PD-L1 CPS may be considered depending on an assessment of fit of the model.
- A plot of estimated log(hazard ratio) with 95% confidence band vs PD-L1 CPS > 0 (X-axis)

The following analysis will be performed to evaluate the evaluation of association between PD-L1 CPS level and ORR (per BICR) among PD-L1 CPS evaluable subjects for each treatment arm.

- A logistic regression model will be fitted for ORR with PD-L1 CPS among all PD-L1 evaluable subjects. An appropriate transformation of PD-L1 CPS may be considered depending on an assessment of fit of the model.
- A plot of estimated response probability with 95% confidence band vs PD-L1 CPS (X-axis)
- Box plot of PD-L1 CPS versus Response Status

Receiver Operating Characteristics (ROC) analysis with ORR (per BICR) will be performed to help assess in-study predictive accuracy of the logistic regression model and whether there is a clinically meaningful threshold of PD-L1 CPS. A plot of the ROC curve and a plot of estimated Youden's index vs PD-L1 CPS will be provided for all PD-L1 CPS evaluable subjects.

7.8.3 Analyses of TPS, GEP and TMB

The following analyses will be performed for TPS, GEP and TMB:

- Summary statistics of the biomarker in all randomized subjects with quantifiable biomarker.
- Summary of BOR and ORR by baseline biomarker in all randomized subjects.
- For each biomarker subgroup (further specified in the DPP) at baseline:
  - OS/PFS curves for each treatment group will be estimated using the Kaplan-Meier product limit method. Two-sided, 95% confidence intervals for median OS/PFS will be constructed based on a log-log transformed CI for the survivor function S(t)
  - Forest plot of Hazard Ratios with 95% CIs.
- For GEP only, a histogram of the distribution of GEP between arms will be provided

Additional analyses may be performed to evaluate the association between these biomarkers and the efficacy.

7.9 Clinical Outcomes Assessments

The outcome research analyses will be performed using all randomized subjects at the time of final analysis of OS.

The analyses of EQ-5D-3L and FACT-HN outcomes will be restricted to all randomized subjects and the randomized subjects with PD-L1 CPS ≥ 20 who have an assessment at baseline and at least one post-baseline assessment.

7.9.1 EQ-5D-3L

The following descriptive analyses will be conducted:

- EQ-5D-3L questionnaire completion rate, defined as the proportion of questionnaires actually received out of the expected number (i.e. number of subjects on treatment or in follow up), will be calculated and summarized for each assessment time point by treatment group.
- A by-subject listing of the level of problems in each dimension, corresponding to EQ-5D-3L health state (i.e., 5-digit vector), EQ-5D-3L utility index score, and EQ-5D-3L VAS score will be provided.
- Proportion of subjects reporting problems for the 5 EQ-5D-3L dimensions at each assessment time point will be summarized by level of problem and by treatment group. Percentages will be based on number of subjects assessed at assessment time point. Separate analyses will be performed including or excluding cases with missing data
- For the EQ-5D-3L utility index and VAS scores, separately:
  - Mean score and mean change from baseline at each assessment time point will be summarized by treatment group using descriptive statistics (N, mean with SD and 95% CI, median, first and third quartiles, minimum, maximum).
  - A line graph summarizing the mean changes from baseline will be produced.

7.9.2 FACT-HN

FACT-HN data will be described by treatment group as randomized in the following ways:

- Scores and post-baseline changes in scores for the following components of the FACT-HN will be summarized at each assessment time point using descriptive statistics (ie, N, mean with SD and 95% CI, median, first and third quartiles, minimum, maximum): FACT-G, PWB, SWB, EWB, FWB, HNC, FHNSI-10, FACT-HN TOI, and FACT-HN total.
- A line graph summarizing the mean changes from baseline will be produced.
- The proportion (N) of subjects with symptomatic deterioration, defined as a clinically meaningful decline in FHNSI-10 score (worsening from baseline ≥ 3 points), will be summarized at each assessment time point.
- TTSD will be estimated using the Kaplan-Meier (KM) product limit method and will be displayed graphically. A table will be produced presenting number of events, number of subjects involved, medians, and 95% CIs for the medians. Median values of TTSD, along with two-sided 95% CI in each treatment group will be computed based on a log-log transformation method.
- The proportion (N) of subjects with symptomatic deterioration, defined as a clinically meaningful decline in FHNSI-10 score (worsening from baseline ≥ 3 points), or death, will be summarized at each assessment time point.
- Time to symptomatic deterioration or death will be estimated using the Kaplan-Meier (KM) product limit method and will be displayed graphically. A table will be produced presenting number of events, number of subjects involved, medians, and 95% CIs for the medians. Median values of time to symptomatic deterioration or death, along with two-sided 95% CI in each treatment group will be computed based on a log-log transformation method.

- Proportion (N) of subjects endorsing each response option for item GP5, "I am bothered by side effects of treatment," will be summarized by treatment group at each assessment time point. Separate analyses will be performed including or excluding cases with missing data.
- Proportion (N) of subjects with post-baseline improvement, stability, or worsening in GP5 item response will be summarized by treatment group at each assessment time point (except baseline). Separate analyses will be performed including or excluding cases with missing data.

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[21]
- 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[22].
- 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 progression 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 progression date will be compared to the date of death. The minimum of the imputed progression date and date of death will be considered as the date of progression.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.

- For date of progression 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 CONTENT OF REPORTS

All analyses describe in this SAP will be included in the final Clinical Study Report. Refer to the Data Presentation Plan for mock-ups of all tables and listings.

10 DOCUMENT HISTORY

Table 10-1: Document Revision History

Version Number	Author(s)	Description
1.0	[redacted]	Initial release dated 21-Oct-2019
2.0		• Per US FDA feedback, removed the second formal interim analysis (at 90% information fraction). • Added to Section 7.9.2 the analysis of time to symptomatic deterioration or death. • Corrected a typo in Table 7.4.1-3 replacing number 75 with 100.

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 A1-1 is summarizing key derivation steps for each type of clustered AEs.

Time-to resolution of AE (any grade) for a specific category is defined as the longest time from onset to complete resolution or improvement to the grade at baseline among all clustered AEs experienced by the subject in this category per adverse event criteria category. Events which worsened into grade 5 events (death) or have a resolution date equal to the date of death are considered unresolved. If a clustered AE is considered as unresolved, the resolution date will be censored to the last known alive date. Improvement to the grade at baseline implies that all different events in the clustered adverse event should at least have improved to the corresponding (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 A1-1: Derivation of Clustered AE

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

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

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

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

APPENDIX 2 MISSING AND PARTIAL RADIOTHERAPY AND SURGERY DATES IMPUTATION ALGORITHMS

Procedures – Imputation Rules.

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

In case of partial date use imputation rules described below:

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

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

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

In case of partial date use imputation rules described below:

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

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

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

Surgeries – Imputation Rules.

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

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

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

- If only day is missing then impute as the first day of the month;
- If both day and month are missing then then impute as 01JAN of the year;
- If date is completely missing or invalid then leave missing.

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

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

APPENDIX 3 IMMUNOGENICITY ANALYSIS: BACKGROUND AND RATIONALE

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

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

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

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

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

Clinical consequences of ADA can range from no apparent clinical effect to lack of efficacy (primary treatment failure), loss of efficacy (secondary treatment failure) or heightened effect due to

altered exposure to the biologic drug, adverse drug reactions (administration-related systemic or site reactions), and severe adverse drug reactions (anaphylaxis and unique clinical problems associated with cross-reactivity and neutralization of endogenous molecules). Thus it becomes important to examine any associations between ADA or any of its attributes with the various clinical sequelae. The presence of ADA may or may not preclude the administration of drug to ADA-positive subjects because the outcome is dependent upon the magnitude of the impact of ADA on PK and PD. Hence, the relationship of ADA with PK/PD is an important additional consideration, but does not necessarily result in a clinically impactful consequence per se.

Immunogenicity Endpoints

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

Terms and Definitions

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

Sample ADA Status**:**

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

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

Subject ADA Status:

- Baseline ADA-positive subject: A subject with baseline ADA-positive sample
- **ADA-positive subject**: A subject with at least one ADA positive-sample relative to baseline at any time after initiation of treatment
  1) *Persistent Positive (PP):* ADA-positive sample at 2 or more consecutive time points, where the first and last ADA-positive samples are at least 16 weeks apart
  2) *Not PP-Last Sample Positive:* Not persistent positive with ADA-positive sample at the last sampling time point
  1) *Other Positive:* Not persistent positive but some ADA-positive samples with the last sample being negative

2) *Neutralizing Positive:* At least one ADA-positive sample with neutralizing antibodies detected

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

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

Population for Analyses

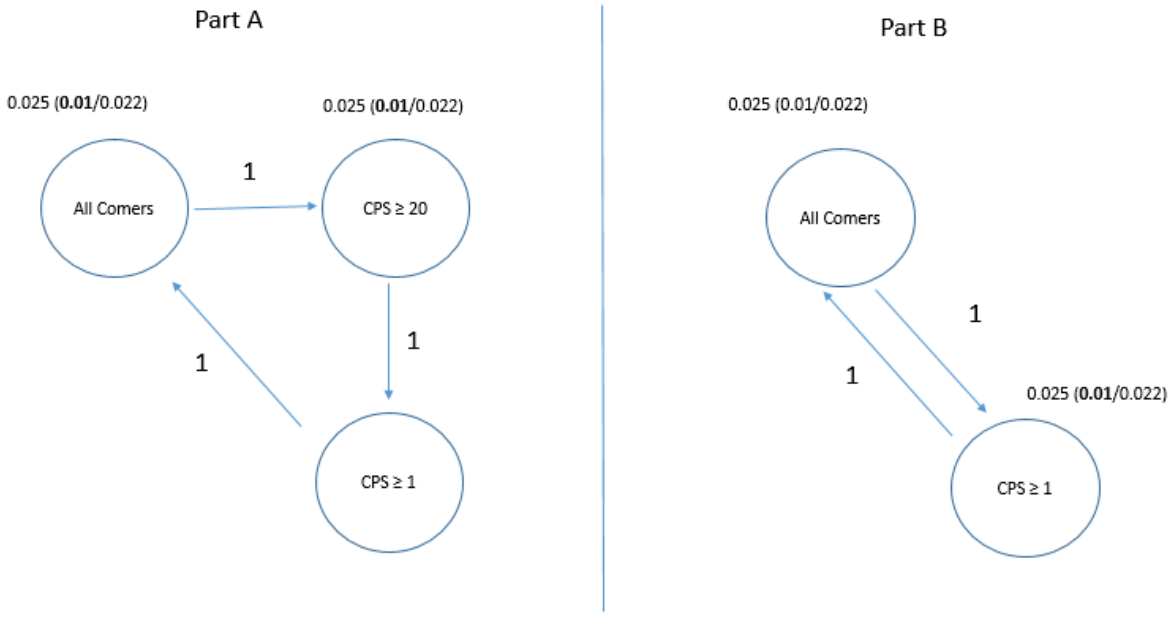
Analysis of immunogenicity data will be based on ADA evaluable subjects defined as all treated subjects with baseline and at least 1 post-baseline immunogenicity assessment. Analysis dataset and data listing will include all available ADA samples. However, subject-level ADA status will be defined based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

APPENDIX 4 SEQUENTIALLY REJECTIVE PROCEDURE

Section 7.5.6 of this SAP specifies the multiple testing procedure that controls family-wise type I error rate of 5% in the strong sense across primary and secondary endpoints. After rejecting an individual hypothesis, the local significance levels and the transition weights of the edges are updated: the rules for that update are determined by Algorithm 1 in Maurer and Bretz (2013)[20]. The resulting sequentially rejective testing procedure is uniquely determined by the graph in Figure 7.5.6-1. Below example illustrates the principles of this iterative procedure.

In this study, two primary endpoints (OS in all randomized subjects and OS in subjects with PD-L1 CPS ≥ 20) and one secondary endpoint (OS in subjects with PD-L1 CPS ≥ 1) will be tested at the interim analysis (Time 1) and final analysis (Time 2) as described in Section 5.

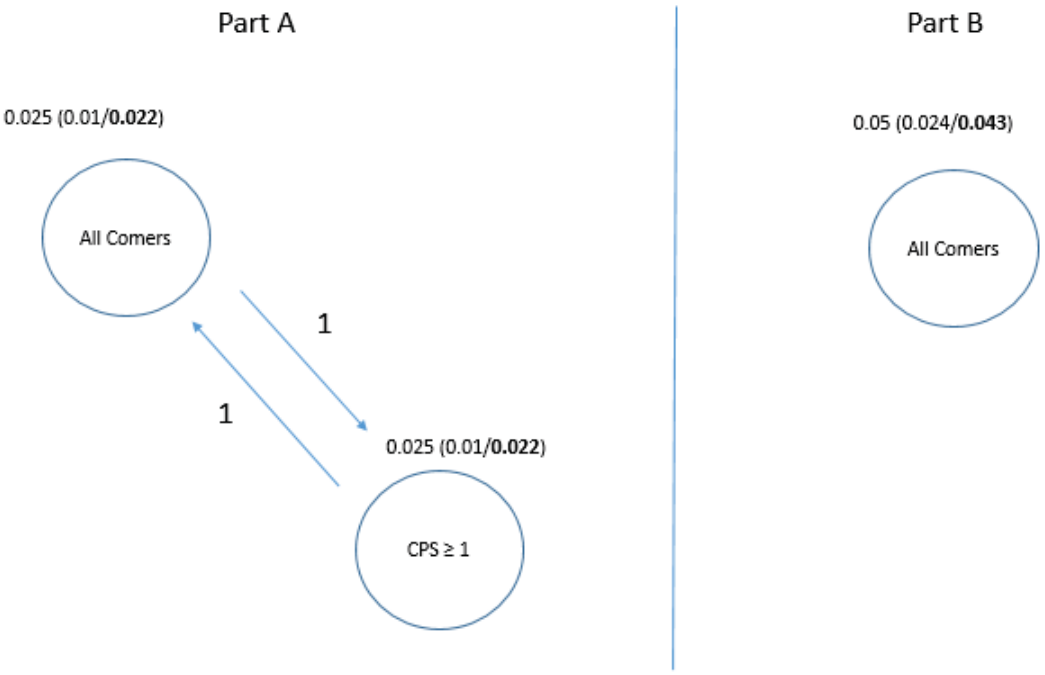
Step 1: At Time 1, OS in all randomized subjects and OS in subjects with PD-L1 CPS ≥ 20 will be respectively tested in nominal alpha 0.01 (obtained from O'Brien and Fleming α-spending function given information fraction of 80% and initial overall alpha is equally split for the two endpoints). Assuming only OS in subjects with PD-L1 CPS ≥ 20 is rejected, it will be removed from the graph and the entire nominal alpha (0.025) for this endpoint will be passed to the secondary endpoint OS in subjects with PD-L1 CPS ≥ 1. The below figure illustrates the update of the graph before (part A) and after (part B) Step 1.



Step 2: At Time 1, OS in subjects with PD-L1 CPS ≥ 1 will be tested at 0.01 (the nominal alpha is determined by the actual information fraction for this population and can vary from 0.01). If it is not significant, graph will not change and continue to Step 3.

Step 3: At Time 2 (final analysis), OS in all subjects and in subjects with PD-L1 CPS ≥ 1 will each be tested at 0.022. Assuming only OS in subjects with PD-L1 CPS ≥ 1 is rejected, it will be removed from the graph and the entire nominal alpha (0.025) for this endpoint will be passed to

the endpoint OS in all randomized subjects. The below figure illustrates the update of the graph before and after Step 3.



Step 4: The overall alpha for OS in all randomized subjects becomes 0.05 and this endpoint can be retested with nominal alpha of 0.043 at Time 2 (final analysis).

Similarly, if OS in all randomized subjects is significant in any of the 2 time points but OS in subjects with PDL1 CPS ≥ 20 is not, the latter will be re-tested with updated nominal alpha obtained from O'Brien and Fleming α-spending function using an overall alpha of 0.05. Subsequently, if OS in subjects with PD-L1 CPS ≥ 20 is significant, the OS in subjects with PD-L1 CPS ≥ 1 will also be tested with nominal alpha obtained from O'Brien and Fleming α-spending function using an overall alpha of 0.05.

11 REFERENCES

1 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011 ; 6(2):69-90

2 NCI SEER Statistics Review http://seer.cancer.gov/statfacts/html/oralcav.htmlaccessed March 5,2016

3 Leemans CR, Braakhuis B, Brakenhoff R. Molecular biology of head and neck cancer. Nature Reviews Cancer Vol 11 Jan 2011

4 NCCN Guidlines Version 2.2014 Head and Neck Cancers

5 Vermoken et al Platinum-based chemotherapy plus cetuximab in head and neck cancer. N Engl J Med 2008;359:1116-27

6 Ferris et al. Nivolumab vs investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck: 2-year long-term survival update of CheckMate 141 with analyses by tumor PD-L1 expression. Oral Oncology 81 (2018) 45–51

7 Curran MA, Montalvo W, Yagita H, et al. PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. PNAS 2010; 107: 4275-80.

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

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

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

11 Yount S, List M, Du H et al. A randomized validation study comparing embedded versus extracted FACT Head and Neck Symptom Index scores. Qual Life Res. 2007; 16: 1615-1626.

12 Robert L. Ferris, George Blumenschein Jr, Jerome Fayettec, Joel Guigay et al.Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. N Engl J Med 2016; 375:1856-1867

13 Burtness B, Harrington KJ, Greil R, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomized, open-label, phase 3 study. Lancet. 2019;394:1915-1928.

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

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

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

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

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

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

[20] W Maurer, and F Bretz (2013). Multiple testing in group sequential trials using graphical approaches. Statistics in Biopharmaceutical Research. Vol 5, 2013.

[21] Adverse Event Domain Requirements Specification Bristol Myers Squibb Co. PRI. Version 2.3.0 April 23, 2018

[22] Non-Study Medication Domain Requirements Specification Bristol Myers Squibb Co. PRI. Version 2.10.0 April 23, 2018