Official Title of Study:

A Phase 3, Randomized, Double-blind Study of Adjuvant Immunotherapy with Nivolumab versus Ipilimumab after Complete Resection of Stage IIIb/c or Stage IV Melanoma in Subjects who are at High Risk for Recurrence

NCT Number: NCT02388906

Document Date (Date in which document was last revised): April 3, 2019

STATISTICAL ANALYSIS PLAN
FOR DMC

A PHASE 3, RANDOMIZED, DOUBLE-BLIND STUDY OF ADJUVANT IMMUNOTHERAPY WITH NIVOLUMAB

VERSUS IPILIMUMAB AFTER COMPLETE RESECTION OF STAGE IIIB/C OR STAGE IV MELANOMA IN SUBJECTS

WHO ARE AT HIGH RISK FOR RECURRENCE

(CHECKMATE 238: CHECKPOINT PATHWAY AND NIVOLUMAB CLINICAL TRIAL EVALUATION 238)

PROTOCOL(S) CA209238

VERSION # 3.0

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 BACKGROUND AND RATIONALE

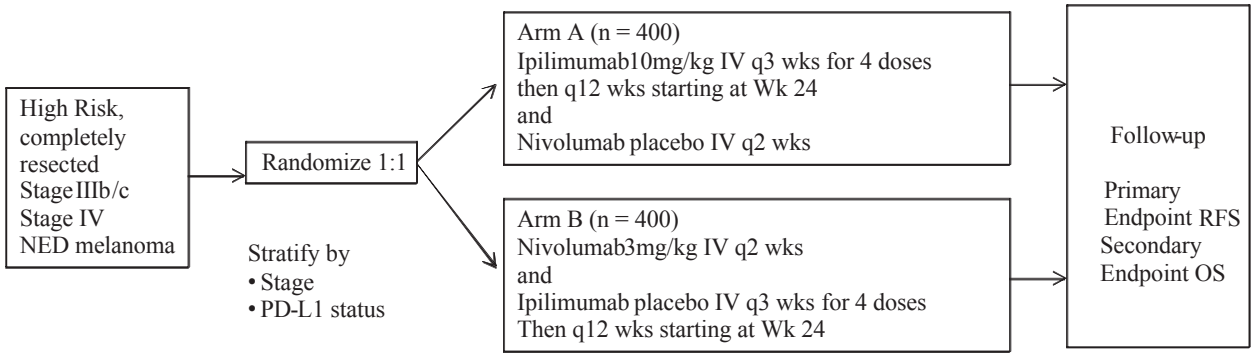
[redacted]

2 STUDY DESCRIPTION

2.1 Study Design

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

Figure 2.1-1: Study Design Schematic



For both arms, the treatment duration is maximum one year.

The subjects will be treated in both arms until disease recurrence, unacceptable toxicity, or subject withdrawal of consent with a maximum of 1-year total duration of study medication.

Timing for safety analysis:

The schedule of DMC meetings are outlined in Table 4.1-1 of the DMC charter[4].

2.2 Treatment Assignment

Refer to protocol[5] Section 4.4.

2.3 Blinding and Unblinding

Refer to DMC charter[4] Section 4.4.

2.4 Protocol Amendments

None.

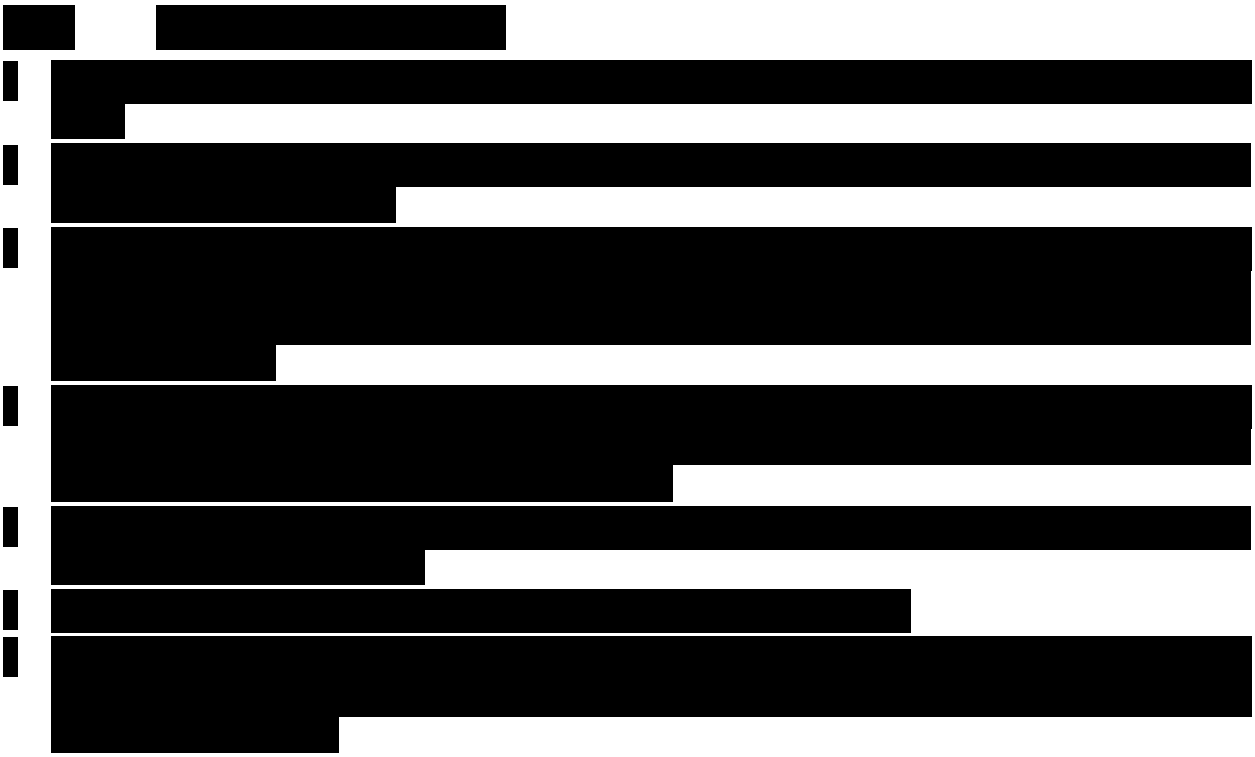
3 OBJECTIVES

3.1 Primary

- To compare the efficacy, as measured by RFS, provided by nivolumab versus ipilimumab in subjects with completely resected Stage IIIb/c or Stage IV NED melanoma who are at high risk for recurrence

3.2 Secondary

- To compare the overall survival of nivolumab vs. ipilimumab in subjects with completely resected Stage IIIb/c or Stage IV NED melanoma who are at high risk for recurrence.
- To assess the overall safety and tolerability of nivolumab and ipilimumab in subjects with completely resected Stage IIIb/c or Stage IV NED melanoma who are at high risk for recurrence.
- To evaluate whether PD-L1 expression is a predictive biomarker for RFS.
- To evaluate the Health Related Quality of Life (HRQoL) as assessed by European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30.



4 ENDPOINTS

4.1 Recurrence-Free-Survival

The primary endpoint is RFS. The primary endpoint of RFS will be programmatically determined based on the disease recurrence date provided by the investigator and is defined as the time between the date of randomization and the date of first recurrence (local, regional or distant metastasis), new primary melanoma, or death (whatever the cause), whichever occurs first. A subject who dies without reported recurrence will be considered to have recurred on the date of death. For subjects who remain alive and whose disease has not recurred, RFS will be censored on the date of last evaluable disease assessment. For those subjects who remain alive and have no recorded post-randomization tumor assessment, RFS will be censored on the day of randomization. Censoring rules for the primary analysis of RFS are presented in Table 4.1-1.

Table 4.1-1: Censoring Scheme for Primary Definition of RFS

Situation	Date of Event or Censoring	Outcome
Recurrence (local, regional, distant, new primary melanoma)	Date of first recurrence	Event
Death without recurrence	Date of death	Event
Disease at baseline	Date of randomization	Event
No baseline disease assessment and no death	Date of randomization	Censored
No on-study disease assessments and no death	Date of randomization	Censored
No recurrence and no death	Date of last evaluable disease assessment	Censored
New anticancer therapy, tumor-directed radiotherapy, or tumor-directed surgery received without recurrence reported prior to or on the same day of disease assessment	Date of last evaluable disease assessment prior to or on the same date of initiation of subsequent therapy	Censored
Second non-melanoma primary cancer reported prior or on the same day of disease assessment	Date of last evaluable disease assessment prior to or on the same date of diagnosis of second non-melanoma primary cancer	Censored

4.2 Overall Survival

Overall survival 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 ("last known alive date").

4.3 Safety Endpoints

The safety endpoints include serious and non-serious adverse events, select adverse events, laboratory evaluations, exposure, deaths, and reasons off treatment.

Definition of Select Adverse Events

To fully characterize AEs of clinical importance that are potentially associated with the use of nivolumab, the Sponsor identified AEs, termed Select AEs, meeting any of the following 4 criteria:

- AEs that may differ from AEs caused by non-immunotherapies.
- AEs that may require immunosuppression as part of their management.
- AEs whose early recognition and management may mitigate severe toxicity.
- AEs for which multiple event terms may be used to describe a single type of AE, thereby necessitating the pooling of terms for full characterization

The select adverse events (select AEs) consist of a list of preferred terms grouped by specific category (e.g. pulmonary events, gastrointestinal events categories, see Table 4.3-1). The select

AE categories and terms are defined by the Sponsor and the list that is the most current at the time of analysis will be used. This list will be based on the most current version of MedDRA at the time of CA209238 database lock.

Table 4.3-1: Select Adverse Events

Category	Preferred Terms
Endocrine Adverse Events	**ADRENAL DISORDER**: ADRENAL INSUFFICIENCY ADRENAL SUPPRESSION BLOOD CORTICOTROPHIN DECREASED BLOOD CORTICOTROPHIN INCREASED HYPOTHALAMIC PITUITARY ADRENAL AXIS SUPPRESSION SECONDARY ADRENOCORTICAL INSUFFICIENCY **DIABETES:** DIABETES MELLITUS LATENT AUTOIMMUNE DIABETES IN ADULTS **PITUITARY DISORDER:** HYPOPHYSITIS **THYROID DISORDER:** AUTOIMMUNE THYROIDITIS BLOOD THYROID STIMULATING HORMONE DECREASED BLOOD THYROID STIMULATING HORMONE INCREASED HYPERTHYROIDISM HYPOTHYROIDISM THYROID FUNCTION TEST ABNORMAL THYROIDITIS THYROIDITIS ACUTE THYROXINE DECREASED THYROXINE FREE DECREASED THYROXINE FREE INCREASED THYROXINE INCREASED TRI-IODOTHYRONINE UPTAKE INCREASED
Hypersensitivity/Infusion Reactions	ANAPHYLACTIC REACTION ANAPHYLACTIC SHOCK BRONCHOSPASM HYPERSENSITIVITY INFUSION RELATED REACTION

Table 4.3-1: Select Adverse Events

Category	Preferred Terms
Gastrointestinal Adverse Events	COLITIS
	COLITIS ULCERATIVE
	DIARRHOEA
	ENTERITIS
	ENTEROCOLITIS
	FREQUENT BOWEL MOVEMENTS
	GASTROINTESTINAL PERFORATION
Hepatic Adverse Events	ACUTE HEPATIC FAILURE
	ALANINE AMINOTRANSFERASE INCREASED
	ASPARTATE AMINOTRANSFERASE INCREASED
	BILIRUBIN CONJUGATED INCREASED
	BLOOD BILIRUBIN INCREASED
	DRUG-INDUCED LIVER INJURY
	GAMMA-GLUTAMYLTRANSFERASE INCREASED
	HEPATIC ENZYME INCREASED
	HEPATIC FAILURE
	HEPATITIS
	HEPATITIS ACUTE
	HYPERBILIRUBINAEMIA
	LIVER DISORDER
	LIVER FUNCTION TEST ABNORMAL
	LIVER INJURY
	TRANSAMINASES INCREASED
Pulmonary Adverse Events	ACUTE RESPIRATORY DISTRESS SYNDROME
	ACUTE RESPIRATORY FAILURE
	INTERSTITIAL LUNG DISEASE
	LUNG INFILTRATION
	PNEUMONITIS
Renal Adverse Events	BLOOD CREATININE INCREASED
	BLOOD UREA INCREASED
	CREATININE RENAL CLEARANCE DECREASED
	HYPERCREATININAEMIA
	NEPHRITIS
	NEPHRITIS ALLERGIC
	NEPHRITIS AUTOIMMUNE
	RENAL FAILURE
	RENAL FAILURE ACUTE
	RENAL TUBULAR NECROSIS
	TUBULOINTERSTITIAL NEPHRITIS
	URINE OUTPUT DECREASED

Table 4.3-1: Select Adverse Events

Category	Preferred Terms
Skin Adverse Events	BLISTER
	DERMATITIS
	DERMATITIS EXFOLIATIVE
	DRUG ERUPTION
	ECZEMA
	ERYTHEMA
	ERYTHEMA MULTIFORME
	EXFOLIATIVE RASH
	PALMAR-PLANTAR ERYTHRODYSAESTHESIA SYNDROME
	PHOTOSENSITIVITY REACTION
	PRURITUS
	PRURITUS ALLERGIC
	PRURITUS GENERALISED
	PSORIASIS
	RASH
	RASH ERYTHEMATOUS
	RASH GENERALISED
	RASH MACULAR
	RASH MACULO-PAPULAR
	RASH PAPULAR
	RASH PRURITIC
	SKIN EXFOLIATION
	SKIN HYPOPIGMENTATION
	SKIN IRRITATION
	STEVENS-JOHNSON SYNDROME
	TOXIC EPIDERMAL NECROLYSIS
	URTICARIA
	VITILIGO

Source: MedDRA version 17.1

Laboratory results, adverse events, and other symptoms will be graded using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0, except where CTCAE grades are not available.

5 SAMPLE SIZE AND POWER

The sample size is calculated to compare RFS between subjects randomized to receive nivolumab vsipilimumab. RFS will be evaluated for a treatment effect at an overall alpha level of 0.05 (two-sided) with approximately 85% power. The number of events and power were calculated assuming a delayed treatment effect and cure fraction.

Approximately 900 subjects have been randomized to the two treatment arms in a 1:1 ratio. Based on current accrual and revised piece-wise accrual rate assumptions, accrual duration is 7.5 months. Taking into account the actual AJCC disease stage distribution (about 80% of Stage III subjects and 20% of Stage IV subjects), higher cure rates, and some early drop-out, the original planned 507 events might not be reached by the final RFS analysis (ie, after all subjects have a minimum of 36 months of follow-up). Approximately 450 RFS events are anticipated at the final RFS analysis, ensuring at least 85% power to detect a hazard ratio of 0.75 (critical hazard ratio of 0.83) with an overall type I error of 0.05 (two-sided). Sample size calculations for this study design were done using EAST 6 (v 6.3.1) and R.

An interim analysis of RFS will take place after all subjects have a minimum of 18 months of follow-up. Approximately 350 RFS events are anticipated at this analysis. The stopping boundary at the interim analysis will be derived based on the exact number of RFS events at interim using Lan-DeMets alpha spending function with O'Brien-Fleming boundaries. With an interim RFS analysis at 350 RFS events (about 78% information fraction), the critical hazard ratio would be 0.78 and the type I error would be 0.022 (two-sided). The type I error to be used for final RFS analysis would be 0.043 (two-sided).

One formal interim analysis of OS will be conducted at the time of the final RFS analysis. The formal interim comparison of OS will be tested via a stratified log-rank test using the interim monitoring feature of EAST software based on a generalization of the Lan-DeMets error spending function approach using an O'Brien-Fleming stopping boundary to reject H0, controlling for a two-sided alpha of 0.05 at the end of the study. The boundary for concluding superiority based on the actual number of OS events pooled across the two arms in the locked database will be computed by the independent statistician in collaboration with BMS. The boundary for concluding superiority at interim will be provided to the DMC by the independent statistician.

5.1 Populations for analyses

The following populations will be used in the DMC analysis reports:

- **All enrolled subjects**: All subjects who signed an informed consent and was assigned a unique ID by the IVRS.
- **All randomized subjects**: All subjects who were randomized to any treatment arm in the study. Analysis of demography, protocol deviations, baseline characteristics, primary efficacy analysis will be performed for this population.
- **All treated subjects**: All subjects who received any dose of study therapy. This is the primary dataset for drug exposure and safety analysis.

6 STATISTICAL ANALYSES

6.1 General Methods

Continuous variables will be summarized using descriptive statistics; i.e. number of non-missing observations (n), mean, standard deviation (STD), median, minimum, maximum and quartiles. Categorical variables will be summarized by frequencies and percentages. The Kaplan-Meier (KM) product limit method will be used to estimate the distribution and median of each time-to-event endpoint in which censoring is involved. If required, the log-log transformation method will be used to compute a 95% confidence interval (CI) for the median of each time-to-event endpoint. If required a Cox proportional hazards model will be used to compute an estimate of the hazard ratio of the investigational to the control arm, for time to event endpoints. Hazard ratios for time-to-event variables will be rounded to two decimal places.

Laboratory results, adverse events, and other symptoms will be graded using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE), version 4.0, except where CTCAE grades are not available. Adverse events will be categorized using the most current version of Medical Dictionary for Regulatory Activities (MedDRA), by system organ class and preferred term. Prior therapies will be summarized using the most current version of the World Health Organization (WHO) drug dictionary.

In the open report, the summary tables and listings will be presented as pooled across the two treatment arms, without revealing the treatment identity. In the closed report, partially un-blinded summaries and listings with treatment arm labeled as “Arm A” and “Arm B will be presented. The contents of these reports are outlined in Section 4.3 of the DMC charter[4].

Statistical analyses will be carried out in SAS (Statistical Analysis System, SAS Institute, North Carolina, USA), unless otherwise indicated.

Unless otherwise specified, analyses described in the next sections will be produced for both closed and open reports.

6.2 Study Conduct

6.2.1 Accrual

The accrual pattern will be summarized per month for all enrolled subjects.

6.2.2 Relevant Protocol Deviations

The following programmable deviations will be considered as relevant protocol deviations and summarized by treatment group and overall for all randomized subjects.

At Entrance:

- No histologically documented stage III or stage IV melanoma as per AJCC staging
- Subject with Baseline Ecog Status > 1
- Subject received prior systemic anti cancer therapy

- Subject with ocular/uveal melanoma

On-treatment:

- Subjects receiving anti-cancer therapy (chemotherapy, hormonal therapy, immunotherapy, standard or investigational agents for treatment of cancer) while on study therapy
- Subjects treated differently than as randomized (subjects who received the wrong treatment, excluding the never treated)

A subject listing with subjects treated differently than randomized will be presented.

6.3 Study Population

Unless otherwise specified, analyses will be performed on the all randomized population.

6.3.1 Subject Disposition

The total number of subjects randomized (treated or not treated) will be summarized.

Number of subjects who discontinued study treatment along with corresponding reason will be tabulated by treatment group as treated. This analysis will be performed only on the all treated subjects population.

The current status of follow-up for overall survival will be summarized.

6.3.2 Demographics and Other Baseline Characteristics

Descriptive statistics will be summarized for the following baseline characteristics by treatment group as randomized.

- Age (descriptive statistics)
- Age category (< 65, ≥ 65 - < 75, ≥ 75)
- Gender (male/female)
- Race (white/black/asian/other)
- Region (North America vs. Western Europe vs. Eastern Europe vs. Asia vs. Australia vs. ROW)
- Baseline ECOG Performance Status (0,1,...)
- Baseline LDH (≤ ULN, > ULN)
- Baseline LDH (≤ 2 x ULN, > 2 x ULN)
- Baseline weight (kg).
- Time from Surgical Resection to Randomization (in weeks): <3, 3 - <6, 6 - <9, 9 - <12, 12 - < 15, 15 - <18, 18 - <21, ≥ 21)
- Baseline CTC grade for laboratory parameters
- Melanoma sub-type (mucosal, cutaneous, acral, ocular/uveal, other)
- Tumor origin (primary vs. recurrent)

- Tumor ulceration (yes/no) in Stage III subjects (Source: eCRF)
- Macro/Microscopic lymph node involvement in Stage III subjects (Source: eCRF)
- M-stage in Stage IV subjects (Source: eCRF)

Similarly the following IVRS data will be summarized by treatment group as randomized.

- Disease Stage (Stage IIIb/c vs. Stage IV M1a-M1b vs. Stage IV M1c)
- PD-L1 status (positive vs negative/indeterminate)

6.4 Extent of Exposure

Unless otherwise specified, the primary population will be the all treated population.

6.4.1 Administration of Study Therapy

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

- Time from randomization to first dose of study therapy (≤ 3 days, 4-5 days, 6-7 days, 8-14 days, 15-21 days, > 21 days) - Open and Closed Reports
- Duration of treatment (in months) - Closed Report Only
- Relative dose intensity (%) - Closed Report Only

6.5 Efficacy - Closed Report Only

Unless otherwise specified, the primary population will be the all randomized population.

RFS

An interim analysis of RFS will take place after all subjects have a minimum of 18 months of follow-up. Approximately 350 RFS events are anticipated at this analysis. The stopping boundary at the interim analysis will be derived based on the exact number of RFS events at interim using Lan-DeMets alpha spending function with O'Brien-Fleming boundaries. With an interim RFS analysis at 350 RFS events (about 78% information fraction), the critical hazard ratio would be 0.78 and the type I error would be 0.022 (two-sided). The type I error to be used for final RFS analysis (with 450 RFS anticipated) would be 0.043 (two-sided).

All analyses of RFS will be performed at the time of this interim analysis only.

The primary RFS analyses will be conducted using a two-sided log-rank test stratified by PD-L1 status and stage at screening (per IVRS) in all randomized subjects. The two-sided log-rank p-value will be reported. The estimate of the RFS hazard ratio, of nivolumab to ipilimumab, will be calculated using a stratified Cox proportional hazards model, with treatment as the single covariate, stratified by the above stratification factors. A two-sided (1-adjusted α)% CI for the hazard ratio will also be presented.

The RFS distribution for each treatment group will be estimated using Kaplan-Meier techniques. Median RFS along with 95% CI will be constructed based on a log-log transformed CI for the survivor function.

Current status of follow-up for RFS will also be provided:

- event/censored
- for subjects with event: source of RFS
  - recurrence
    - ♦ local recurrence
    - ♦ regional recurrence (in-transit metastasis or regional node recurrence)
    - ♦ distant metastasis
    - ♦ new primary melanoma
  - death
- for subjects who are censored, the time between data cut-off and last tumor assessment

The status of subjects who are censored in the RFS KM analysis will be tabulated using following categories:

- Censored on randomization date
  - no baseline disease assessment
  - no on-study disease assessment and no death
- Censored on date of last disease assessment on-study
  - Received subsequent anti-cancer therapy
  - Second non-melanoma primary cancer
  - Still on treatment
  - In follow-up
  - Off study
    - ♦ lost to follow-up
    - ♦ subject withdrew consent
    - ♦ other

OS

The OS distributions will be estimated by treatment group and graphically displayed using Kaplan-Meier methodology. Disposition status of subjects censored in the OS analysis will be also summarized.

6.5.1 Follow-up Therapy - Closed Report only

Number and percentage of subjects receiving subsequent therapies including radiotherapies, surgeries and systemic therapies will be reported.

6.6 Safety Analyses - Closed Report Only

Safety analyses will be performed in all treated subjects. Descriptive statistics of safety will be presented using NCI CTCAE version 4.0 by treatment group. All on-study AEs, drug-related AEs, SAEs, and drug-related SAEs, AEs leading to discontinuation, drug-related AEs leading to discontinuation, select AEs and drug-related select AEs will be tabulated using worst grade per NCI CTCAE v 4.0 criteria by system organ class and preferred term. On-study lab parameters including hematology, chemistry, liver function, and renal function as well as electrolytes will be summarized using worst grade NCI CTCAE v 4.0 criteria.

Unless otherwise noted, the analyses of AEs will include events occurring on-study (i.e., from first day of study treatment and no later than 30 days following the last day of study treatment, or events entered on an AE form but with a missing onset date).

AEs with a relationship to study therapy of related or unknown (not reported) will be considered drug-related.

The frequency of on-study AEs, coded using MedDRA system organ classes and preferred terms will be calculated.

Laboratory parameters will include hemoglobin, platelets, neutrophils, WBC, lymphocytes, AST, ALP, ALT, total bilirubin, creatinine, amylase and lipase as well as electrolytes (sodium, calcium, magnesium and potassium).

The following summaries will be produced:

- Summary of Any Adverse Events by Worst CTC Grade
- Summary of Drug-Related Adverse Events by Worst CTC Grade
- Summary of Serious Adverse Events by Worst CTC Grade
- Summary of Drug-Related Serious Adverse Events by Worst CTC Grade
- Summary of Adverse Events Leading to Discontinuation by Worst CTC Grade
- Summary of Drug-Related Adverse Events Leading to Discontinuation by Worst CTC Grade
- Summary of Any Select Adverse Events by Worst CTC Grade and by Category
- Summary of Drug-Related Select Adverse Events by Worst CTC Grade and by Category
- Intensity Rate of Serious Adverse Events by Onset Period
- Summary of incidence of deaths and reason for death.
- On-study Laboratory Parameters Summary by Worst CTC grade: Changes from Baseline Grade - SI Units

Death, reason for death and adverse events leading for discontinuation will be also listed.

6.7 Interim Analysis of Overall Survival

At the time of formal interim analysis of OS, the following outputs will solely be created:

The primary OS analyses will be conducted using a two-sided log-rank test stratified by PD-L1 status and stage at screening (per IVRS) in all randomized subjects. The two-sided log-rank p-value will be reported. The estimate of the OS hazard ratio, of nivolumab to ipilimumab, will be calculated using a stratified Cox proportional hazards model, with treatment as the single covariate, stratified by the above stratification factors. A two-sided (1-adjusted α)% CI for the hazard ratio will also be presented.

The OS distribution for each treatment group will be estimated using Kaplan-Meier techniques. Median OS along with 95% CI will be constructed based on a log-log transformed CI for the survivor function.

Current status of follow-up for OS will also be provided:

- event/censored
- for subjects who are censored, the time between data cut-off and last tumor assessment

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

- Still on treatment
- In follow-up
- Off study
  - lost to follow-up
  - subject withdrew consent
  - other

7 CONTENTS OF REPORTS

Refer to Section 4.3 of DMC Charter[4] and Section 1.2 of DMC DPP[6].