Official Title of Study:

A RANDOMIZED, DOUBLE-BLIND PHASE 2/3 STUDY OF RELATLIMAB COMBINED WITH NIVOLUMAB VERSUS NIVOLUMAB IN PARTICIPANTS WITH PREVIOUSLY UNTREATED METASTATIC OR UNRESECTABLE MELANOMA

PROTOCOL(S) CA224-047

NCT Number: NCT03470922

Document Date (Date in which document was last revised): October 25, 2021

STATISTICAL ANALYSIS PLAN
FOR CSR

A RANDOMIZED, DOUBLE-BLIND PHASE 2/3 STUDY OF RELATLIMAB COMBINED WITH NIVOLUMAB VERSUS NIVOLUMAB IN PARTICIPANTS WITH PREVIOUSLY UNTREATED METASTATIC OR UNRESECTABLE MELANOMA

PROTOCOL: CA224-047

VERSION # 5.0

DATE: 25-Oct-2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 BACKGROUND AND RATIONALE

Individually targeting immune checkpoint receptors such as CTLA-4 and PD-1 has been clinically successful across multiple tumor types. It is possible, however, that combination therapies could potentially lead to greater depth of response and overall survival as has been noted with anti-PD-1 and anti-CTLA-4 combination therapy in advanced melanoma patients. Targeting additional checkpoints, such as LAG-3, is considered a promising, novel approach. Blockage of LAG-3 in combination with anti-PD-1 has potential to improve efficacy in comparison with blocking anti-PD-1 alone without adding significant toxicity.

This study aims to demonstrate that treatment with BMS-986213 (fixed-dose combination relatlimab/nivolumab at a 1:3 ratio) shows improved PFS compared with nivolumab monotherapy in participants with previously untreated, unresectable, or metastatic melanoma. Additional objectives of the study include characterization of safety and tolerability, pharmacokinetics, potential predictive biomarkers, and changes in patient-reported outcomes for quality-of-life assessments.

Research Hypothesis:

Treatment with BMS-986213 will improve PFS when compared to nivolumab monotherapy in previously untreated participants with unresectable or metastatic melanoma.

Schedule of Analyses:

PFS is the primary endpoint for this study. The analysis for the primary PFS endpoint is planned to occur when 365 participants have had a PFS event. Total enrollment will take approximately 27 months (including a potential 6-month halt due to the interim efficacy analysis).

An independent Data Monitoring Committee (DMC) will have access to periodic interim safety and efficacy reports to allow for a risk/benefit assessment. In addition, the DMC will evaluate an interim analysis of PFS. This interim analysis will be performed when a minimum follow-up of 12 weeks is achieved for approximately 400 randomized participants or at least 150 PFS events have been observed using BICR. The analysis will include all available data at the time of the lock and will be performed by the DMC to maintain blinding of the sponsor. If the interim analysis is positive (see DMC Charter for specific criteria), then the study will start to randomize participants again to complete the Phase 3 study and it will remain blinded. In this case, the final Phase 2 PFS analysis will not be performed. Additional participants will be randomized (300 more) for approximately 700 total in the study. If the interim analysis does not support continuing to the Phase 3 study (see DMC Charter for specific criteria), no more participants will be randomized in the study, the Phase 2 data will mature, and then the sponsor will be unblinded for the PFS analysis.

Two interim analyses of OS are planned in the event the primary analysis of PFS is statistically significant (see Sections 2.3 and 5.2.1 for more details).

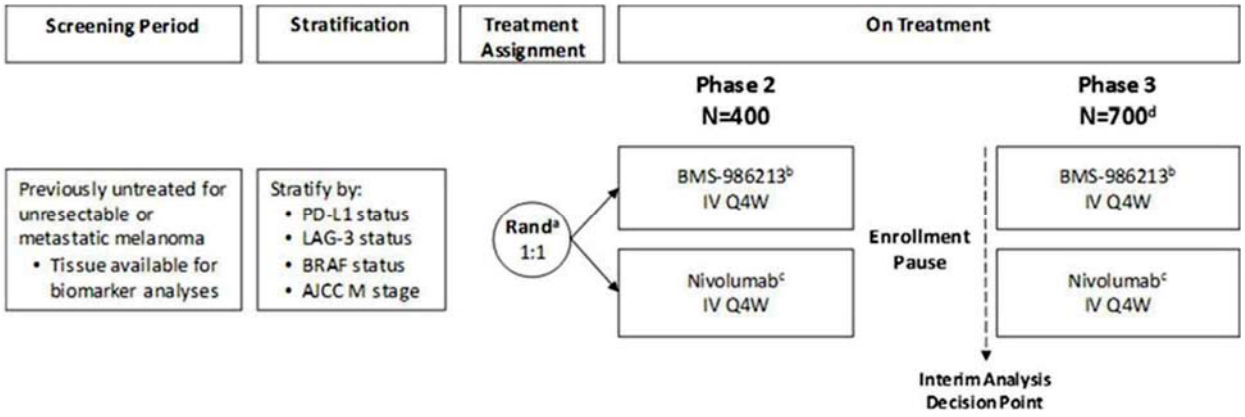
2 STUDY DESCRIPTION

2.1 Study Design

This is a Phase 2/3, randomized, double-blind study of BMS-986213 (FDC relatlimab/nivolumab at a 1:3 ratio) versus nivolumab monotherapy in adult and adolescent participants (≥ 12 years of age) with previously untreated, unresectable, or metastatic melanoma. Participants must have unresectable Stage III or Stage IV melanoma, per the 8th edition of the American Joint Committee on Cancer (AJCC) staging system, and must not have received prior systemic therapy for the treatment of unresectable or metastatic melanoma.

The study design schematic is presented in Figure 2.1-1

Figure 2.1-1: Study Design Schematic



A pre-treatment tumor sample to determine PD-L1 and LAG-3 status is required to be submitted from all participants prior to randomization. The sample must be obtained within 3 months prior to enrollment from a metastatic tumor lesion or from an unresectable primary tumor lesion that has not been previously irradiated; no intervening treatment may have been administered between the time of biopsy/surgery and study entry. The tumor sample will be submitted as either a formalin-fixed paraffin-embedded (FFPE) block (preferred) or minimum of 20 slides requested, obtained from core biopsy, punch biopsy, excisional biopsy, or surgical specimen. If sufficient tissue is not available from within 3 months prior to enrollment, then a fresh biopsy will be required during the screening period. Samples will be submitted to the analytical laboratory for PD-L1 and LAG-3 testing. The analytical laboratory must provide the interactive response technology (IRT) with the related results prior to randomization.

Participants must have a documented BRAF status prior to randomization. Those participants enrolling in this study without documented results must have testing performed locally and result (wild type or mutant) be available prior to randomization.

Depending on the treatment arm, the participant will receive BMS-986213 or nivolumab until disease progression, treatment discontinuation, withdrawal of consent, or the study ends.

On-study tumor assessments will begin 12 weeks from randomization and will continue every 8 weeks up to Week 52 and every 12 weeks thereafter. Tumor assessments should continue until disease progression confirmed by the BICR or treatment discontinuation, whichever occurs later. Treatment beyond initial investigator-assessed RECIST v1.1-defined progression is permitted if the participant has investigator-assessed clinical benefit and is tolerating study treatment.

The decision on whether to complete the Phase 3 enrollment (N=700) or stop at the Phase 2 enrollment (N=400) will be recommended by the Data Monitoring Committee (DMC) based upon the interim PFS analysis (PFS IA). If the Phase 2 enrollment is complete before a decision on completing the Phase 3 enrollment has been made, enrollment to the study will pause. Given estimates of accrual, this pause may be up to 5 to 6 months but could be shorter if enrollment is slower than anticipated.

2.2 Treatment Assignment

After the subject's initial eligibility is established and informed consent has been obtained, the subject must be enrolled into the study by accessing an Interactive Response Technologies web-based system (IRT) to obtain the subject number. Every subject that signs the informed consent form must be assigned a subject number in IRT. The investigator or designee will register the subject for enrollment by following the enrollment procedures established by BMS. The following information is required for enrollment:

- Date that informed consent was obtained
- Date of birth
- Sex at birth.

Once enrolled in IRT, enrolled subjects that have met all eligibility criteria will be ready to be randomized through the IRT. The following information is required for subject randomization:

- Subject number
- Date of birth
- PD-L1 Expression: PD-L1 ≥1% tumor cell surface expression vs. PD-L1<1% tumor cell surface expression /non-quantifiable (entered at IRT)
- LAG-3 Status: LAG-3 ≥ 1% expression/ LAG-3<1% expression (entered at IRT)
- BRAF Status: Mutation Positive/ Mutation Wild-type at IRT
- AJCC M Stage: M0/M1any[0] / M1any[1]) at IRT

Subjects meeting all eligibility criteria will be randomized in a 1:1 ratio to BMS-986213 or Nivolumab monotherapy and stratified by PD-L1 expression level (PD-L1 ≥1% expression vs PD-L1 < 1% expression), LAG-3 expression (LAG-3 ≥ 1% expression vs LAG-3 < 1% expression), BRAF (Mutation Positive/ Mutation Wild-type) and M stage (M0/M1any[0] / M1any[1]). The randomization procedures will be carried out via permuted blocks within each stratum.

2.3 Blinding and Unblinding

The sponsor, participants, investigator, and site staff will be blinded to the study therapy administered. Each investigative site must assign an unblinded pharmacist/designee.

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

Before breaking the blind of an individual participant's treatment, the investigator should determine that the unblinded information is necessary, i.e., that it will alter the participant's immediate management. In many cases, particularly when the emergency is clearly not related to the investigational product, the problem may be properly managed by assuming that the participant is receiving active product. It is highly desirable that the decision to unblind treatment assignment be discussed with the medical monitor, but the investigator always has ultimate authority for the decision to unblind. The principal investigator should only call in for emergency unblinding after the decision to discontinue the participant has been made.

For information on how to unblind in an emergency, please consult the IRT manual.

Any request to unblind a participant's treatment assignment for non-emergency purposes should be discussed with the medical monitor.

In cases of accidental unblinding, contact the medical monitor and ensure every attempt is made to preserve the blind.

The DMC will assess safety and risk benefit on an ongoing basis, and will have access to unblinded treatment codes for individual subjects. An analysis team external to BMS (at Axio Research LLC), including a reporting statistician and programming support, who are not involved with the conduct of the study, will provide analyses to the DMC. The procedures to be respected and the reasons for unblinding of the DMC are discussed in the DMC charter.

An interim analysis of PFS (PFS IA) will be performed when a minimum follow-up of 12 weeks is achieved for approximately 400 randomized subjects or at least 150 PFS events have been observed using BICR. If the decision from the interim analysis is not to proceed to Phase 3, the study will stop enrollment and remain blinded until the final analysis of the Phase 2 study which will be performed when a minimum of 183 events have been observed.

If PFS IA meets the pre-specified threshold and the study proceeds to Phase 3, the final PFS analysis will be performed when at least 365 PFS events have occurred per BICR. The sponsor will be unblinded at the time of the PFS FA (to PFS results and all individual treatment assignments) but participants, investigator, and site staff will remain blinded until a positive interim OS result or until final analysis of the OS endpoint, whichever comes first.

Assuming a statistically significant PFS result, there will be interim analyses of OS. The first interim analysis of OS (OS IA1) will be performed at the time of PFS FA. If the OS IA1 is not

statistically significant, the second interim analysis of OS (OS IA2) will be performed when a minimum of 90% (270/300) of the expected deaths have occurred. If OS IA2 is not statistically significant, the final OS analysis will be performed when at least 300 deaths have occurred.

Designated staff of BMS Research & Development may be unblinded prior to final PFS results to facilitate the bioanalytical analysis of pharmacokinetic, immunogenicity, and pharmacodynamic samples.

To further minimize potential bias, the participants and the investigative clinical site staff are blinded to results from PD-L1 and LAG-3 analysis.

2.4 Protocol Amendments

Table 2.4-1 summarizes changes in the protocol that affect the analyses.

Table 2.4-1: Summary of Changes to the Protocol

Description of Change	Brief Rationale
Amendment 1:	
Phase 2 primary endpoint from ORR to PFS.	
Updated interim PFS analysis (PFS IA) to include all participants from Phase 2 and clarified results of the interim analysis.	PFS is the better predictor of final trial success Better use of the information collected to make a decision.
Revised some endpoint language for better clarity	Revised for clarity.
Updated tumor assessments from Week 49 to 52.	Better timing for procedure.
Amendment 2:	
Added PFS2 as an exploratory objective.	
Amendment 3:	
Synopsis Table 4-1 Objectives and Endpoints	Revised "Median duration of response (DOR)" to "DOR" to remove Median DOR as an endpoint. Added a separate analysis of GP5 in the FACT-M.
Synopsis Figure 5.1-1 Study Diagram 10.2.2. Efficacy Analyses 10.2.5 Interim Analyses	Revised text related to the interim and final analyses The study recently proceeded to phase 3 at the interim PFS analysis indicating that it met or exceeded a PFS HR threshold of 0.8. Assuming a statistically significant PFS result at PFS FA, added that there will be interim analyses of OS. The first interim analysis of OS (OS IA1) will be performed at the time of a statistically significant PFS result. If OS IA1 is not statistically significant, the second interim analysis of OS (OS IA2) will be performed when a minimum of 90% (270/300) of the expected deaths have occurred. If OS IA2 is not statistically significant, the final OS analysis will be performed when at least 300 deaths have occurred.
7.3 Blinding	Specified that the analysis team reporting to the Data Monitoring Committee (DMC) is external to BMS

Table 2.4-1: Summary of Changes to the Protocol

Description of Change	Brief Rationale
10.1 Sample Size Determination	Added text on non-proportional hazards for sample size justification for progression free survival (PFS) estimation
10.2.2. Efficacy Analyses	Revised the Phase 3 secondary endpoints hierarchy (1. Overall survival [OS] 2. Overall response rate [ORR]). In the event of a statistically significant PFS result, evaluating OS more formally would be considered more clinically meaningful than ORR. OS is considered a more relevant endpoint for evaluation of clinical benefit in the melanoma landscape compared to ORR. In study CA209-067 in previously untreated metastatic melanoma, overall survival at 5 years was remarkable at 52% in the nivolumab-plus-ipilimumab group and 44% in the nivolumab group, as compared with 26% in the ipilimumab group.1 In addition, in the event of a statistically significant PFS result, evaluating OS more formally would be considered more clinically meaningful than ORR.2 Therefore OS will be tested hierarchically at the time of statistically significant PFS followed by ORR (if OS is statistically significant and at the time that ORR data are mature).Therefore OS will be tested hierarchically at the time of positive PFS (if DMC states PFS is statistically significant) followed by ORR (if OS is positive and when ORR reaches maturity). Provided more detail on the Phase 2 Primary statistical methods Divided the "Testing Strategy" into separate sections for Phase 2 and Phase 3 Updated the Phase 3 Testing Strategy text to include the plan and conditions for analysis of OS Updated the Phase 2 Testing Strategy text to include when the sponsor would be unblinded in the case of a Phase 2 study

2.5 Data Monitoring Committee

A DMC will be established to provide oversight of safety and efficacy considerations and to provide advice to the Sponsor regarding actions the committee deems necessary for the continued protection of participants enrolled in this study. The DMC will be charged with assessing such actions in light of an acceptable benefit/risk profile for relatlimab and nivolumab. The DMC will act in an advisory capacity to BMS and will monitor participant safety and evaluate the available efficacy data for the study. The oncology therapeutic area of BMS has primary responsibility for design and conduct of the study.

The DMC will convene to assess the results of the interim analyses and provide its recommendations. Additional details concerning DMC oversight are provided in the DMC charter.

In addition, a BICR will be utilized in this study for determination of BICR-assessed endpoints. The BICR will review all available tumor assessment scans for all treated participants. Details of BICR responsibilities and procedures are specified in the BICR charter.

3 OBJECTIVES

Table 3-1 summarizes the study objectives and endpoints. Section 4 goes into more detail about the endpoint definitions.

Table 3-1: Study Objectives and Endpoints

Objectives	Endpoints
Phase 3 Primary Objective	
The Phase 3 primary objective is to compare PFS of BMS-986213 to nivolumab monotherapy in participants with previously untreated, unresectable, or metastatic melanoma.	PFS time as assessed by a Blinded Independent Central Review (BICR), using RECIST v1.1. PFS is defined as the time between the date of randomization and the first date of documented progression, or death due to any cause, whichever occurs first.
Phase 3 Secondary Objectives	
To compare OS of BMS-986213 to nivolumab monotherapy in participants with previously untreated, unresectable or metastatic melanoma.	OS is defined as the time between the date of randomization and the date of death due to any cause.
To compare ORR of BMS-986213 to nivolumab monotherapy in participants with unresectable or metastatic melanoma	ORR as assessed by a BICR. The ORR is defined as the number of subjects with a best overall response (BOR) of complete response (CR) or partial response (PR) divided by the number of randomized subjects for each treatment group. The BOR is defined as the best response designation, recorded between the date of randomization and the date of objectively documented progression per RECIST v1.1 or the date of subsequent anti-cancer therapy, whichever occurs first.
Phase 3 Tertiary/Exploratory Objective	
To evaluate duration of and time to objective response	Duration of response (DOR) is defined as the time between the date of first response to the date of first documented tumor progression (per RECIST v1.1) or death due to any cause. Time to objective response (TTR) is defined as the time from randomization to the date of the first documented CR or PR.
To evaluate PFS, ORR, DOR, and OS of BMS-986213 to nivolumab monotherapy in subgroups based on combinations of LAG-3 expression (≥ 1% expression versus < 1% expression) and PDL-1 status (≥ 1% tumor cell surface expression versus < 1% tumor cell surface expression) among participants with unresectable or metastatic melanoma treated with BMS-986213 compared to those treated with nivolumab monotherapy.	PFS time as assessed by a BICR using RECIST v1.1. ORR as assessed by a BICR. DOR. TTR. OS.
To evaluate the difference in symptom burden between participants with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy.	Time to meaningful symptomatic deterioration (TTSD) as measured by the Functional Assessment of Cancer Therapy – Melanoma (FACT-M) Melanoma Subscale (MS).

Table 3-1: Study Objectives and Endpoints

Objectives	Endpoints
To characterize the pharmacokinetics (PK) of relatlimab and nivolumab in subjects with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy.	PK parameters, influence of intrinsic and extrinsic covariates will be characterized using population PK models.
To characterize immunogenicity of relatlimab and nivolumab in subjects with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy.	Incidence of anti-drug-antibody (ADA) to relatlimab and nivolumab when administered in combination with nivolumab.
To explore potential exposure-response relationships in subjects treated with BMS- 986213.	Potential exposure response relationship (PD effect, efficacy, and select safety) will be evaluated using integrated analysis.
To explore potential biomarkers associated with clinical efficacy (ORR, PFS, and OS) by analyzing biomarker measures within the tumor microenvironment and periphery (e.g., tumor, serum and PBMCs) as related to clinical outcomes.	Association measures of select biomarkers and key efficacy endpoints (e.g., ORR, PFS, and OS)
To characterize changes in cancer-related symptoms and quality of life for participants with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy	Summary measures of FACT-M total and subscale scores (and changes) following treatment.
To characterize participant perceptions of the bothersomeness of symptomatic AEs, based on FACIT GP5 item as found in the FACT-M.	• Summary changes and frequency of responses in FACIT GP5 item measuring bother due to side effects of treatment.
To characterize changes in overall health status and health utility for participants with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy.	Summary measures of EQ-5D-3L visual analog scale (VAS) and utility index scores.
To characterize changes in work productivity and activity impairment for participants with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy	Summary measures of Work Productivity and Activity Impairment: General Health (WPAI:GH) work productivity loss, activity impairment, presentism, and absenteeism scores.
To evaluate PFS, PFS2, ORR, and DOR of BMS-986213 to nivolumab monotherapy in participants with previously untreated, unresectable, or metastatic melanoma	PFS and PFS2 time as assessed by investigator using RECIST v1.1. ORR as assessed by investigator. DOR.
To evaluate treatment-free interval (TFI) defined as time from last dose of study treatment to the start of subsequent systemic therapy or the last known date alive (for those who never received subsequent cancer therapy)	TFI
Phase 3 Safety Objective:	
To assess the overall safety and tolerability of BMS-986213 and of nivolumab monotherapy.	Rate of AEs, SAEs, AEs leading to discontinuation of treatment, deaths, and laboratory abnormalities.

Table 3-1: Study Objectives and Endpoints

Objectives	Endpoints
Phase 2 Primary Objective:	
The Phase 2 primary objective is to compare PFS of BMS-986213 to nivolumab monotherapy in participants with previously untreated, unresectable, or metastatic melanoma.	PFS time as assessed by BICR, using RECIST v1.1. PFS time is defined as the time between the date of randomization and the first date of documented progression, or death due to any cause, whichever occurs first.
Phase 2 Secondary Objectives:	
To estimate the treatment effect, measured by ORR, as determined by BICR using RECIST v1.1 in all-comers and in subgroups based on combinations of LAG-3 expression (≥ 1% expression versus < 1% expression) and PDL-1 status (≥ 1% tumor cell surface expression versus < 1% tumor cell surface expression) among participants with unresectable or metastatic melanoma treated with BMS-986213 compared to those treated with nivolumab monotherapy.	ORR as assessed by BICR.
To evaluate DOR and PFS rates at pre- specified time points (e.g., 24 weeks) based on BICR assessments using RECIST v1.1 in the randomized population (for DOR) and in subgroups based on combinations of LAG-3 expression (≥ 1% expression versus < 1% expression and PDL-1 status (≥ 1% tumor cell surface expression versus < 1% tumor cell surface expression) among subjects with unresectable or metastatic melanoma treated with BMS-986213 compared to those treated with nivolumab monotherapy	DOR. PFS time as assessed by BICR using RECIST v1.1.
To assess the 1- and 2-year OS rate in the randomized population and in subgroups based on combinations of LAG-3 expression (≥ 1% expression versus < 1% expression) and PDL-1 status (≥ 1% tumor cell surface expression versus < 1% tumor cell surface expression) among subjects with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy.	OS is defined as the time between the date of randomization and the date of death due to any cause.
Phase 2 Exploratory Objectives:	
To characterize the PK of relatlimab and nivolumab in subjects with unresectable or metastatic melanoma treated with BMS- 986213 and those treated with nivolumab monotherapy.	Summary measures of PK parameters based on population PK models including influence of intrinsic and extrinsic covariates
To characterize the immunogenicity of relatlimab and nivolumab in subjects with unresectable or metastatic melanoma treated with BMS-986213 and those treated with nivolumab monotherapy.	Incidence of ADA to relatlimab and ADA of nivolumab when administered in combination with relatlimab

Table 3-1: Study Objectives and Endpoints

Objectives	Endpoints
To assess the PD effects of BMS-986213 based on select biomarkers in the peripheral blood and tumor biopsy specimens.	Summary measures of change (or % change) from baseline in various biomarkers.
To characterize T-cell function during treatment of BMS-986213.	Summary of pre- treatment levels and of changes in T cells levels observed on- treatment.
To explore potential exposure-response relationships (e.g., with efficacy, receptor occupancy, pharmacodynamic effects) in subjects treated with BMS-986213.	Measures of Potential PK exposure response relationship (with pharmacodynamic effects, efficacy and select safety)
To evaluate ORR, DOR, and PFS and PFS2 rates at pre-specified time points (e.g., 24 weeks) based on investigator assessments using RECIST in the randomized population and in subgroups based on combinations of LAG-3 expression (≥ 1% expression versus < 1% expression) and PDL-1 status (≥ 1% tumor cell surface expression versus < 1% tumor cell surface expression) among subjects with unresectable or metastatic melanoma treated with BMS- 986213 and those treated with nivolumab monotherapy.	ORR. DOR. TTR. PFS and PFS2 time using RECIST v1.1.
Phase 2 Safety Objective:	
To assess safety and tolerability among participants with unresectable or metastatic melanoma treated with BMS-986213 compared to those treated with nivolumab monotherapy.	Rate of AEs, SAEs, AEs leading to discontinuation of treatment, deaths, and laboratory abnormalities.

4 ENDPOINTS

4.1 Efficacy

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

The first on-study tumor assessment is scheduled to be conducted at 12 weeks (± 1 week) following randomization. Subsequent tumor assessments are scheduled every 8 weeks (± 1 week) up to 52, then every 12 weeks (± 1 week) until disease progression.

4.1.1.1 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.

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 in Figure 4.1.1.1-1 and in Table 4.1.1.1-1.

Figure 4.1.1.1-1: PFS Primary Definition

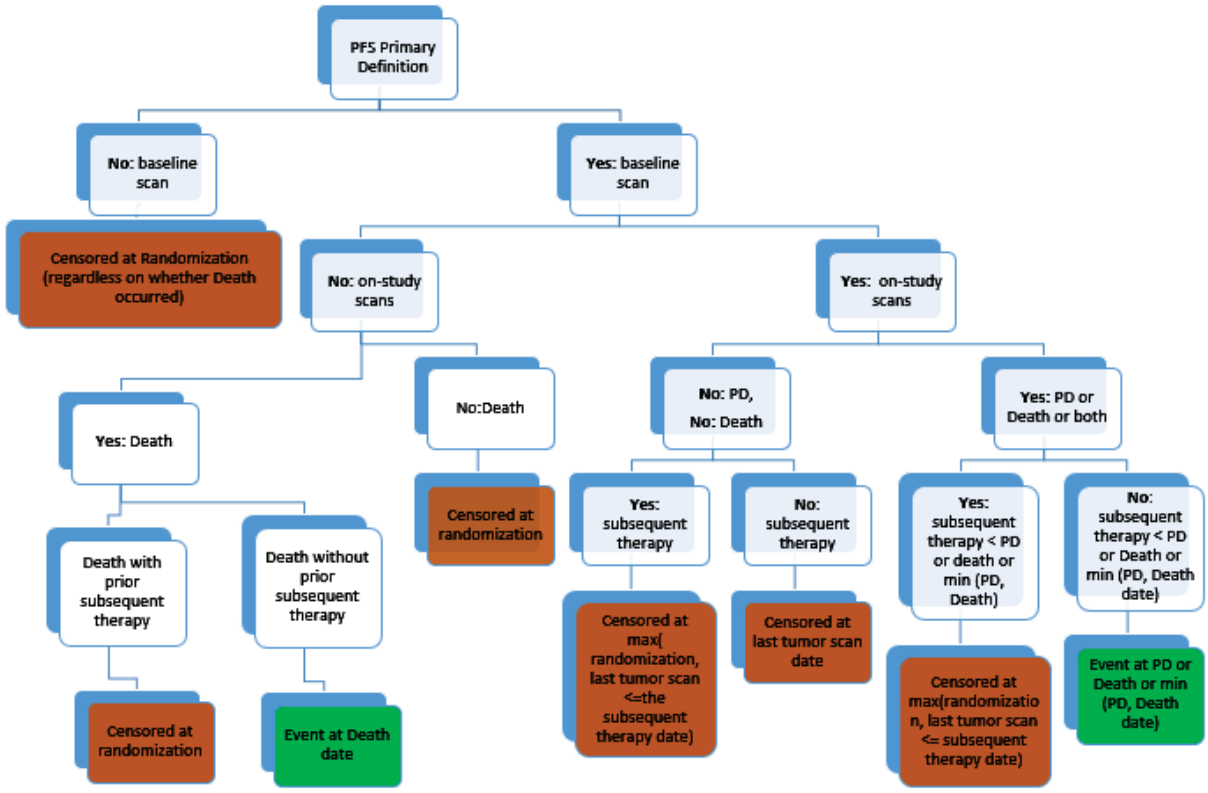


Table 4.1.1.1-1: Censoring Scheme used in Primary Definition of PFS

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. Note: subjects with measurable disease per Investigator but no measurable disease per BICR will be included in the analysis.

4.1.1.2 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.

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 Figure 4.1.1.2-1 and in Table 4.1.1.2-1.

Figure 4.1.1.2-1: PFS Secondary Definition

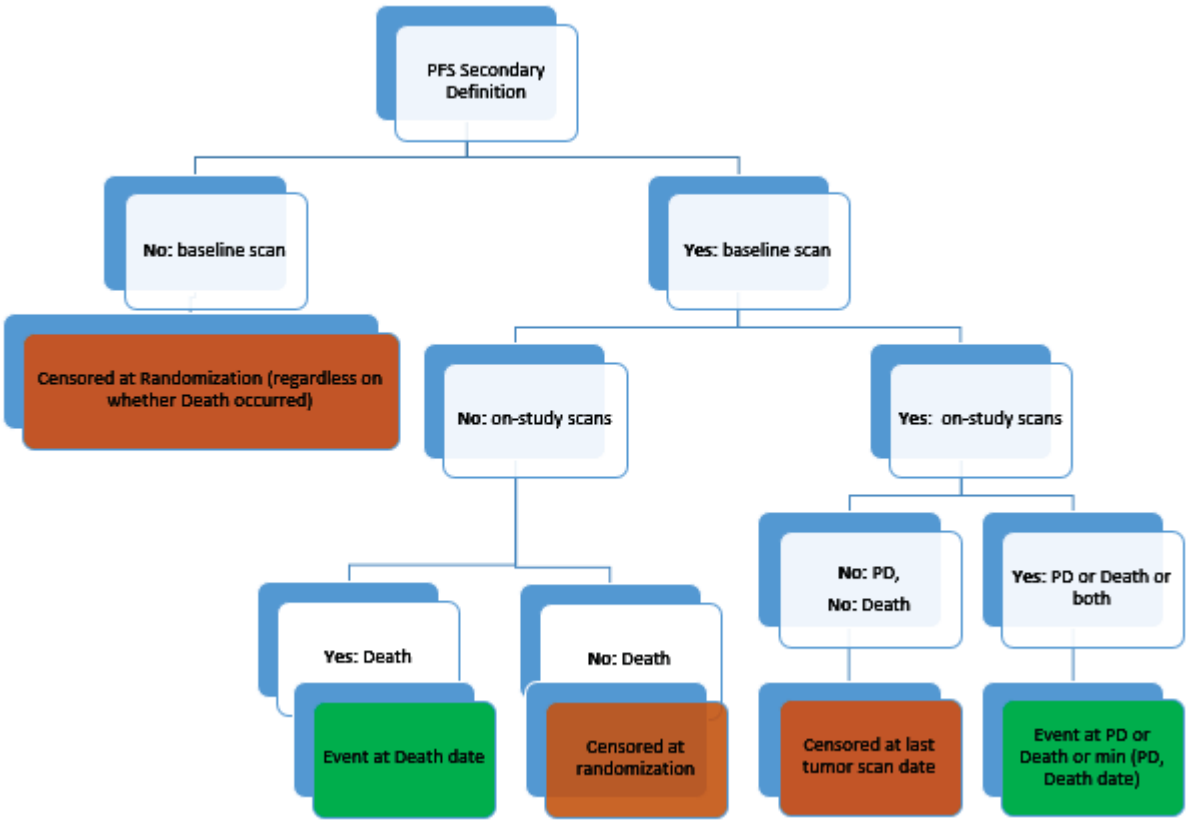


Table 4.1.1.2-1: Censoring Scheme for Secondary definition of PFS

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

Note: subjects with measurable disease per Investigator but no measurable disease per BICR will be included in the analysis.

4.1.2 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"). Overall survival will be censored for subjects at the date of randomization if they were randomized but had no follow-up.

The first survival follow-up will be conducted 3 months (±14 Days) after subject's FU 2 visit, subsequent survival FU visits every 3 months (±14 days).

4.1.3 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 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.

ORR will be summarized using a binomial response rate. The Clopper-Pearson method will be used to estimate the two-sided 95% confidence interval. BOR will be summarized by response category. Summary statistics of time to objective response will be provided for subjects who achieve a BOR of PR or CR, as assessed by the BICR.

Objective Response Rate as assessed by investigator will be summarized in the same manner as per BICR.

4.1.4 Time to and Duration of Response

Time to Response (TTR) 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.

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 (CR or PR) only.

Duration of Response and time to response as assessed by investigator will be summarized in the same manner as per BICR.

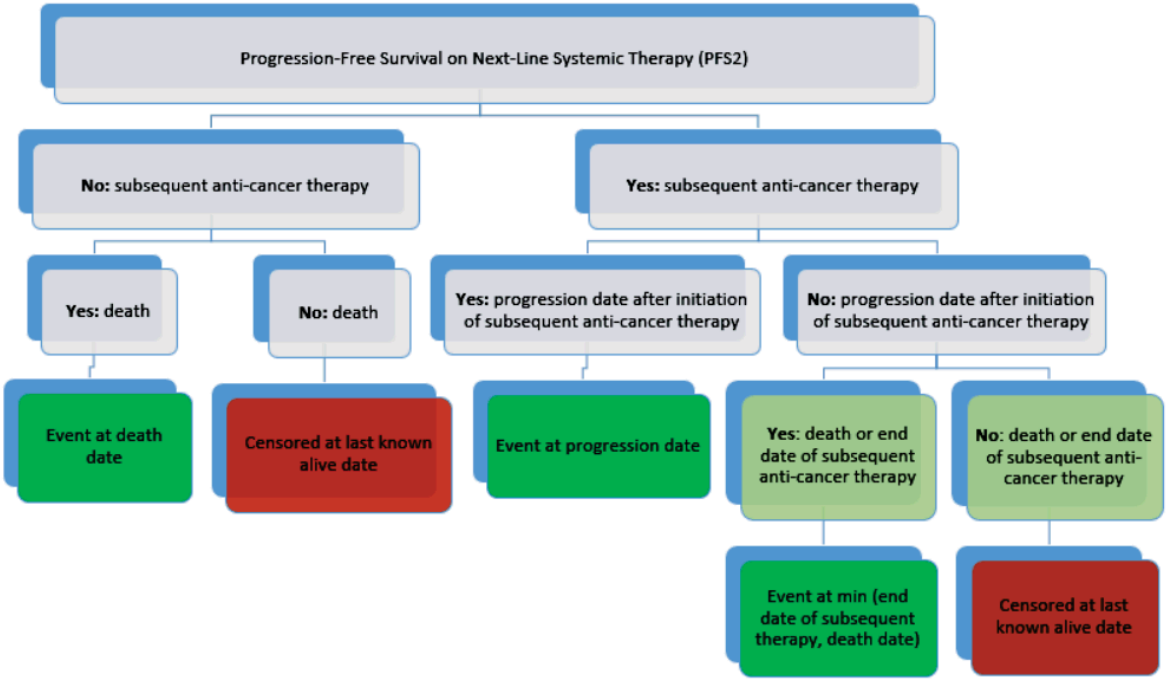
4.1.5 PFS2

PFS on next-line therapy (PFS2) is defined as the time from randomization to objectively documented progression date after the next line of therapy, per investigator assessment, 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:

- 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 his last known alive date.
- 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 discontinued subsequent anti-cancer therapy, the date of min (death, discontinuation of subsequent anti-cancer therapy) is the event date;
  - Else the subject's PFS2 is censored at his last known alive date.

Figure 4.1.5-1: PFS2 Definition



4.1.6 TFI

Treatment-Free Interval (TFI) and Treatment-Free Survival (TFS) are defined only in randomized participants who are off study treatment and continued to be followed in study prior to initiation of subsequent systemic therapy. TFI is defined as time from end of study therapy until last known alive date in those who never received subsequent systemic therapy, and defined as time from end of study therapy until subsequent systemic therapy in those who received subsequent systemic therapy. For TFS, event is defined as receiving subsequent therapy or death, whichever comes first. For TFS, time is TFI and event is receiving subsequent therapy.

4.2 Safety endpoints

The assessment of safety will be based on the incidence of adverse events (AEs), serious adverse events (SAEs), adverse events leading to discontinuation, adverse events leading to dose modification, select adverse events (select AEs) for EU 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 Other Endpoints

4.3.1 Pharmacokinetics

Relatlimab and nivolumab concentration-time data at scheduled trough (Ctrough) and end-of-infusion timepoints will be evaluated. Measurements will be collected on treatment (up to 12

cycles) and for up to second post-treatment follow-up. These data may also be pooled with other dataset for population PK and exposure-response analysis which will be presented in a separate report.

4.3.2 Biomarkers

Biomarker endpoints from peripheral blood will generally be measured at multiple timepoints, and evaluated as both predictive and pharmacodynamic markers in the context of the exploratory biomarker objectives. These may include measures such as levels and change from baseline in levels of the following at each scheduled timepoint:

- Serum soluble factors including cytokines and soluble LAG3
- The proportion of specific lymphocyte subsets/expression levels of T cell co-stimulatory/inhibitory markers assessed using flow cytometry

Biomarker endpoints from tumor biopsies will be explored predominantly in an effort to identify baseline markers predictive of efficacy, since they are only measured at baseline for most subjects.

For the subset of subjects who have both pre-treatment and on-treatment biopsies, pharmacodynamic associations may be explored. Endpoints may include measures such as pre-treatment levels and change in levels observed on-treatment of:

- Expression of selected genes and gene signatures
- IHC assessment of the presence/absence and intensity of T-cell infiltration and expression of MHCI, MHC class II, CD3 and CD8
- Tumor mutation burden

Appropriate functional transformation of these exploratory measures will be applied as necessary. Measures not available at the time of the clinical study report may be summarized in a separate report.

4.3.2.1 PD-L1 and LAG-3

PD-L1 expression is defined as the percent of tumor cells demonstrating plasma membrane PD-L1 staining of any intensity using an immunohistochemistry (IHC) assay. LAG-3 expression is defined as the percent of nucleated cells demonstrating LAG-3 staining of any intensity using an immunohistochemistry (IHC) assay.

Additional biomarkers potentially associated with clinical endpoints will be measured by analyzing tumor and blood samples. Biomarker endpoints include, but are not limited to, single-nucleotide polymorphisms (SNPs), proteins in tumor specimens and serum, and immune cell populations.

PD-L1 expression and LAG-3 expression will be collected in the IRT as well as in the clinical database. Please see Section 6.2.1 for more details on which values will be used for the analyses.

4.3.3 Immunogenicity

Incidence of ADA to either relatlimab or nivolumab will be assessed during treatment and for up to second follow-up post treatment. Baseline ADA positive subject is defined as a subject with positive seroconversion detected in the last sample before initiation of treatment. ADA-positive subject is a subject with at least 1 ADA-positive sample relative to baseline after initiation of the treatment.

At the sample level, individual samples will be characterized as ADA-positive or ADA-negative. A subject will be considered to have a positive sample at baseline if the last sample prior to the initiation of treatment is ADA-positive. To examine the potential relationship between immunogenicity and safety, a table summarizing the frequency and type of AEs of special interest may be explored by immunogenicity status.

4.3.3.1 ADA Status of a Sample

- 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 (1) than baseline positive titer
- ADA-Negative Sample: After initiation of treatment, ADA not positive sample relative to baseline

4.3.3.2 ADA Status of a Subject

- Baseline ADA-Positive Subject: A subject with baseline ADA-positive sample
- ADA-Positive Subject: A subject with at least one ADA-positive sample at any time after initiation of treatment
- *Persistent Positive (PP)*: ADA-positive sample at 2 or more consecutive timepoints where the first and last ADA-positive samples are at least 16 weeks apart

Not PP - *Last Sample Positive*: Not persistent positive with ADA-positive sample in the last sampling timepoint

- *Other Positive*: Not persistent positive but some ADA-positive samples with the last sampling being negative

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

4.3.4 Outcomes Research

4.3.4.1 EuroQoL EQ-5D-3L

Subjects' overall health status will be assessed using the EuroQoL Group's self-reported health status measure (EQ-5D-3L)[1]. EQ-5D-3L essentially has 2 components: the EQ-5D-3L descriptive system and the EQ visual analogue scale (EQ-VAS).

The EQ-5D-3L descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 3 levels: no problems, some problems, and severe problems. Once the data have been collected and a database created, a scoring function can be used to assign a value (i.e., EQ-5D-3L utility index score) to self-reported health states from a set of population-based preference weights. For this study, EQ-5D-3L utility index values will be computed using a scoring algorithm based on the UK MVH-A1 TTO value set.

The EQ-VAS records the subject's self-rated health state on a 101-point visual analogue scale (0= worst imaginable health state; 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 for the EQ-5D-3L.[2]

4.3.4.2 FACT-M

The FACT-M questionnaire[3] will be used to assess the effects of disease symptoms on functioning and well-being. The FACT-M includes the 27-items from the 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). In addition, the FACT-M includes a 16-item disease-specific melanoma subscale (MS) and an 8-item melanoma surgery scale (MSS). The GP5 item, which will also be analyzed on its own is: "I am bothered by side effects of treatment".

Each FACT-M item is rated on a 5-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 MS scores can be combined to produce a total score for the FACT-M, 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, can be derived by summing the PWB, FWB, and MS scores. All scores are scaled so that higher values indicate better functioning as well as lower symptom burden.

For any given scale, the minimally important difference represents the smallest difference in score that informed patients might perceive as important and that would warrant consideration of a change in management. With regard to the MS, the minimally important difference has been estimated as a change of 2-4 points.[4] For this trial, we will use the midpoint of the range (i.e., a change ≥ 3 points) to define a minimally important change in MS score. This interpretation has been applied in previous investigations.[5]

Data will be scored according to the algorithm described in the FACT-M scoring manual.

If there are missing items, subscale scores can be prorated. This can be done using the following formula:

> **Prorated subscale score =** [Sum of item scores] **x** [N of items in subscale] ÷ [N of items answered]

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the items were answered. The total score is then calculated as the sum of the un-weighted subscale scores. The FACT scale is considered to be an acceptable indicator of patient quality of life as long as overall item response rate is greater than 80% (e.g., at least 22 of 27 FACT-G items completed). This is not to be confused with individual subscale item response rate, which allows a subscale score to be prorated for missing items if greater than 50% of items are answered. In addition, a total score should only be calculated if ALL of the component subscales have valid scores.

4.3.4.3 WPAI:GH

The WPAI:GH is a 6-item measure of impairment in work productivity and usual activities. Responses to the questionnaire's items are used to derive a score measuring the amount of absenteeism (work time missed), presentism (impairment at work/reduced on-the-job effectiveness), work productivity loss (overall work impairment/absenteeism plus presentism), and daily activity impairment attributable to general health. Outcomes are expressed as impairment percentages with higher numbers indicating greater impairment or less productivity.

Data will be scored according to the algorithm as described in the WPAI:GH scoring manual. Questions:

1 = currently employed

2 = hours missed due to health problems

3 = hours missed other reasons

4 = hours actually worked

5 = degree health affected productivity while working

6 = degree health affected regular activities

Multiply scores by 100 to express in percentages.

- Percent work time missed due to health: Q2/(Q2+Q4)
- Percent impairment while working due to health: Q5/10
- Percent overall work impairment due to health: Q2/(Q2+Q4)+[(1-Q2/(Q2+Q4))x(Q5/10)]
- Percent activity impairment due to health: Q6/10

5 SAMPLE SIZE AND POWER

The sample size for the study is based on a primary endpoint of PFS using BICR for either a Phase 2 or a Phase 3 study. The sample size for the Phase 2 study as justified below is approximately 400 randomized participants. The sample size for the Phase 3 study as justified below is 700 randomized participants, comprised of the Phase 2 population and an additional 300 participants.

The Phase 2 portion of the study serves multiple purposes. First, it enables the decision whether there is sufficient signal of efficacy in an all-comers population to proceed seamlessly to Phase 3. This PFS IA will be performed on the Phase 2 population when at least 150 events have been observed or all Phase 2 participants have been followed for at least 12 weeks. Second, if the PFS IA does not indicate moving to Phase 3 in an all-comers population, the final Phase 2 analyses will include the analyses of biomarker subgroups. Third, if the PFS IA indicates enrollment should be re-opened, the Phase 2 population will be part of the subsequent Phase 3 analyses.

The overall alpha for the Phase 3 study is 0.05 (two-sided). An administrative alpha penalty of 0.001 will be used for the PFS IA, although there is no plan to stop for efficacy at that analysis. If PFS IA meets the pre-specified threshold (HR $\leq$ 0.8) and the study proceeds to Phase 3, the final PFS analysis (PFS FA) will be performed when at least 365 PFS events have occurred per BICR and will use all remaining unspent alpha (0.049).

Note: on 26AUG2019 the study proceeded to phase 3 following the interim PFS analysis. This indicated that the HR was <= 0.8.

5.1 Sample Size Justification for Phase 2 PFS Estimation

The primary objective of the Phase 2 portion of the study is to demonstrate preliminary clinical evidence of the treatment effect, measured by PFS, as determined by BICR using RECIST v1.1 in randomized participants with unresectable or metastatic melanoma treated with BMS-986213 compared to those treated with nivolumab monotherapy. The number of events is based on results from study CA209067 with a median PFS of 6.9 months for nivolumab monotherapy and 11.8 months for nivolumab with relatlimab. With 183 events, there will be approximately 80% power to detect a hazard ratio (HR) = .73 with a type 1 error of 0.1 (1-sided). The cure rates were assumed to be 30% in the nivolumab arm and 40% in the nivolumab with relatlimab arm. The power is also affected by non-proportional hazards, since it is driven by the number of events, not the number of study participants, and some fraction of the participants in each arm are assumed will remain event-free for the duration of the study. Approximately 400 participants will be randomized to the Phase 2 study to ensure enough power to analyze the biomarker subgroups (see section below on the secondary endpoint of ORR in subgroup).

An interim analysis of PFS (PFS IA) will be performed when a minimum follow-up of 12 weeks is achieved for approximately 400 randomized participants or at least 150 PFS events have been observed using BICR. This analysis will include all available data at the time and will be performed by the DMC to maintain blinding of the sponsor. If the recommendation of the DMC is to proceed to the Phase 3 portion of the study then the study will start to re-enroll and the final PFS analysis for Phase 2 will not be performed.

Based on an anticipated screen failure rate of 30%, approximately 575 participants will need to be screened in order to randomize approximately 400 participants who fit the eligibility criteria.

5.1.1 Phase 2 Testing Strategy

If the interim PFS indicates the study will stop at Phase 2 (PFS HR > 0.8), the study will not enroll any additional participants, but members of the immediate study team (site-facing BMS team members) will remain blinded until the data are sufficiently mature (183 PFS events per BICR and a minimum 6 months of follow-up). The final Phase 2 analysis includes multiple subgroups of interest. The Benjamini-Hochberg method will be used to account for multiplicity by controlling the False Discovery Rate (FDR) when making comparisons in multiple subgroups.

5.1.2 Sample Size Justification for Secondary Objective of Subgroup Comparison in Phase 2 Trial

If the PFS IA does not support continuing to a Phase 3 study, the study will complete as a Phase 2 trial. The primary objective of the Phase 2 study using patients with previously untreated, unresectable or metastatic melanoma is to assess preliminary clinical evidence of the treatment effect measured by PFS as determined by BICR that may represent substantial improvement over available therapies. As a secondary objective, the Phase 2 will examine ORR information in biomarker subgroups based on LAG-3 expression (≥ 1% expression versus < 1% expression) and PDL-1 status (≥ 1% tumor cell surface expression versus < 1% tumor cell surface expression/non-quantifiable). ORR information and biomarker prevalence rates (for LAG-3 positive and negative expression, and for PDL-1 positive and negative status) used in the sample size calculations were based on information from study CA209067 and interim information from study CA224020. Table 5.1.2-1 shows the power for varying ORRs and varying prevalence rates for subgroups, assuming a Type 1 error rate of 0.1 (one-sided) and an overall N of 400 subjects. The Benjamini-Hochberg procedure will be used to control the False Discovery Rate (FDR) when testing multiple subgroups.

Table 5.1.2-1: Power for Subgroups with Different N and Prevalence

Overall N for Phase 2 Study	Subgroup N	Subgroup ORR	Subgroup Prevalence	Power (one-sided alpha .1 and 20% delta)
400	200	0.7	0.5	0.991
	140	0.7	0.35	0.962
	50	0.5	0.125	0.576
	80	0.5	0.2	0.72
	50	0.33	0.125	0.57
	80	0.33	0.2	0.713
	100	0.23	0.25	0.815
	120	.23	0.3	0.865

5.2 Sample Size Justification for Phase 3 PFS Comparison

The sample size is calculated in order to compare PFS among participants randomized to receive BMS-986213 versus nivolumab. The number of events required is simulated based on results from study CA209067 with a median PFS of 6.9 months for nivolumab monotherapy and 11.8 months for nivolumab with relatlimab, incorporating 35% of subjects with durable response in the combined groups, and a piecewise hazard ratio resulting in an effective hazard ratio of approximately 0.73.

Based on these assumptions, the study requires at least 365 PFS events to ensure approximately 85% power to detect a hazard ratio of 0.73 with an overall type I error of 0.05. Approximately 700 participants will be randomized to the two treatment arms in a 1:1 ratio. The final PFS analysis is planned to occur when 365 participants have had a PFS event. Total enrollment will take approximately 27 months (including a potential 6-month pause due to the PFS IA).

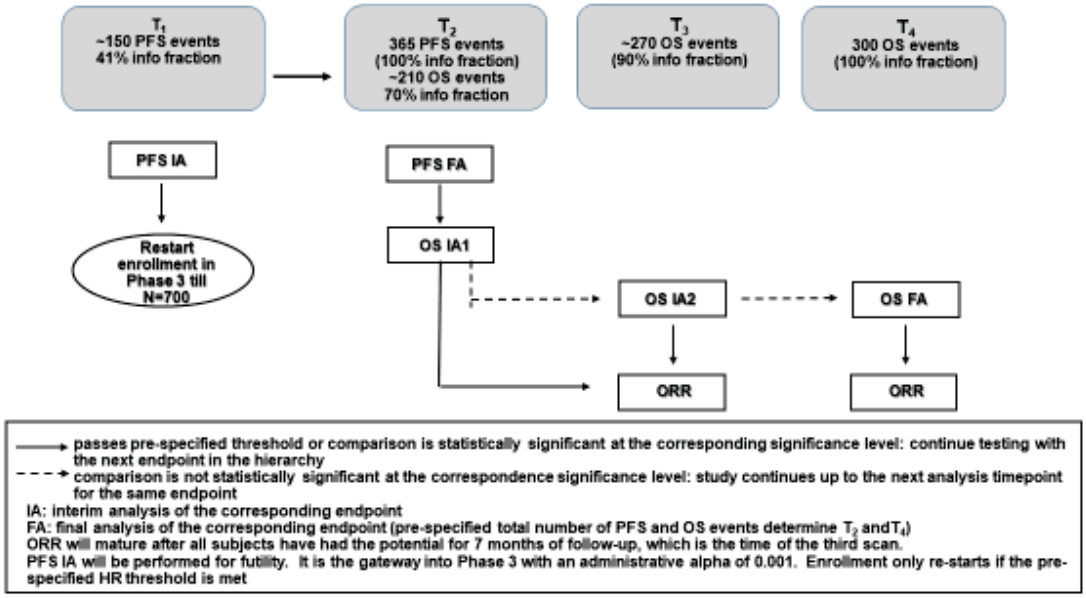
Based on an anticipated screen failure rate of 30%, approximately 1000 participants will need to be screened in order to randomize 700 participants who meet the eligibility criteria.

5.2.1 Phase 3 Testing Strategy

The overall alpha for the Phase 3 study is 0.05 (two-sided).

For the Phase 3 study, type I error control across endpoints will be performed hierarchically (see Figure 5.2.1-1). The primary analysis is for PFS. If that analysis is significantly superior, then the secondary endpoints may be tested in the order of OS followed by ORR. That is, if the results comparing PFS between treatment groups are significant at the applicable alpha level, then results comparing OS between treatment groups will be interpreted. If the results comparing OS between treatment groups are significant, then results comparing ORR between treatment groups will be interpreted. Other endpoints will not be tested formally.

Figure 5.2.1-1: Phase 3 Hierarchical Procedure with Group Sequential Testing in All Randomized Subjects



Approximately 700 subjects will be randomized in 1:1 ratio to BMS-986213 or nivolumab monotherapy arm respectively.

- Approximately 365 PFS events in all randomized subjects will provide around 85% power to detect an average HR of 0.73 with an overall type I error of 5.0% (two-sided). The target hazard ratio was obtained via simulation using above modeling.

Table 5.2.1-1 summarizes the analysis of PFS in Phase 3. The interim PFS analysis (PFS IA) will be performed when a minimum follow-up of 12 weeks is achieved for approximately 400 randomized subjects or at least 150 PFS events have been observed per BICR. PFS IA will determine whether the study stops at Phase 2 (PFS HR > 0.8) or continues as Phase 3 (PFS HR ≤ 0.8). An administrative alpha penalty of 0.001 will be used for the PFS IA, although there is no plan to stop for efficacy at that analysis.

If PFS IA meets the pre-specified threshold and the study proceeds to Phase 3, the final PFS analysis will be performed when at least 365 PFS events have occurred per BICR and will use all remaining unspent alpha.

Table 5.2.1-1: Schedule of Analyses of Primary Endpoint

	Interim PFS Analysis (PFS IA)	Final PFS Analysis (PFS FA)
Population	All randomized subjects	
Conditions	When 150 events of PFS are observed or all Phase 2 randomized subjects have been followed for at least 12 weeks (time of 1st scan)	When 365 events of PFS are observed
Expected timing	~15 months	~34 months
Nominal significance level	.001	0.049
Critical Hazard Ratio	HR $\leq$ 0.8*	HR $\leq$ 0.81

*Not formally tested (administrative alpha); had to pass this boundary for the study to continue to Phase 3.

Analysis of the Key Secondary Endpoint of OS

The first key secondary endpoint of OS will be formally compared using a hierarchical procedure using Lan-DeMets alpha spending function with the O'Brien-Fleming type of boundary in East v6 for each endpoint. This testing procedure is summarized in Figure 5.2.1-1. OS will only be formally compared if the PFS FA analysis (primary endpoint) is statistically significant. This comparison could be either:

a) at PFS FA (T_2) that will occur when at least 365 (100%) PFS events are observed. We are estimating that approximately 210 (70%) of OS events will have occurred

b) after 270 (90%) OS events have been observed (T_3)

c) after 300 (100%) OS events have been observed (T_4).

The timing of OS IA2 is based on an estimate that it will be halfway between the first interim analysis of OS and the final OS analysis. The second key secondary endpoint of ORR will be formally compared using a hierarchical procedure. If the PFS (primary endpoint) and OS (first key secondary endpoint) analyses are statistically significant, ORR will be tested when it has matured. ORR will not be considered mature until all subjects have had the potential for 7 months of follow-up, which is the time of the third on-study scan.

Table 5.2.1-2: Schedule of Analyses of OS, First Key Secondary Endpoint

	Interim OS Analysis #1 (OS IA1)	Interim OS Analysis #2 (OS IA2)	Final OS Analysis OS FA
Population	All randomized subjects		
Assumed medians	Nivo mono arm: 36.9 months; BMS-986213 arm: 49.2 months		
Conditions*	At PFS FA (T_2) when ~ 210 (70%) OS events may be observed	After 270 (90%) OS events have been observed (T_3)	After 300 (100%) OS events have been observed (T_4)
Expected timing	~34 months (T_2)	~42 months (T_3)	~50 months (T_4)
Nominal significance level (2-sided) - OS efficacy boundary	0.014	0.031	0.038
Cumulative alpha spent	0.014	0.036	0.049
Critical Hazard Ratio	HR ≤ 0. 713	HR ≤ 0. 0.769	HR ≤ 0.787
Cumulative Power	~36%	~60%	`69%

* Primary condition for OS analysis is that the primary PFS endpoint analysis must be statistically significant before any formal comparison of OS.

Note: interim alpha spending and OS efficacy boundary will be determined based on the actual number of events/deaths at the time of database lock.

04AUG2021 Update to SAP: The primary endpoint analysis was tested on 09MAR2021 and it was statistically significant. The first interim analysis of OS (OS IA1) was performed at the same time by the DMC and it was not statistically significant. As pre-specified, the significant results for the primary endpoint led to unblinding at the subject level. OS and ORR, however, were not analyzed, nor summarized by treatment arm. The positive results of the PFS analysis led to submission of the data to health authorities. On 04AUG2021, the decision was made to unblind response data at the time of the second interim OS analysis (OS IA2), anticipated for OCT2021. ORR will be unblinded and reported descriptively until (and if) OS is statistically significant and then it will be tested according to the hierarchy.

25OCT2021 Update to SAP:

With a target of 270 deaths for the second interim analysis of OS (OS IA2), BMS observed 271 deaths in the blinded database on 30-AUG-2021. Consequently, BMS began preparing for a database lock planned to occur 19-OCT-2021, based on a standard 6-week timeline to allow time for final data cleaning, image reconciliation, and a survival sweep. During the course of that survival sweep, 20 additional deaths were entered into the database, resulting in 291 observed deaths as of 18-OCT-2021. Given that the target number of deaths for the final OS analysis (OS FA) is 300, BMS believes that it is impractical and unnecessary to perform an interim and final analysis so close together. The second interim analysis of OS will no longer be performed. We will proceed directly to the final OS analysis and all deaths in the database will be used in the analysis. As originally planned for the final OS analysis BMS, and not the DMC, will analyze the OS data and be unblinded to OS results.

Table 5.2.1-3 gives details on the alpha and critical hazard ratio if no second interim analysis is performed and there are 300 deaths at the time of lock. However, these will be calculated based on the actual data in the database at the time of lock.

Table 5.2.1-3: Schedule of Analyses of OS in the Event that Second Interim Analysis Is Not Performed.

	Interim OS Analysis #1 (OS IA1)	**Final OS Analysis OS FA**
Population	All randomized subjects	
Assumed medians	Nivo mono arm: 36.9 months; BMS-986213 arm: 49.2 months	
Conditions*	At PFS FA (T_2) 227 (75.7%) OS events were observed	Approximately 300 (100%) OS events have been observed
Nominal significance level (2-sided) OS efficacy boundary	0.019	0.043
Cumulative alpha spent	0.019	0.049
Critical Hazard Ratio	HR ≤ 0. 733	HR ≤ 0.792
Cumulative Power	~43%	`69%

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

Each treated subject will go through two study periods: baseline period and post baseline period.

- 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 on the same date and time of the first dose of study treatment will be considered as baseline evaluations.
  - 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, the assessment that is closest to day (and time if collected) of the first dose of study treatment will be used as the baseline in the analyses. If multiple assessments are collected on the same date (and time, if collected), the last assessment will be considered as baseline.

- Post baseline period:
  - On-treatment AEs will be defined as AEs with an onset date and time on or after the date and time of the first dose of study treatment (or with an onset date on or after the day of first dose of study treatment if time is not collected or is missing). For subjects who are off study treatment, AEs will be included if event occurred within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of study treatment. No "subtracting rule" will be applied when an AE occurs both pre-treatment and post-treatment with the same preferred term and grade.
  - On-treatment evaluations (laboratory tests, 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 30 days (or 100 days depending on the analysis) after the last dose of study treatment.
  - If more than one tumor biopsy specimen is available, baseline PD-L1 and LAG-3 expression will be determined from the most recently collected specimen (prior to first dose of study treatment) with a measurable result

6.2 Treatment Regimens

The treatment group "as randomized" corresponds to the treatment group assigned by the IRT system.

- Arm A: Experimental arm: BMS-986213
- Arm B: Control arm: nivolumab

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".

6.2.1 Stratification

Subjects were stratified based on 4 different variables:

- PD-L1 Expression: PD-L1>=1% tumor cell surface expression / PD-L1<1% tumor cell surface expression (entered by vendor in IRT)
- LAG-3 Status: LAG-3>=1% expression / LAG-3<1% expression (entered by vendor in IRT)
- BRAF Status: Mutation Positive/ Mutation Wild-type at screening
- AJCC M Stage: M0/M1any[0] / M1any[1]) at screening

During the monitoring of prospectively-assessed PD-L1status, results revealed a lower rate of positivity (30-40%) than observed in historical studies of melanoma subjects (55-65%). Details regarding the initial scoring of PD-L1 at randomization vs. the historical rates of positivity

in melanoma studies can be found in the electronic protocol master file. The observed difference was hypothesized to be due to the testing site's pathologists as all previous data was generated at [redacted] and the study CA224047 data was generated at [redacted].

To address the impact on efficacy assessment within the PD-L1/LAG-3-defined biomarker subsets included as secondary endpoints for Phase 2, [redacted] will enable re-scoring of PD-L1 by pathologists from a second testing site [redacted] for any randomized subjects who had been evaluated for PD-L1 at the [redacted] site. The re-scored PD-L1 results will be used in the primary efficacy analyses. If a rescore value is not applicable, indeterminate, or missing, it will be grouped together withPD-L1<1% and categorized as PD-L1<1% /non-quantifiable tumor cell surface expression if the number accounts for < 5% of the scores among all randomized subjects. If the number of non-quantifiable PD-L1 scores accounts for >=5% of the scores among all randomized subjects, they should be placed in a third category of PDL1 NON-QUANTIFIABLE and analyzed as such.

The evaluations of PD-L1 status were changed to [redacted] site at the start of Phase 3 so all subjects randomized during Phase 3 were scored via [redacted] pathologists at the [redacted] site. Thus, the PD-L1 values within the IRT system for all Phase 3 subjects will be used for the primary efficacy analyses. In addition, two subjects who transferred into the study during Phase 2 will also use the PD-L1 value placed in IRT. A sensitivity analysis will be performed using the IRT PD-L1 original scores for all subjects to assess the impact of the discrepancies.

During the monitoring of M-stage stratification, results showed that sites were not including LDH lab values into their classification. For this reason, the analyses will use the RAVE database for the M-stage stratification factor.

6.3 Populations for Analyses

For purposes of analysis, the following populations are defined:

Population	Description
Enrolled Participants	All participants who sign informed consent and were registered into IRT
Randomized	All participants who are randomized to any treatment group.
Treated	All participants who received at least one dose of double-blind study medication.
Safety	All randomized participants who take at least 1 dose of double blind study treatment. Data in this data set will be analyzed based on randomized treatment, except in the following cases: • If a participant received the same incorrect treatment throughout the study, then the participant will be analyzed based on the treatment received. • If a participant received study drug from more than one treatment group, and none of the administrations were

Population	Description
	consistent with the assigned randomized treatment group, then the participant will be analyzed based on the first treatment received.
PK	All randomized participants with available serum time-concentration data.
Immunogenicity	All randomized participants with available ADA data.
Biomarker	All randomized participants with available biomarker data.
Phase 2	All participants randomized prior to 26AUG2019, date of DMC recommendation to proceed to Phase 3.
Phase 3	All participants randomized after 26AUG2019, date of DMC recommendation to proceed to Phase 3.

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

Laboratory results, adverse events, and other symptoms will be graded using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE), Version 5.0, 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.

[redacted]

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". For studies for which some of the study medication is an oral dose, adverse event that led to dose delay of the oral drug (similarly defined as dose omission or dose interruption) will be coded with action "Drug was interrupted".

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[7] in order to collapse adverse event records into unique records based on the preferred term. These data will be presented as the rate per 100 person-years of exposure. These analyses will take into account all on-treatment events (allowing more than 1 event per subject) and the total exposure time. The person-year exposure will be computed as the sum over the subjects' exposure expressed in years where the exposure time is defined as

- (Date of last dose of study treatment - date of first dose of study treatment +31 days (or 101 days, depending on the analysis))/365.25, for subject who are off study treatment and were

followed for at least 30 days (or 100 days, depending on the analysis) after last dose of study treatment.

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

7.1.1.1 Select Adverse Events (EU Submission)

The select Adverse Events (select AEs) consist of a list of preferred terms grouped by specific category (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/sub-categories.

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 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 (e.g., Myositis/Rhabdomyolysis Event, Myocarditis Event, Demyelination Event, Guillain-Barre Syndrome, Pancreatitis Event, Uveitis Event, Encephalitis Event, Myasthenic Syndrome, Graft Versus Host Disease, Troponin Elevation, and Other Meningitis). 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. Immune-mediated AEs are specific events (or groups of PTs describing specific events) that include pneumonitis, diarrhea/colitis, hepatitis, nephritis/renal dysfunction, rash, endocrine (adrenal insufficiency, hypothyroidism/thyroiditis, hypothyroidism, thyroiditis, hyperthyroidism, diabetes mellitus, and hypophysitis), and other specific events, considered as potential immune-mediated events by investigator, that meet the definition summarized below:

- those occurring within 100 days of the last dose,

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

The list of MedDRA preferred terms used to identify Immune-Mediated adverse events 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 for US Submission.

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 sub-category and laboratory test code sequence number.

7.1.3 Immunogenicity data

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

7.2 Study Conduct

7.2.1 Relevant Protocol Deviations

The following programmable deviations will be considered as relevant protocol deviations and summarized and listed based on the all randomized subjects, by treatment group and overall. Non-programmable relevant eligibility and on-treatment protocol deviations, as well as significant (both programmable and non-programmable) eligibility and on-treatment protocol deviations will be reported through ClinSIGHT listings.

Eligibility:

- Wrong Cancer Diagnosis: no histologically confirmed Stage III (unresectable) or Stage IV melanoma, per the AJCC staging system
- Baseline image performed outside required window per RECIST 1.1 and per protocol (beyond 28 days + 2 week window)

Inadequate Performance Status at Baseline:

- ECOG > 1 for adults, Lansky performance < to 80% for adolescents
- Verify Target Population: Prohibited Prior Therapies/Required Prior Therapies not Received No prior systemic anticancer therapy given as primary therapy for unresectable or metastatic disease (except anti PD-1, anti CTLA-4, BRAF- or MEK- inhibitor-containing regimens as adjuvant or neoadjuvant with at least 6 months between the last dose and the recurrence and interferron with at least 6 weeks prior to randomization). No Prior Treatment with Relatlimab or any other LAG-3 agents
- LAG-3, PD-L1, or BRAF mutation status unevaluable, indeterminate or not available prior to randomization. Note PD-L1 values should be based off of the IRT information and not the re-scored value. PD-L1 was re-scored posthoc and should not be considered an ineligibility if the value is non quantifiable.

On-study:

- 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 by-subject listing of batch numbers will be also 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) 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 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 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 subject listing for all enrolled subjects will also be provided, showing whether the subject was randomized/ along with the reason for not being randomized.

Enrollment by country and site, and enrollment by month will be summarized for all enrolled subjects. Randomization date, first dosing date, country, investigational site will be presented in a by subject listing of accrual.

Baseline stratification factors from IRT (for PDL-1 we will use [redacted] scoring) will be summarized by randomization arm. Accrual will also be summarized by stratification factor, as stratified at IRT (with the exception of PDL-1 which will use [redacted] scoring).

7.3.2 Demographics and Other Baseline Disease Characteristics

The following baseline characteristics will be summarized by treatment group as randomized. The demography and baseline characteristics table will also be summarized for the Phase 2 and Phase 3 populations to examine consistency for all randomized participants.

- Age (continuous)
- Age categorization (>=12 and <18, ≥ 18 and < 65, ≥ 65 and < 75, ≥ 75 and < 85, ≥ 85, ≥ 75, ≥ 65)
- Sex (Male vs. Female)
- Race (as collected)
- Ethnicity (Hispanic/Latino and Not Hispanic/Latino)
- Region (USA/Canada, Europe, Latin America (Central/South America), Australia/NZ)
- Initial disease stage (I, II, III, IV)
- AJCC Stage v8 at study entry (Unresectable Stage III or Metastatic Stage IV)

  Programming notes for above: AJCC stage III vs IV at study entry

  Stage III (which will include all the sub categories IIIA, IIIB, IIIC, and IIID)

  Stage IV (which is a collapse of (which will include all the sub categories IV, IVA, IVB, IVC, IVD)

- Baseline Metastasis Stage:
  M0

  M1a (programming notes M1a= M1a, M1a (0) LDH not elevated, M1a (1) LDH elevated)

  M1b (programming notes M1b= M1b, M1b (0) LDH not elevated, M1b (1) LDH elevated)

  M1c (programming notes M1c= M1c, M1c (0) LDH not elevated, M1c (1) LDH elevated)

  M1d (programming notes M1d= M1d, M1d (0) LDH not elevated, M1d (1) LDH elevated)

- Melanoma subtype classification (mucosal, cutaneous acral, cutaneous non-acral, other)
- Baseline m stage status from CRF: M0, M1a, M1a (0), M1a (1), M1b, M1b (0), M1b (1), M1c, M1c (0), M1c (1), M1d, M1d (0), M1d (1)
- Baseline BRAF from clinical database (Mutation Positive/ Mutation Wild-type)
- Baseline PD-L1 status base on fresh vs archival biopsy
- Baseline PD-L1 from IRT (use re-scored values)
  - Baseline PD-L1+ status based on a 1% cut off (≥ 1% vs. < 1%/non-quantifiable)
  - Baseline PD-L1+ status based on a 5% cut off (≥ 5% vs. < 5%/non-quantifiable)

  - Baseline PD-L1+ status based on a 10% cut off (≥ 10% vs. < 10%/non-quantifiable)
- Baseline LAG-3 from IRT:
  - Baseline LAG-3 expression status based on 1% cut off (≥ 1% vs. < 1%)
  - Baseline LAG-3 expression status based on 5% cut off (≥ 5% vs. < 5%)
- Time from initial disease diagnosis to randomization (< 1 year, 1 - < 2 Years, 2 - < 3 Years, 3- < 4 Years, >= 5 Years, unknown)
- ECOG performance status or Karnofsky performance status (70, 80, 90, 100) (for age < 18)
- Type of prior adjuvant/neoadjuvant therapy (interferon, nivolumab/pembrolizumab, ipilimumab, braf/mek inhibitor, other)
- Prior systemic therapy regimen setting (adjuvant, neo-adjuvant, metastatic)
- Time from completion of prior adjuvant/neoadjuvant therapy (<6 months/ >= 6 months)
- Time from completion of prior adjuvant/neoadjuvant therapy (<6 months/ >= 6 months) by the categories of the prior adjuvant therapy (interferon, nivolumab/pembrolizumab, ipilimumab, braf/mek inhibitor, other)
- History of Brain metastasis
- Prior surgery (Yes vs. No)
- Prior radiotherapy (Yes vs. No)
- Baseline LDH level (≤ ULN, > ULN)
- Baseline LDH level (≤ 2 x ULN, >2 x ULN)
- Baseline hemoglobin (< LLN, ≥ LLN)
- Sites of diseases (all lesions, BICR and separately the Investigator assessment)
- Number of disease sites per subject (all lesions, BICR and separately the Investigator assessment)
- Tumor burden: sum of the diameters of target lesions at baseline (all lesions, BICR and separately the Investigator assessment)
- TUMOR BURDEN AT BASELINE PER BICR and per INV (<Quartile 1, quartile 1 to <Quartile 3 and >=Quartile 3)
- Pretreatment tumor assessment (all lesions, BICR and separately the Investigator assessment)

The IRT data will be summarized by treatment group as randomized (with the exception of PD-L1 expression which will use the score from [redacted]):

PD-L1 Expression: PD-L1>=1% tumor cell surface expression vs PD-L1<1% tumor cell surface expression/non-quantifiable.

LAG-3 Status: LAG-3>=1% expression vs LAG-3<1% expression

BRAF Status: Mutation Positive vs Mutation Wild-type

AJCC M Stage: M0/M1any [0] vs M1any[1]

PD-L1 Expression x LAG3 Status

Summary table (cross-tabulation) by treatment group for stratification factors will be provided to show any discrepancies between what was reported through IRT vs. other data sources at baseline (especially looking at PD-L1 scores from CMBP vs. the PD-L1 values re-scored by [redacted]). This summary will be performed on all randomized subjects.

7.3.3 Medical History

General medical history will be listed by subject.

7.3.4 Prior Therapy Agents

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

7.3.5 Baseline 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 (BMS986213 vs nivolumab):

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

Table 7.4.1-1: Administration of Study Therapy - Definition of Parameters

Parameter	BMS-986213 or Nivolumab
Dosing schedule per protocol	BMS986213: Relatlimab 160 mg/nivolumab 480 mg IV Q4W,
Adolescents	≥ 40 kg will receive adult dosing; for adolescents < 40 kg, dosing is relatlimab 2 mg/kg/nivolumab 6 mg/kg. or Nivolumab: 480 mg IV Q4W. Adolescents ≥ 40 kg will receive adult dosing; for adolescents < 40 kg, dosing is 6 mg/kg.
Dose	Dose (mg/kg or mg) is defined as Total Dose administered (mg/kg). Dose administered in mg/kg (or mg) at each dosing date on the CRF.
Cumulative Dose	Cum dose (mg/kg or mg) is sum of the doses administered to a subject during the treatment period.
Relative dose intensity (%)	[Cum dose (mg/kg or mg)/(Last dose date - Start dose date + 28)/7 x (planned dose)] x 100
Duration of study therapy	Last dose date - Start dose date +1, where the last dose date is event date for subjects who discontinued study therapy. Subjects who are still on therapy will be censored on their last dose date.

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose Delays

Each BMS-986213 or nivolumab infusion 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 BMS-986213 or nivolumab. 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 delay.

7.4.2.2 Infusion Interruptions and Rate Changes

Each BMS-986213 or nivolumab infusion can be interrupted and/or the IV infusion rate can be reduced. This information will be retrieved from CRF dosing pages.

The following parameters will be summarized by treatment group:

- Number of subjects with at least one dose infusion interruption, the reason for interruption, and the number of infusion interruptions per subject.

- Number of subjects with at least one IV infusion rate reduction, the reason for reduction and the number of infusion with IV rate reduction per subject.

7.4.2.3 Dose Escalations

Dose escalations (within subject) are not permitted for either BMS-986213 or nivolumab.

7.4.2.4 Dose Reductions

Dose reductions (within subject) are not permitted for either BMS-986213 or nivolumab.

7.4.2.5 Dose Omissions

Dose omissions are not permitted for either BMS-986213 or nivolumab.

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 WHO Drug Dictionary.

The following summary table will be provided:

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

A by-subject listing will accompany the table.

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

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

7.4.3.2 Subsequent Cancer Therapy

Number and percentage of subjects receiving subsequent cancer therapies will be summarized. 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

Below is a brief summary table (Table 7.5-1) of planned statistical analyses of the primary and secondary efficacy endpoints. Following the summary table are more details about each analysis in individual sections.

Unless stated otherwise, whenever a stratified analysis is specified, the following stratifications factors will be used PD-L1, LAG-3, BRAF, AJCC M Stage. Each of the 4 stratification factors has 2 levels, with 16 strata in total. Anticipating the study population may not be equally distributed across levels in a given stratification factor, it is likely that the number of subjects in some strata may be small. With 2 treatment arms in this study, the stratum size with less than 10 subjects would likely cause unreliable estimates in the stratified analyses [8,9,10,11] Therefore, in the case of less than 10 subjects in a strata or in the case of unexpected model convergence issues due to small strata with too small event numbers, stratification factors would be removed one at a time from the model until convergence is achieved. This will happen in the following order: (1) PD-L1 and (2) the stratification factor with the lowest prevalence in a level. A sensitivity analysis of PFS will be performed using the original stratification variables as per IRT.

- PD-L1 Expression: PD-L1>=1% tumor cell surface expression vs PD-L1<1% tumor cell surface expression/ non-quantifiable (from LabCorp LA)
- LAG-3 Status: LAG-3>=1% expression vs LAG-3<1% expression (from IRT)
- BRAF Status: Mutation Positive/ vs Mutation Wild-type (from IRT)
- AJCC M Stage: M0/M1any[0] vs M1any[1]) (from lab value and eCRF)

All p-values reported will be two-sided.

CIs for the efficacy endpoints will be at the two-sided 95% level. The p-values presented in the CSR will be rounded to the fourth decimal place. Point estimates and confidence bounds for efficacy variables will be rounded to the second decimal place.

Table 7.5-1: Summary Table of Efficacy Analyses

Endpoint	Statistical Analysis Methods
Phase 3 Primary	
• PFS time, as assessed by a BICR, using RECIST v1.1. PFS is defined as the time between the date of randomization and the first date of documented progression, or death due to any cause, whichever occurs first.	• Two-sided log-rank test stratified by LAG-3 expression (≥ 1% vs < 1%), (PD-L1 status (≥ 1% vs < 1%), BRAF status, and AJCC (8th edition) M Stage in randomized participants to compare the PFS of relatlimab + nivolumab arm (BMS-986213) and the nivolumab alone arm. Hazard ratios and corresponding 2-sided 95% CI will be estimated using a Cox proportional hazards model, with treatment group as a single covariate, stratified by the above factors. • PFS curves, PFS medians with 95% CIs, and PFS rates at 6 and 12 months with 95% CIs will be estimated using Kaplan-Meier methodology.
Phase 3 Secondary	
• OS time is defined as the time between the date of randomization and the date of death due to any cause.	• This test will only be interpreted if the primary analysis is significantly superior. Two-sided log-rank test stratified by LAG-3 expression (≥ 1% vs < 1%), (PD-L1 status (≥ 1% vs < 1%), BRAF status, and AJCC (8th edition) M Stage will be used to compare the relatlimab + nivolumab arm (BMS-986213) and the nivolumab arm. Hazard ratio and corresponding two-sided 95% CI will be estimated using a Cox proportional hazards model, with treatment group as a single covariate, stratified by the above factors. Additionally, OS curves, OS medians with 95% CIs, and OS rates at 12 months with 95% CIs will be estimated using Kaplan-Meier methodology.
• ORR as assessed by BICR. The ORR is defined as the number of subjects with a BOR of CR or PR divided by the number of randomized subjects for each treatment group. The BOR is defined as the best response designation, recorded between the date of randomization and the date of	• This test will only be interpreted if the primary (PFS) and first secondary (OS) analyses are significantly superior. Two-sided Cochran-Mantel-Haenszel (CMH) test stratified by LAG-3 expression (≥ 1% vs < 1%), (PD-L1 status (≥ 1% vs < 1%), BRAF status, and AJCC (8th edition) M Stage to compare the BMS-986213 arm and the nivolumab

Table 7.5-1: Summary Table of Efficacy Analyses

Endpoint	Statistical Analysis Methods
objectively documented progression per RECIST v1.1 or the date of subsequent anti-cancer therapy, whichever occurs first.	monotherapy arm. Associated odds ratios and 95% CIs will be calculated. Additionally, ORRs and corresponding 95% exact CIs will be calculated using the Clopper Pearson method.
Phase 3 Exploratory	Will be described in the SAP finalized before database lock
Phase 2 Primary	
• PFS time, as assessed by BICR, using RECIST v1.1. PFS time is defined as the time between the date of randomization and the first date of documented progression, or death due to any cause, whichever occurs first.	• Two-sided log-rank test stratified by LAG-3 expression (≥ 1% vs < 1%), (PD-L1 status (≥ 1% vs < 1%), BRAF status, and AJCC (8th edition) M Stage in randomized participants to compare the PFS of relatlimab + nivolumab arm (BMS-986213) and the nivolumab alone arm. Hazard ratios and corresponding 2-sided 95% CI will be estimated using a Cox proportional hazards model, with treatment group as a single covariate, stratified by the above factors. • PFS curves, PFS medians with 95% CIs, and PFS rates at 6 and 12 months with 95% CIs will be estimated using Kaplan-Meier methodology.
Phase 2 Secondary	
• ORR as assessed by BICR.	• Summary of ORR with corresponding 2-sided 95% CI in each arm and for the difference between arms, along with each category of BOR. Multiple comparisons in population subgroups will be handled using the Benjamini-Hochberg method to control the False Discovery Rate (FDR) at 0.05.
• DOR is defined as the time between the date of first response to the date of first documented tumor progression (per RECIST v1.1) or death due to any cause. TTR is defined as the time from randomization to the date of the first documented CR or PR. • PFS time and PFS rate as assessed by BICR, using RECIST v1.1.for subgroups	• Summary of DOR with median (95% CI) and range (min, max) by K-M method. Summary of median Time to Response (mTTR) (95% CI) • Summary of PFS with median (95% CI) by Kaplan-Meier method and PFS rate
• OS is defined as the time between the date of randomization and the date of death due to any cause.	• OS with median (95% CI) and range (min, max) by Kaplan-Meier method

7.5.1 Analysis of Progression-Free Survival

The primary objective of the study is to compare the progression-free survival (as determined by BICR) between treatment groups in all randomized subjects.

The analysis of PFS (as determined by BICR) will be to compare the two treatment groups via a stratified log-rank test among all randomized subjects. The primary definition of PFS adjusting for subsequent anticancer therapy will be used in this analysis. The two-sided log-rank p-value will be reported. The estimate of the PFS hazard ratio between treatment groups will be calculated

using a stratified Cox proportional hazards model. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio will also be presented.

The PFS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median PFS in each treatment group will be computed via the log-log transformation method. PFS rates at fixed time points (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[12] formula for variance derivation and on log-log transformation applied on the survivor function[13].

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 randomized 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

PFS will also be reported as per investigator.

7.5.2 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:

1) Delayed effect of immunotherapy interventions may cause a late separation in the PFS KM curves and non-proportional hazards. The principal sensitivity analysis of PFS (as determined by BICR) will be to compare the treatment groups via two-sided 0.05 stratified weighted log-rank test among subjects. The primary definition of PFS will be used in this analysis. 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[14].

   The Fleming Harrington test can be unstable, so it is possible, though uncommon, that the p-value for this trial will not be estimable. In this case, the principle sensitivity analysis will default to using the two sided 0.05 stratified log-rank test among all randomized subjects.

   The estimate of the PFS hazard ratio in the period before and following 6 months will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. In this model, period is a binary variable indicating pre- vs. post-6 months. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio will also be presented.

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:

- Age category (< 65, ≥ 65- <75, ≥ 75)
- Sex (male and female)
- Baseline ECOG Performance Status (0 and 1)
- History of Brain Metastases (Yes, No)
- Baseline LDH (≤ ULN, > ULN)

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

This same analysis will be performed with the CRF stratification information instead of the IRT stratification.

1) PFS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the randomized subjects.
2) PFS using the investigator's assessment. The hazard ratio associated with treatment and median PFS will be presented along with the associated two-sided 95% CIs.
3) PFS using an un-stratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.
4) PFS using an unstratified Cox proportional hazards model, adjusted, using as covariates only the four stratification factors used in randomization. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.
5) PFS 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 95% CIs.
6) During the course of this study a global pandemic occurred known as COVID-19. A sensitivity analysis will be performed if 10% of all PFS events were due to COVID-19. Subjects with a COVID-19 PFS event (i.e., death) will be censored on the previous evaluable image. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.

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.

PFS2 will be analyzed similarly to PFS:

- Median values based on KM method, along with two-sided 95% CI using Brookmeyer and Crowley method will be calculated. The estimate of standard error will be calculated using the Greenwood formula;
- PFS2 will be graphically displayed along with the median and 95% CI. By-subject listing of PFS and PFS 2 will be provided.

7.5.3 Subset Analyses of Progression-Free Survival

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

- Subgroups in the following order: (Display 95% CI)
- OVERALL
- LAG3 STATUS AT BASELINE USING 1% CUTOFF (>=1% vs <1%) source: IRT
- LAG3 STATUS AT BASELINE USING 5% CUTOFF ((>=5% vs <5%) source: IRT
- PD-L1 STATUS AT BASELINE USING 1% CUTOFF (>=1% vs <1%/non-quantifiable) (Source: combination IRT and re-score)
- PD-L1 STATUS AT BASELINE USING 5% CUTOFF ((>=5% vs <5%/non-quantifiable) (Source: combination IRT and re-score)
- PD-L1 STATUS AT BASELINE USING 10% CUTOFF ((>=10% vs <10%/non-quantifiable) (Source: combination IRT and re-score)
- INTERACTION OF LAG-3 EXPRESSION AND PDL-1 STATUS (Source: combination IRT and re-score)
- Subjects with LAG-3 >= 1% and PD-L1 >=1%
- Subjects with LAG-3 >= 1% and PD-L1 <1% (or non-quantifiable)
- Subjects with LAG-3 < 1% and PD-L1 >=1%
- Subjects with LAG-3 < 1% and PD-L1 <1% (or non-quantifiable)
- BRAF MUTATION STATUS (BRAF MUTANT; BRAF WILDTYPE) source: IRT
- AJCC STAGE (M0/M1any[0] ; M1any1) (use RAVE database – combination of CRF variable metastasis and lab data for LDH).
- BASELINE METASTASIS STAGE: (source: CRF variable metastasis and lab data for LDH)
- M0
- M1a (programming notes M1a= M1a, M1a(0) LDH not elevated, M1a(1) LDH elevated)
- M1b (programming notes M1b= M1b, M1b(0) LDH not elevated, M1b(1) LDH elevated)
- M1c (programming notes M1c= M1c, M1c(0) LDH not elevated, M1c(1) LDH elevated)
- M1d (programming notes M1d= M1d, M1d(0) LDH not elevated, M1d(1) LDH elevated)
- DISEASE STAGE AT STUDY ENTRY (III vs IV)

- HISTOLOGY (DISEASE SUBTYPE) Four categories: cutaneous acral, cutaneous non-acral, mucosal, and other
- BASELINE LDH (≤ ULN; > ULN)
- BASELINE LDH (≤ 2x ULN; > 2x ULN)
- HISTORY OF BRAIN METASTASES (YES, NO) from disease history CRF page
- TUMOR BURDEN AT BASELINE PER BICR and per INV (<Quartile 1, quartile 1 to <Quartile 3 and >=Quartile 3)
- BASELINE ECOG PS (0, 1) from CRF
- SMOKING STATUS (CURRENT/FORMER, NEVER SMOKED)
- AGE CATEGORIZATION:
- >= 12 and < 18
- >= 18 and < 65
- >= 65 and <75
- >=65
- >=75
- SEX (MALE, FEMALE)
- RACE (WHITE, BLACK OR AFRICAN AMERICAN, ASIAN, AND OTHER)
- REGION (USA/Canada, Europe, Latin America (Central/South America), Australia/NZ)

A forest plot of the PFS 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.

7.5.4 Analysis of Overall Survival

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

Overall survival will be compared between the treatment groups at the interim and final analyses, using stratified log-rank test. An O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses. The stratified hazard ratio between the treatment groups will be presented along with 100*(1- α) % CI (adjusted for interim). In addition, two-sided p-value will also be reported for the analysis of OS only during the analysis of the primary endpoint of PFS and during the final 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)

During the course of this study a global pandemic occurred known as COVID-19. A sensitivity analysis will be performed if 10% of all OS events were due to COVID-19. Subjects with a death due to COVID-19 will be censored on the start date of their COID-19 adverse event. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.

7.5.5 Subset Analyses of Overall Survival

To assess consistency of treatment effects in different subsets, a "forest" plot of the OS unstratified hazard ratios (and 95% CIs) will be produced for the same variables as in the PFS analysis (see Section 7.5.3). If a subgroup category has less than 10 subjects per treatment group, HR will not be computed/displayed.

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

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

7.5.6 Current Status of PFS and OS Follow-up

Time from last evaluable tumor assessment to cutoff date in months will be summarized descriptively (median, min, max) 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, Kaplan-Meier plots of time from randomization to post-baseline tumor assessment will be produced by treatment arm for the first twelve assessments

The current status 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 descriptively (median, min, max) 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 current status of 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) in months for all randomized subjects.

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.

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

7.5.7 Analysis of Objective Response Rate

One of the secondary 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[15] will be presented, by treatment group. A two sided 95% CI for difference of response rate between the treatment groups will also be computed.

A 2-sided, 95% CI for the difference of ORR between treatment groups will also be computed for all randomized subjects by the method of DerSimonian and Laird[16], using a fixed-effects model (setting Δ^2 equal to zero), adjusting for the stratification factors. The weighted response rate difference and 95% CI can be determined using the following formula:

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

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

A two sided 95% CI for odds ratio of response between the treatment groups will also be computed.

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. 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, progression, last dose received, and death.
- 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.

7.5.8 Subset Analyses of Objective Response

To assess consistency of treatment effects in different subsets, a "forest" plot of treatment effect on ORR per BICR in the subgroups (the same variables as in the PFS analysis (see Section 7.5.1) will be produced. The un-weighted differences in ORR between the two treatment groups and corresponding 95% two-sided CI using the method of Newcombe[17,18] will be provided.

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

7.5.9 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 randomized 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. Cumulative Response Rates will be tabulated for Week 8, Month 4, 6, 8, and 12, and overall response rate will be provided.

The following graphic will also be provided:

- For response evaluable subjects (randomized subjects with baseline and at least one on-study tumor assessment),
  - A bar plot showing event chart for time course of response (time to response, duration of response, first response, last dose)

7.5.10 Treatment-free interval

Treatment free interval and survival will be summarized in months (median and 95% CI) and rates will be provided at 1, 2, 3, 6, 9, 12, 24, 36, 48, and 60 month intervals. Kaplan-Meier plot of treatment-free interval and survival will be provided.

7.6 Safety

7.6.1 Deaths

Deaths will be summarized by treatment group:

- All deaths, reasons for death

- Deaths within 30 days of last dose received, reasons for death
- Deaths within 100 days of last dose received, reasons for death

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.

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. 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 (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT

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

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 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.

The following analyses will be conducted using the 30 days safety window and repeated using the 100 days safety window:

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

By-subject AE listing will be provided. 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. Some analyses may also be repeated by subcategory of endocrine events.

7.6.7 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 summary of any serious select AEs by worst CTC grade (any grade, grade 3-4, grade
- 5) presented by Category or Subcategory /PT
- Overall summary of drug-related serious select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category or Subcategory /PT
- Overall summary 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-subject select AE listing will be provided.

7.6.8 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.

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

7.6.10 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 AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by immune mediate Category / PT. This summary includes all AEs that are qualified for IMAE preferred terms list, without requirement of either usage of immune modulating medications or accounting for immune mediated etiology or immune mediated component. Overall summaries of IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT.
- Overall summaries of endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- Overall summaries of serious IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by Category / PT.

- Overall summaries of endocrine serious IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- Overall summaries of 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 summaries of endocrine IMAEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- Overall summaries of 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 summaries 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 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.

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 listing of Adverse Events Considered as Immune-Mediated Events per Investigator but not Qualified for Immune-Mediated Adverse Events Definition will also be provided.

In addition, for all nivolumab treated subjects who experienced at least one immune-mediated adverse event, the following data presentation will be provided:

- Summary of subjects who were re-challenged with nivolumab by immune-mediated adverse event category, with extended follow-up
- Summary of subjects who were re-challenged with BMS986213 or nivolumab by immune-mediated adverse event category, with extended follow-up

For these, re-challenge is considered to have occurred when last BMS986213 or nivolumab infusion was administered after the onset of an IMAE.

7.6.11 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

By-subject listing of OEOSI will be provided.

7.6.12 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.

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.

7.6.13 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.

7.6.14 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.

7.6.15 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.

7.6.16 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.

7.6.17 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
- ALP > 1.5 x ULN
- Concurrent (within 1 day) ALT or AST > 3 x ULN and total bilirubin > 1.5 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 1.5 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.

7.6.18 Vital Signs

Vital signs collected on the CRF will be provided in separate listings.

7.6.19 Physical Measurements

Physical measurements will be listed by subject.

7.6.20 Pregnancy

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

7.6.21 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 (12 and < 18 vs ≥ 18 < 65 vs. 65 - < 75 vs. 75 - < 85 vs. ≥ 85 vs. ≥ 75 vs. ≥ 65)
- Region (Europe, North America, Rest of the World, Asia)

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

7.6.22 Immunogenicity Analysis

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

The number (%) of participants with the following anti-drug responses will be reported by dose, if applicable, and overall.

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

Listing:

All collected immunogenicity samples will be listed with flags indicating baseline-positive sample, ADA-positive sample or ADA-negative sample, together with the associated drug concentration.

Clinical Implications:

Clinical implications of positive ADA may be explored by a comparison of ADA-positive participants to ADA-negative participants.

Effect of immunogenicity on clearance of study drugs may be explored by comparison of clearance estimates (determined by PPK analysis), if appropriate. Effect of immunogenicity on safety will be explored by examining the frequency and type of AEs of special interest such as hypersensitivity/infusion reaction. Summary tables for incidence of each preferred terms and overall as a category of AEs will be provided, if the number of participants is of sufficient size (e.g., at least 10 participants). Otherwise, individual participant's safety profile will be examined and described based on a listing. Association between trough concentrations of study drugs and ADA assessments may be explored, as needed.

7.6.23 Electrocardiogram

All of the available ECG parameter values, from each subject will be included in the ECG data set. All recorded ECG parameter values will be included in the data listings.

Baseline values are defined as the last recorded values prior to the first dosing. Although individual values for QT, QTcB, and QTcF will be presented in the data listings, only QTcF will be analyzed and discussed in the report. For other ECG parameters such as heart rate (HR), QT, QRS, and PR, summary measures (n, mean, standard deviation, median, minimum, and maximum) will be provided.

Analysis of Central Tendency: Summary statistics (n, mean, standard deviation, median, minimum, and maximum) will be presented for each ECG parameters and the corresponding changes from baseline by treatment and time point.

Listing:

- A by-subject listing of all ECG measures
- A listing of abnormal ECG interpretations

7.7 Pharmacokinetic Analyses

The PK parameters and concentration data of relatlimab and nivolumab will be summarized and listed by study part, treatment, cycle, study day and time. All derived PK parameter values will be included in the PK dataset and reported, but only subjects with adequate PK profiles will be included in summary statistics and statistical analysis.

7.7.1 Pharmacokinetic Concentrations

Summary:

Analyses for this section will be provided by the CPAR group.

Summary statistics will be provided for the following pharmacokinetic serum concentrations by treatment, cycle, study day and time for based on the relatlimab and Nivolumab PK Subjects.

- Serum concentrations of relatlimab
- Serum concentrations of Nivolumab

Listing:

- relatlimab and nivolumab serum concentrations

These data may also be pooled with other datasets for population PK analysis, which will be presented in a separate report.

The relationship of time with concentration will be investigated graphically (e.g., individual plots of change/percent change over time). Further analyses may be performed to characterize the relationship by fitting appropriate models (e.g., e-max model).

7.7.2 Ctrough and Ceoi

Ctrough and Ceoi of BMS-986213 and Nivolumab will be tabulated by treatment, cycle and study day using summary statistics. To evaluate the steady state of relatlimab concentration in the body, the geometric means of Ctrough vs. cycle and day will be plotted. Ceoi (end of infusion) and Ctrough will be listed, summarized and plotted by cycle and study day.

7.7.3 Population PK

Population PK analysis and exposure-response analysis will be performed by CP&P, if needed. The concentration vs time data obtained in this study may 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 relatlimab and to determine measures of individual exposure (such as steady-state peak, trough, and time-averaged concentration). Model determined exposures will be used for exposure-response analyses of selected efficacy and safety end points. Results of population PK and exposure-response analyses will be reported separately.

7.8 Biomarkers

Analyses listed below may or may not be performed depending on data availability (additionally some analyses in the study may be subject to change as technologies and assay methods evolve). Additional types of analyses may be conducted on a post-hoc basis pending review of data. Not all analyses will be included in the Clinical Study Report (CSR) unless they represent meaningful findings or are relevant to subject management.

To describe any changes (or percent changes) in biomarkers, summary statistics of biomarkers and their corresponding changes (or percent changes) from baseline will be tabulated by planned timepoint and dose (i.e., treatment). The time course of these biomarker measures will be investigated graphically (e.g., individual plots of change/percent change from baseline vs. time. If there is any indication of exposure-response relationship, further analyses may be performed to characterize the relationship by fitting appropriate models (e.g., e-max model).

Analyses to assess the association of clinical efficacy and safety with selected markers in peripheral blood and tumor microenvironment, may be conducted on a hierarchical basis. If a biomarker measurement undergoes meaningful change over time, further analyses may be conducted to address additional hypotheses and explore relationship between the specific change and efficacy outcomes (as categorical or continuous variables) or safety events. The analysis

methods include methods specified in this SAP and additional exploratory methods not specified in the SAP upon review of the data.

7.8.1 Biomarkers Assessed by Flow Cytometry

Summary statistics for activated flow cytometry (e.g., Immunophenotyping/Tetramer) variables and their corresponding changes from baseline will be tabulated by planned timepoints. The time course of these biomarker measures will be investigated graphically (e.g., individual plots of change/percent change from baseline vs. time). If there is any indication of exposure-response relationship, further analyses may be performed to characterize the relationship by fitting appropriate models (e.g., e-max model).

Any potential relationships between the biomarker levels or changes in biomarker levels and clinical outcomes such as tumor response, PFS, and adverse event incidence may be explored using logistic (or Cox) regression, if sufficient samples are collected.

7.8.2 Soluble Factors

Soluble factors will be analyzed similarly to flow cytometry variables. Cytokines including but not limited to IL-2Ra and IP-10 will be summarized by time point.

7.8.3 Gene expression profiling

Differences in genomic expression patterns in tumor and peripheral blood sample may be summarized for genes of particular interest.

7.8.4 Tumor mutation burden

Potential association between pre-treatment tumor mutation burden and clinical outcomes may be explored.

7.8.5 Tumor Biopsy Immunohistochemistry (IHC)

For association between efficacy and IHC assessment of T-cell infiltration or expression of markers such as MHCI, MHCII, CD3 and CD8 (IHC marker), the below analyses are expected to be performed:

Figures:

- Scatter plot of intensity of IHC markers with clinical outcome annotated (ORR)
- Change of intensity of IHC markers from pre-treatment to on-treatment when available, with clinical outcome annotated (ORR)

7.8.5.1 PD-L1 and Lag-3 Expression

Analyses of PD-L1 and LAG-3 expression (PDL-1/LAG-3) are descriptive in nature and intended to examine the distribution of PD-L1 and/or LAG-3 expression and assess potential associations between PD-L1/LAG-3 expression and efficacy measures. If there is an indication of a meaningful association, future work will evaluate PD-L1/LAG-3 expression as a predictive biomarker, including selection of an optimal PD-L1/LAG-3 expression cut-off to classify subjects as PD-L1/LAG-3 positive or PD-L1/LAG-3 negative.

PD-L1/LAG-3 expression in this section will be defined based on the validated immunoassay method for PD- L1/LAG- 3 IHC assays. PD- L1/LAG- 3 status is a categorical variable by X% cut off for quantifiable PD-L1/LAG-3 expression:

- Positive: ≥ X% PD- L1/LAG-3 expression
- Negative: < X% PD- L1/LAG-3 expression

where X denotes the possible PD- L1 expression levels1%, 5% and 10% cutoff levels and LAG-3 expression cut-off values of 1%, 5%. Additional cut off values may also be explored, but the 1% cutoff for PD- L1/LAG-3 will be considered primary.

Analyses of PD- L1/LAG-3 will include:

- Examine the distribution of PD- L1/LAG-3 expression
- Assess potential association between PD- L1/LAG-3 status and efficacy measures
- Evaluate the potential predictive relationship of the PD- L1/LAG- 3 status and efficacy measures
- Test performance statistics such as sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)
- Assess potential association between PD- L1/LAG-3 status and overall AEs

7.8.5.2 Analyses

- Descriptive statistics of PD- L1/LAG-3 expression and PD- L1/LAG-3 status, analyses will be based on all evaluable PD-L1/LAG-3 subjects if not otherwise specified:
- Listing of all PD-L1/LAG-3 IHC data.
- Summary of tumor specimen acquisition and characteristics, all randomized subjects.
- Summary statistics of PD-L1/LAG-3 expression by treatment groups of select subgroups and overall.
- Box plot of PD-L1/LAG-3 expression by treatment group and overall.
- Cumulative distribution plot of PD-L1/LAG-3 expression versus population percentile by treatment group and overall.
- Waterfall plots of individual PD-L1/LAG-3 expression by treatment group and overall.

  Note: Selected subgroups are identical to the subgroups used for PFS subgroup analysis (Section 7.5.1)

Evaluation of associations between PD-L1/LAG-3 status and efficacy measures. Each analysis will be performed for the subgroups listed below if not otherwise specified:

- Each PD-L1/LAG-3 status and combination thereof subgroup Analyses for PFS endpoint:

For each of the subgroups:

- PFS curves for each treatment group will be estimated using the Kaplan- Meier product limit method. Two-sided, 95% confidence intervals for median 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

Analyses for OS endpoint:

For each of the subgroups:

- OS curves for each treatment group will be estimated using the Kaplan-Meier product limit method. Two-sided, 95% confidence intervals for median OS will be constructed based on a log-log transformed CI for the survivor function S(t).
- Forest plot of Hazard Ratios with 95% CIs
- Analyses for ORR (BOR)
- Box plots of PD-L1/LAG-3 expression versus Response Status by treatment group

For each of the subgroups:

- Frequency and percentage of BOR will be summarized for each treatment group.
- ORR will be computed by treatment group along with exact 95% CIs using the Clopper-Pearson method.

Analyses DOR:

For each subgroup duration of response (DOR) will be summarized by treatment group.

Evaluation of the potential predictive relationship of PD-L1/LAG-3 status for efficacy measures all PD-L1/LAG-3 evaluable subjects

Analyses for PFS endpoint:

A Cox proportional hazards regression model will be fitted for PFS with treatment, PD-L1 status, LAG-3 status, PD-L1*LAG-3 status, and treatment by PD-L1/LAG-3 status interaction. Although the study is not designed to have substantial power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- Hazard ratio of treatment vs. control and its associated 95% CI for each of the PD-L1/LAG-3 status subgroup

- Hazard ratio PD-L1/LAG-3 positive vs. negative and its associated 95% CI within each treatment group. For the combination of the two biomarkers, we will designate one of the four cells to be the reference group(the PD-L1-/LAG-3-) and report the ratio of the hazard ratios

Analyses for OS endpoint:

A Cox proportional hazards regression model will be fitted for OS with treatment, PD- L1 status, LAG-3 status, PD-L1*LAG-3 status, and treatment by PD-L1/LAG-3 status interaction. Although the study is not designed to have substantial power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- Hazard ratio of treatment vs. control and its associated 95% CI for each of the

PD- L1 status subgroup

- Hazard ratio PD-L1/LAG-3 positive vs. negative and its associated 95% CI within each treatment group. For the combination of the two biomarkers, we will designate one of the four cells (the PD-L1-/LAG-3-) to be the reference group and report the ratio of the hazard ratios.

Analyses for ORR endpoint:

A logistic regression model will be fitted for response (yes=CR or PR, No=SD or PD or unknown) with treatment, PD-L1/LAG-3 status and the treatment by PD-L1/LAG-3 status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- Odds ratio of treatment vs. control and its associated 95% CI will be reported for each of the PD-L1/LAG-3 status subgroup
- Odds ratio of PD-L1/LAG-3 positive vs. negative and its associated 95% CI will be reported for each treatment group. For the combination of the two biomarkers, we will designate one of the four cells (the PD-L1-/LAG-3-) to be the reference group and report the ratio of the odds ratio.

Test performance statistics for PD-L1/LAG-3 status vs. efficacy measures on all PD- L1/LAG-3 evaluable subjects:

For each treatment group the following will be produced:

- Contingency table of PD-L1/LAG-3 status by response status (yes=CR or PR; No=SD or PD or unknown).
- Sensitivity, specificity, PPV and NPV will be reported along with the contingency table.

Association of all select AE and PD-L1/LAG-3 expression, all PD-L1/LAG-3 treated subjects Overall summary of any select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT by treatment group for following subgroups will be provided for each PD-L1/LAG-3 status subgroup.

7.9 Outcomes Research Analyses

7.9.1 EuroQol EQ-5D-3L

Unless otherwise specified, the analysis of EQ-5D will be performed in all randomized subjects who have an assessment at baseline and at least one or more post-baseline assessments. EQ-5D utility index values will be computed using a scoring algorithm based on the UK MVH-A1 TTO value set. The following descriptive analyses will be conducted.

- EQ-5D 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 EQ-5D with the problem levels for each of the 5 dimensions (mobility, selfcare, usual activities, pain/discomfort and anxiety/depression), health state (5 dimensions digits combined in a 5-digit number) and EQ-VAS will be provided.
- Proportion of subjects reporting problems for the 5 EQ-5D dimensions at each assessment timepoint will be summarized by level of problem and by treatment group. Percentages will be based on number subjects assessed at assessment time point.
- For the EQ-5D Index and visual analog scale (VAS), separately:
  - Mean score and mean change from baseline at each assessment time point will be summarized by treatment group using descriptive statistics (N, mean, SD and 95% CI, median, 25th and 75th percentiles, minimum, maximum).
  - A plot summarizing the mean change from baseline will be presented.

7.9.2 FACT-M

Unless otherwise specified, the analysis of FACT-M will be performed in all randomized subjects who have an assessment at baseline and at least one or more post-baseline assessments. The following descriptive analyses will be conducted.

- FACT-M 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.
- For the total score, subscales and GP5, separately:
  - Mean score and mean change from baseline at each assessment time point will be summarized by treatment group using descriptive statistics (N, mean, SD and 95% CI, median, 25th and 75th percentiles, minimum, maximum).
  - A plot summarizing the mean change from baseline will be presented.

- For GP5 only: Number and proportion of subjects endorsing each response option at each assessment timepoint. Percentages will be based on number of subjects assessed at assessment time point.

7.9.3 Time to Meaningful Symptomatic Deterioration of FACT-M MS

The time to meaningful symptomatic deterioration is defined as the time (in months) between the date of randomization and the first date of a 3-point or more decrease from baseline in FACT-M Melanoma Subscale (MS). The focus of this analysis is the period while patients remain on treatment; subjects who did not meet the criteria of a 3-point or more increase from baseline in FACT-M MS will be censored on the date of their last FACT-M assessment. Patients without any PRO measurement are censored at randomization.

Time to meaningful symptomatic deterioration will be analyzed using a Cox proportional hazards model with baseline as a covariate. The estimate of the hazard ratio along with a two-sided 95% CI for the hazard ratio will be calculated.

7.9.4 WPAI:GH

Unless otherwise specified, the analysis of WPAI:GH will be performed in all randomized subjects who have an assessment at baseline and at least one or more post-baseline assessments. The following descriptive analyses will be conducted:

- WPAI:GH 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.
- Employment status (Q1, Yes/No) will be summarized at each assessment time point using
- frequency counts and percentages.
- Hours actually worked during the past 7 days (Q4) will be summarized at each assessment time
- point as a categorical variable (> 0 hour/0 hour) using frequency counts and percentages
- For the four subscores, separately:
  - Absenteeism, presenteeism, work productivity loss, and activity impairment scores and postbaseline changes in scores will be summarized at each assessment time point using descriptive statistics (i.e., N, mean with SD and 95% CI, median, first and third quartiles, minimum, maximum). Absenteeism and work productivity loss scores will be summarized for participants who were employed (Q1=Yes). Presenteeism will be summarized for participants who actually worked during the past 7 days (Q4 > 0 hour). Activity impairment will be summarized for all participants. A plot summarizing the mean change from baseline will be presented.

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[19]
- 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[20].
- 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.

[redacted]

9 CONTENT OF REPORTS

All analyses described in this SAP will be included in the Clinical Study Report(s) except where otherwise noted. Refer to the Data Presentation Plan for mock-ups of all tables and listings.

10 DOCUMENT HISTORY

Table 10-1: Document History

Version Number	Author(s)	Description
1.0	[redacted]	Original publication
2.0		23NOV2020: 1. Section 2.4.1 - added details of protocol amendment 2 (Protocol revision 3). 2. Added note to Tables 4.1.1.1-1 and 4.1.2.1-1 that subjects with assessments with measurable disease per Investigator but no measurable disease per BICR (but who were evaluated per BICR) will be included in the analysis. 3. Section 4.1.3 and 4.1.4 - added notes that BOR, time to response, and duration of response per investigator will be summarized in the same manner as per BICR. 4. Section 7.2.1 - added relevant protocol deviation of baseline imaging performed outside required window per RECIST 1.1 and per protocol (beyond 28 days + 2 week window). 5. Section 7.3.2: removed time from CNS metastasis to randomization as date is not collected, but added in history of brain metastasis. Added details on the AJCC and Baseline Metastasis Stage display and programming. 50% PD-L1 cutoff was replaced with the 10% cutoff. Updated race to be as collected. 6. Section 6.3 - added in Phase 2 and Phase 3 populations and definitions. 7. Section 7.1.1.2 - updated Other Events of Special Interest which combined myositis/rhabdomyolysis together and changed aseptic meningitis to other meningitis 8. Section 7.3.2 - updated age, race, and region categories; added details about AJCC Stage v8 at study entry and • Baseline Metastasis Stage; removed time from CNS metastasis to randomization as we did not collect date; specified to use CRF disease history page to report brain metastasis; changed PD-L1 50% cutoff to 10% cutoff; added display PD-L1 Expression x LAG3 Status. Added in additional summaries of the demography and baseline characteristics tables for the Phase 2 and Phase 3 populations to examine consistency for all randomized participants. 9. Tables 7.4.1-1 - updated the duration formula to be used: last dose - first dose +1 (instead of +28) 10. Section 7.5.3 - updated or added details for categories of the subgroups for efficacy: age, region, m-stage, baseline metastasis stage, tumor burden at baseline per BICR (quartiles), histology. Specified the source for the data. 11. Sections 7.5.2 and 7.5.4 - During the course of this study a global pandemic occurred known as COVID-19. Sensitivity analyses of PFS and OS were specified due to COVID-19. These will only be performed if there are more than 10% of all PFS/OS events due to COVID-19. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs. 12. Section 7.6.1.2 - added region categories and lower bound for age category (of age 12). 13. Section 7.6.23 - added in analysis of ECGs 14. Section 7.7.1 removed plots of concentration by dose/ 15. Section 7.8.5.1 - removed the 10% cutoff analysis for LAG3, and detailed PD-L1 to use 1, 5, and 10% cutoffs. 16. Section 7.9.2 - added in separate analysis of FACT-M GP5 item. 17. Sections 4.3.2.1, 6.2.1, 7.2.1, and 7.3.2: Added details regarding any subject who had a PD-L1 value originally scored at CMBP will have a re-score. If PD-L1 was rescored, it will be used as the primary value for analyses/stratification.

Table 10-1: Document History

Version Number	Author(s)	Description
		The PD-L1 re-score will not be used to determine relevant protocol deviations/ineligibility, as it was scored posthoc. Also, if the re-scored value is evaluated as non-quantifiable, these will be categorized with PD-L1 negatives, unless greater than 5% of all randomized subjects are non-quantifiable, in which case there will be a 3rd category for PD-L1 stratification categories. Added in non-quantifiable whenever presenting/analyzing PD-L1 because of the rescored values. Also added in details about m-stage at IRT stratification did not accurately capture LDH value so the RAVE database will be used for the m-stage stratification value/ category. 18. Changed OS to the first secondary endpoint in the hierarchical testing strategy. Order is PFS, OS, ORR 19. Added interim analyses of OS at the time of positive PFS FA (OS IA1 at ~70%) and OS IA2 at ~90% of events. 20. Added GP5 single item endpoint and objective (Table 3-1 and Sections 4.3.4.2 and 7.9.2) 21. Added TFI as an endpoint and objective (Table 3-1 and Sections 4.1.6, 7.5.10) 22. Section 7.1.1 - removed sentence on dose reduction.
2.1	[redacted]	03MAR2021 1. Clarified in Section 7.5 what to do in the case of non-convergence due to small number of events in one of the 16 strata or if there are <10 subjects in a strata in the efficacy model. 2. Corrected Table 5-2.1-2, which mistakenly showed the (one-sided) cumulative alpha spent instead of the nominal p-value boundaries in the examples used to illustrate the Lan-Demets /O'Brien-Fleming spending function used for interim OS analyses. 3. Section 7.9 patient-reported outcome analyses modified to use All Randomized Subjects
3.0	[redacted]	04AUG2021 update to SAP: Section 5.2.1 was updated with information that the primary endpoint analysis was tested on 09MAR2021 and it was statistically significant. The first interim analysis of OS (OS IA1) was performed at the same time by the DMC and it was not statistically significant. As pre-specified, the significant results for the primary endpoint led to unblinding at the subject level and the reporting of PFS and all safety data. Also as pre-specified, OS and ORR were not analyzed, nor summarized by treatment arm. The positive results of the PFS analysis led to submission of the data to health authorities, with a filing on 04AUG2021. The current update to the SAP is the decision has been made to unblind response data at the time of the second interim OS analysis (OS IA2), anticipated for OCT2021. ORR will be unblinded and reported descriptively until (and if) OS is statistically significant and then it will be tested according to the hierarchy.
4.0	[redacted]	12OCT2021 update to SAP: Section 5.2.1 was updated to include the following: In the event that the database has at least 290 deaths before the second interim analysis is performed (i.e., deaths accumulate more quickly than anticipated), the second interim analysis will no longer be performed. Instead, the analysis will automatically become the final OS analysis and all deaths in the database will be used in the analysis. If this occurs, BMS, and not the DMC, will analyze the OS data and be unblinded to OS results. Table 5.2.1-3 gives details on the alpha and critical hazard ratio if no second interim analysis is performed and there are 300 deaths at the time of lock.

Table 10-1: Document History

Version Number	Author(s)	Description
5.0	[redacted]	However, this will be calculated based on the actual data in the database at the time of lock. The update in section 5.2.1 was revised to be more clear that the specification to remove the second interim OS analysis and go straight to the final OS analysis was made with full knowledge of how many deaths were currently in the database, but with no knowledge of the OS analysis by treatment arm.

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

Time-to onset definition

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

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

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

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

Time-to resolution definition

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

Time-to resolution of AE (any grade) for a specific category is defined as the longest time from onset to complete resolution or improvement to the grade at baseline among all clustered AEs in this category experienced by the subject. 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 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 category.	Collapse any on-treatment AE from the same 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).

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 timepoints, 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 timepoint
3) *Other Positive:* Not persistent positive but some ADA-positive samples with the last sample being negative

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

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

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

Population for Analyses

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

11 REFERENCES

1 EuroQol Group. EuroQol--a new facility for the measurement of health related quality of life. Health Policy 1990;16:199–208

2 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.

3 Cormier JN, Ross MI, Gershenwald JE, et al. Prospective assessment of the reliability, validity, and sensitivity to change of the Functional Assessment of Cancer Therapy-Melanoma questionnaire. Cancer. 2008;112:2249–2257.

4 Askew RL, Xing Y, Palmer JL, et al. Evaluating minimal important differences for the FACT-Melanoma quality of life questionnaire. Value in Health. 2009 Nov-Dec; 12(8): 1144-50.

5 Robinson DW, Cormier JN, Zhao N, et al. Health-related quality of life among patients with metastatic melanoma: results from an international phase 2 multicenter study. Melanoma Research. 2012 Feb; 22(1): 54-62.

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

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

8 Schoenfeld, David A., and Anastasios A. Tsiatis. "A modified log rank test for highly stratified data." *Biometrika* 74.1 (1987): 167-175.

9 Kernan, Walter N., et al. "Stratified randomization for clinical trials." *Journal of clinical epidemiology* 52.1 (1999): 19-26.

10 De Stavola, B. L., and D. R. Cox. "On the consequences of overstratification." *Biometrika* 95.4 (2008): 992-996.

11 Feng, Changyong, Hongyue Wang, and Xin M. Tu. "Power loss of stratified log-rank test in homogeneous samples." *Journal of Quality and Reliability Engineering* 2010 (2010).

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

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

14 Fleming TR, Harrington DP 1991: Counting processes and survival analysis. New York: John Wiley and Sons

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

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

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

[18] Newcombe, R. G. Interval estimation for the difference between independent proportions: comparison of eleven methods, Statistics in Medicine, 17, 873-890 (1998)

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

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