Official Protocol Title:	Adjuvant Therapy with Pembrolizumab versus Placebo in Resected Highrisk Stage II Melanoma: A Randomized, Double-blind Phase 3 Study (KEYNOTE 716)
NCT number:	NCT03553836
Document Date:	19-JAN-2021

Supplemental Statistical Analysis Plan (sSAP)

Prepared by

CONFIDENTIAL, PROPRIETARY PROPERTY OF MERCK SHARPE & DOHME CORP., A SUBSIDIARY OF MERCK & CO., INC. (COLLECTIVELY "MERCK")

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

This supplemental SAP (sSAP) is a companion document to the protocol. In addition to the information presented in the protocol SAP which provides the principal features of confirmatory analyses for this trial, this supplemental SAP provides additional statistical analysis details/data derivations and documents modifications or additions to the analysis plan that are not "principal" in nature and result from information that was not available at the time of protocol finalization. The patient reported outcome (PRO) analysis plan is also included in this sSAP. Separate analysis plans (i.e., separate documents from this sSAP) may be developed for PK/modeling analysis, biomarker analysis, and genetic data analysis.

2 SUMMARY OF CHANGES

This sSAP amendment 01 is based on protocol amendment 03 to support recurrence free survival (RFS) interim analysis. Changes from sSAP amendment 01 are listed below:

Section Number	Description of Changes from Protocol
Section 3.1	Aligned with SAP in protocol amendment 03 and current department standard sSAP template.
Section 3.3	Aligned Objective/Hypothesis with protocol amendment 03 section 3
Section 3.4.1	Added TTST and PRFS2 to align with protocol amendment 03.
Section 3.4.2	Aligned with current department standard sSAP template.
Section 3.4.3, 3.5.3, 3.6.3	Updated PRO sections to align with most current sSAP template.
Section 3.6.1	Added clarity to analyses of DMFS, OS, TTST and PRFS2. Aligned with current department standard sSAP template.
Section 3.6.1.1	Clarified stratum 1 (pediatric participants) will be combined with other strata in case of small number of participants. Aligned with current department standard sSAP template. Added a sensitivity analysis to include new primary melanomas as RFS events.
Section 3.6.1.4, 3.6.1.5	Added sections for analyses of TTST and PRFS2.
Section 3.6.2, 3.6.3	Aligned with current department standard sSAP template.
Section 3.6.4	Added summary of COVID-19 impact and aligned with current department standard sSAP template.
Section 3.7, 3.8, 3.9	Aligned with SAP in protocol amendment 03 and current department standard sSAP template.
Section 3.10	Added subgroup analysis by Region. Aligned with current department standard sSAP template regarding subgroup analysis.
Section 3.11	Aligned with current department standard sSAP.

3 ANALYTICAL AND METHODOLOGICAL DETAILS FOR GLOBAL STUDY

3.1 Statistical Analysis Plan Summary

Key elements of the statistical analysis plan are summarized below; the comprehensive plan is provided in Sections 3.2-3.12.

Study Design Overview	A randomized, placebo-controlled, parallel-group, crossover/rechallenge, multi-center study of adjuvant pembrolizumab in participants 12 years of age and older with resected Stage IIB or IIC cutaneous melanoma.
Treatment Assignment	Part 1 of this trial will be conducted as a double-blind trial. Approximately 954 participants will be randomized in about 15 months (double-blind) in a 1:1 ratio between 2 treatment arms: (1) Pembrolizumab as adjuvant therapy or (2) Placebo as adjuvant therapy. Stratification factors are as follows: one stratum for pediatric (age 12-17) participants and 3 strata for adult (age 18 and over) participants defined by T-stage (T3b, T4a, and T4b). Part 2 of this trial (at recurrence: cross-over treatment with pembrolizumab for participants initially randomized to placebo; re-challenge with pembrolizumab for participants initially randomized to pembrolizumab) is open-label. Eligible participants in Part 2 will receive pembrolizumab.
Analysis Populations	Efficacy: Intention-to-Treat Population (ITT) Safety: All Participants as Treated (APaT)
Primary Endpoint	Recurrence-free survival (RFS)
Key Secondary Endpoint(s)	1. Distant Metastasis-free Survival (DMFS) 2. Overall Survival (OS)
Statistical Methods for Key Efficacy Analyses	The primary hypotheses will be evaluated by comparing pembrolizumab with placebo arm with respect to RFS, DMFS and OS using a stratified log-rank test. The hazard ratio (HR) will be estimated using a stratified Cox model. Event rates over time will be estimated within each treatment group using the Kaplan-Meier method.
Statistical Methods for Key Safety Analyses	The analysis of safety will follow a tiered approach. The tiers differ with respect to the analyses that will be performed. There are no events of interest that warrant elevation to Tier 1 events in this study. Tier 2 parameters will be assessed via point estimates with 95% confidence intervals (CIs) provided for between-treatment comparison; only point estimates by treatment group are provided for Tier 3 safety parameters. The 95% CIs for the between-treatment differences in percentages will be provided using the Miettinen and Nurminen method.

Interim Analyses	Five efficacy interim analyses (IAs) are planned in this study. Results will be reviewed by an external Data Monitoring Committee (eDMC). Details are provided in Section 3.7 –Interim Analyses.
	
	Efficacy IAs
	
	• Interim Analysis 1 (IA1):
	  ○ Timing: to be performed after ~128 RFS events have been observed, ~33 months after the first participant randomized
	  ○ Primary purpose: First interim analysis of RFS
	• Interim Analysis 2 (IA2):
	  ○ Timing: to be performed after ~179 RFS events have been observed, ~48 months after the first participant randomized
	  ○ Primary purpose: Final analysis of RFS
	• Interim Analysis 3 (IA3):
	  ○ Timing: to be performed after ~146 DMFS events have been observed, ~60 months after the first participant randomized
	  ○ Primary purpose: First interim analysis of DMFS
	• Interim Analysis 4 (IA4):
	  ○ Timing: to be performed after ~195 DMFS events have been observed, ~108 months after the first participant randomized
	  ○ Primary purpose: Final analysis of DMFS
	• Interim Analysis 5 (IA5):
	  ○ Timing: to be performed after ~154 OS events have been observed, ~120 months after the first participant randomized
	  ○ Primary purpose: First interim analysis of OS
	• Final analysis:
	  ○ Timing: to be performed after ~204 OS events have been observed, ~180 months after the first participant randomized
	  ○ Primary purpose: Final analysis of OS

Multiplicity	The overall type-I error over the three hypotheses (RFS, DMFS and OS) is strongly controlled at 2.5% (one-sided), with 2.5% initially allocated to the RFS hypothesis.
Sample Size and Power	The planned sample size is approximately 954 participants. It is estimated there will be ~179 events at the RFS final analysis (ie, IA2 of the study). With 179 RFS events, the study has ~92% power for detecting a HR of 0.6 at 2.5% (1-sided) significance level. It is estimated there will be ~195 events at the DMFS final analysis (ie, IA4 of the study). With 195 DMFS events, the study has ~84% power for detecting a HR of 0.65 at 2.5% (1-sided) significance level. It is estimated there will be ~204 events at the OS final analysis (ie, FA of the study). With 204 OS events, the study has ~80% power for detecting a HR of 0.67 at 2.5% (1-sided) significance level.

3.2 Responsibility for Analyses/In-House Blinding

The statistical analysis of the data obtained from this study will be the responsibility of the Clinical Biostatistics Department of the Sponsor.

The Sponsor will generate the randomized allocation schedule(s) for study treatment assignment for this protocol, and the randomization will be implemented in IRT.

Part 1 of this trial will be conducted as a double-blind trial under in-house blinding procedures. Blinded sponsor personnel will remain blinded throughout study conduct and the official, final database will not be unblinded until medical/scientific review has been performed, protocol deviations have been identified, and data have been declared final and complete. In addition, the site radiologist(s) will perform the imaging review without knowledge of treatment group assignment and provide their results to the site investigator.

Blinding with respect to planned efficacy interim analyses is described in Section 3.7 – Interim Analyses. Protocol-specified blinding to treatment assignment will be maintained at all investigational sites.

Part 2 of this trial is open-label, so individuals participating in Part 2 will not remain blinded throughout the entire conduct of the trial.

3.3 Hypotheses/Estimation

3.3.1 Primary Objective(s) and Hypothesis(es)

Objective: To compare Recurrence-free Survival (RFS) between treatment arms.

Hypothesis: Pembrolizumab is superior to placebo with respect to RFS as assessed by the site investigator.

3.3.2 Secondary Objective(s) & Hypothesis(es)

1) **Objective:** To compare DMFS between treatment arms.

   Hypothesis: Pembrolizumab is superior to placebo with respect to DMFS as assessed by the site investigator.

2) **Objective:** To compare OS between treatment arms.

   Hypothesis: Pembrolizumab is superior to placebo with respect to OS.

3) **Objective:** To assess the safety and tolerability of pembrolizumab compared to placebo in the proportion of AEs.

3.3.3 Exploratory Objectives

1) **Objective:** To compare average change from baseline during the adjuvant treatment period (up to 21 days after last administration) in global quality of life between the 2 treatment arms using the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) global health status/QoL scale.

2) **Objective:** To compare average change from baseline after the adjuvant period (from 21 days after last administration) in global quality of life between the 2 treatment arms using the EORTC QLQ-C30 global health status/QoL scale.

3) **Objective:** To characterize health utilities using the EuroQoL-5 Dimension Questionnaire (EQ-5D-5L) healthy utility scores.

4) **Objective**: To compare the time to subsequent therapy (TTST) between treatment arms.

5) **Objective**: To compare Progression/recurrence-free Survival 2 (PRFS2) between treatment arms.

6) **Objective**: To identify molecular (genomic, metabolic, and/or proteomic) biomarkers that may be indicative of clinical response/resistance, safety, pharmacodynamic activity, and/or the mechanism of action of pembrolizumab.

3.4 Analysis Endpoints

3.4.1 Efficacy Endpoints

3.4.1.1 Primary Efficacy Endpoint

Recurrence-free Survival (RFS)

Recurrence-free Survival is defined as the time from randomization to any of the following events: recurrence of melanoma at any site (local, in-transit or regional lymph nodes or distant recurrence) or death due to any cause. New incident cases of melanoma and second cancer diagnoses are not counted as events for recurrence-free survival. RFS analysis will be done for Part 1.

3.4.1.2 Secondary Efficacy Endpoints

Distant Metastasis-free Survival (DMFS)

Distant Metastasis-free Survival is defined as the time from randomization to the first diagnosis of a distant metastasis. Distant metastasis refers to cancer that has spread from the original (primary) tumor and beyond local tissues and lymph nodes to distant organs or distant lymph nodes.

Overall Survival (OS)

Overall survival is defined as the time from randomization to death due to any cause.

3.4.1.3 Tertiary/Exploratory Efficacy Endpoints

Time to Subsequent Therapy (TTST)

Time to subsequent therapy is defined as time from randomization to the date of first subsequent therapy (eg, surgery, radiation therapy, antineoplastic therapy) or death (whatever the cause) whichever occurs first.

Progression-free survival 2 (PRFS2)

Progression/recurrence-free Survival 2 is defined as the time between the date of randomization and the earliest of the following:

- date of 1st disease progression per RECIST1.1 beyond the initial unresectable disease recurrence (unresectable local-regional disease recurrence or unresectable distant metastatic disease recurrence);
- date of 2nd recurrence in Participants without evidence of disease after surgery of a resectable 1st recurrence (resectable local regional recurrences or resectable distant metastatic disease recurrence);
- date of death.

3.4.1.4 Exploratory Endpoints Associated with Protocol Part 2

Progression-Free Survival

Progression-free survival (PFS) is defined as the time from the date of 1st dose in Part 2 to the first documented disease progression or death due to any cause, whichever occurs first.

The following analyses will be done for Part 2.

- PFS per RECIST 1.1 assessed by investigators
- PFS per iRECIST assessed by investigators.

Objective response rate

Objective response rate (ORR) is defined as the proportion of the participants in the analysis population who have a confirmed complete response (CR) or partial response (PR).

The following analyses will be done for Part 2.

- ORR per RECIST 1.1 assessed by investigators
- ORR per iRECIST assessed by investigators.

Duration of Overall Response (DOR)

For participants who demonstrated CR or PR, response duration is defined as the time from first documented evidence of CR or PR until disease progression or death due to any cause, whichever occurs first.

The following analyses will be done for Part 2.

- DOR per RECIST 1.1 assessed by investigators
- DOR per iRECIST assessed by investigators.

Overall Survival (OS)

For Part 2, overall survival is defined as the time from the date of 1st dose in Part 2 to death due to any cause.

3.4.2 Safety Endpoints

Safety measurements are described in protocol Section 4.2.1 – Rationale for Endpoints and Section 8.3 – Safety Assessments. Safety and tolerability will be assessed by clinical review of all relevant parameters including AEs, AEs leading to discontinuation, SAEs, fatal AEs, laboratory values and vital signs.

3.4.3 Patient-Reported Outcome Endpoints

The PRO instruments in this study are the EORTC QLQ-C30 and EuroQol-5D (EQ-5D). The EQ-5D will provide data for use in economic models and analyses including developing health utilities or quality-adjusted life years (QALYs). The following PRO endpoints will be assessed.

- Mean change from baseline in EORTC QLQ-C30 global health status/quality of life scores and physical functioning score at pre-specified timepoint (see definition in section 3.6.3.2)

- Mean change from baseline in EQ-5D-5L visual analogue scale (VAS) at pre-specified timepoint

The EORTC QLQ-C30 global health status/quality of life scores and physical functioning score will also be assessed by:

- **Overall improvement**

  Improvement is defined as a 10-point or more increase in score (in the positive direction) from baseline at any time during the study and confirmed by a 10-point or more improvement at a visit scheduled at least 6 weeks later (Osoba et al., 1998 [1], King 1996 [2]).

- **Overall improvement/stability**

  Stability is defined as, when the criteria for improvement are not met, a less than 10 points worsening in score from baseline at any time during the study and confirmed by a less than 10 points worsening at a visit scheduled at least 6 weeks later.

  Overall improvement/stability is defined as the composite of improvement and stability.

The definition of improvement as a change of 10 points or greater from baseline is per Osoba et al. (1998) [1] and King (1996) [2]. Changes from baseline in EORTC QLQ-C30 scores will also be interpreted according to recent subscale-specific guidelines, which indicate that clinically meaningful differences vary by scale (Cocks et al., 2012).

- **Deterioration**

  Deterioration is defined as, greater than 10 points worsening in score from baseline at any time during the study.

3.5 Analysis Populations

3.5.1 Efficacy Analysis Populations

For Part 1, the intention-to-treat (ITT) population will serve as the population for primary efficacy analyses. All randomized participants will be included in this population. Participants will be included in the treatment group to which they are randomized.

For Part 2, participants who are eligible to receive retreatment with pembrolizumab according to the criteria in protocol Section 6.7 will serve as the population for efficacy analyses. Details on the approach to handling missing data are provided in Section 3.6 – Statistical Methods.

3.5.2 Safety Analysis Populations

The All Participants as Treated (APaT) population will be used for the analysis of safety data. The APaT population in Part 1 consists of all randomized participants who received at least one dose of study treatment in Part 1. Similarly, the APaT population in Part 2 consists of all Part 2 participants who received at least one dose of study treatment in Part 2.

For part 1, participants will be included in the treatment group corresponding to the study treatment they actually received for the analysis of safety data using the APaT population. For most participants this will be the treatment group to which they are randomized. Participants who take incorrect study treatment for the entire treatment period will be included in the treatment group corresponding to the study treatment actually received. Any participant who receives the incorrect study treatment for one cycle, but receives the correct treatment for all other cycles, will be analyzed according to the randomized treatment group and a narrative will be provided for any events that occur during the cycle for which the participant is incorrectly dosed.

In Part 1 (Part 2), at least one laboratory or vital sign measurement obtained subsequent to at least one dose of study treatment of Part 1 (Part 2) is required for inclusion in the analysis of each specific parameter in Part 1 (Part 2). To assess change from baseline in Part 1 (Part 2), a baseline measurement in Part 1 (Part 2) is also required. See baseline definitions in section 3.6.4 of this document.

3.5.3 PRO Analysis Populations

PRO analyses are based on the PRO Full Analysis Set (FAS) population, defined as all randomized participants who have at least one PRO assessment available for the specific endpoint and have received at least one dose of study intervention. Participants will be analyzed in the treatment group to which they are randomized.

3.6 Statistical Methods

3.6.1 Statistical Methods for Efficacy Analyses

Statistical testing and inference for safety analyses are described in Section 3.6.2. Efficacy results that will be deemed to be statistically significant after consideration of the Type I error control strategy are described in Section 3.8, Multiplicity. Nominal p-values will be computed for other efficacy analyses, but should be interpreted with caution due to potential issues of multiplicity. For efficacy endpoints associated with Part 2, only descriptive summary statistics will be provided. The summary will be done separately by the randomized treatment group in Part 1.

The stratification factors used for randomization (see protocol Section 6.3.1.1 - Stratification) will be applied to all stratified analyses, in particular, the stratified log-rank test, stratified Cox model, and stratified Miettinen and Nurminen method [3]. In the event that there are small strata, for the purpose of analysis, strata will be combined to ensure sufficient number of participants, responses and events in each stratum. Due to the small number of pediatric participants enrolled (2 participants), stratum 1 (pediatric participants) will be combined with other strata according to the T-stage level. This is applicable to all stratified analyses throughout this sSAP.

For analyses of DMFS, OS, TTST and PRFS2, follow-up will continue whether or not participants go from Part 1 to Part 2, i.e. for Participants go from Part 1 to Part 2 and have first event in Part 2, their event date or the date of the last assessment date in Part 2 (censoring date if no documented event at the time of the analysis) will be used for analysis.

3.6.1.1 Recurrence-free Survival (RFS)

For Part 1 analysis, the non-parametric Kaplan-Meier method will be used to estimate the RFS curve in each treatment group. The treatment difference in RFS will be assessed by the stratified log-rank test. A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (ie, HR) between the treatment arms. The HR and its 95% CI from the stratified Cox model with Efron's method of tie handling and with a single treatment covariate will be reported. Kaplan-Meier estimates and the corresponding 95% CIs at specific follow-up time-points will be provided for RFS. The stratification factors used for randomization (see protocol Section 6.3.1.1 – Stratification) will be applied to both the stratified log-rank test and the stratified Cox model.

For the primary RFS analysis, the true date of event will be approximated by the date of the first assessment at which event is objectively documented. Participants who do not experience an event at the time of analysis will be censored at the last disease assessment date.

Since disease assessment is performed periodically, events such as disease recurrence and metastatic disease recurrence can occur any time in the time interval between the last assessment where the event was not documented and the assessment when the event is documented. In order to evaluate the robustness of the RFS endpoint, a sensitivity analysis with a different set of censoring rules will be performed. The sensitivity analysis is the same as the primary analysis except that events after 2 consecutive missed disease assessments or after new anti-cancer therapy if any should be censored at last disease assessment prior to the earlier date of ≥2 consecutive missed disease assessments and new anti-cancer therapy. The censoring rules for primary and sensitivity analyses of RFS are summarized in Table 1.

Table 1 Censoring Rules for Primary and Sensitivity Analyses of RFS

Situation	Primary Analysis	Sensitivity Analysis
Recurrence or death documented after ≤1 missed disease assessment, and before new anti-cancer therapy, if any	Event at earliest date of documented recurrence or death	Event at earliest date of documented recurrence or death
Recurrence or death documented immediately after ≥2 consecutive missed disease assessments or after new anti-cancer therapy, if any	Event at earliest date of documented recurrence or death	Censored at last disease assessment prior to the earlier date of ≥2 consecutive missed disease assessment and new anti-cancer therapy, if any
No recurrence and no death; and new anticancer treatment is not initiated	Censored at last disease assessment	Censored at last disease assessment
No recurrence and no death; new anticancer treatment is initiated	Censored at last disease assessment	Censored at last disease assessment before new anticancer treatment

As indicated in Section 3.4.1.1, new primary melanomas will not be counted as RFS events for the primary RFS analysis. A sensitivity analysis to include new primary melanomas as RFS events will be performed. Additional supportive unstratified analyses may also be provided.

3.6.1.2 Distant Metastases-free Survival (DMFS)

Non-parametric cumulative incidence curves will be used to estimate the 'time to metastatic disease' curves. The treatment difference in risk for metastatic disease will be assessed by the stratified log-rank test. A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (ie, the HR). The HR and its 95% CI from the stratified Cox model with a single treatment covariate will be reported. The stratification factors used for randomization will be applied, as stratification factors used for analysis, to both the stratified log-rank test and the stratified Cox model. Participants without documented metastatic disease diagnosis will be censored at the date of their last disease assessment.

3.6.1.3 Overall Survival (OS)

The non-parametric Kaplan-Meier method will be used to estimate the survival curves. The treatment difference in survival will be assessed by the stratified log-rank test. A stratified Cox proportional hazard model with Efron's method of tie handling will be used to assess the magnitude of the treatment difference (the HR). The HR and its 95% CI from the stratified Cox model with a single treatment covariate will be reported. The stratification factors used for randomization (See protocol Section 6.3.1.1 – Stratification) will be applied to both the stratified log-rank test and the stratified Cox model. Kaplan-Meier estimates and the

corresponding 95% CIs at specific follow-up time-points will be provided for OS. Participants without documented death at the time of the analysis will be censored at the date of the participant was last known to be alive.

Additional supportive unstratified analyses may also be provided.

3.6.1.4 Time to subsequent therapy (TTST)

The non-parametric Kaplan-Meier method will be used to estimate the distribution curve in each treatment group. 95% CIs for the median and point estimates at various follow-up times from randomization date will be calculated. Participants who remain alive and do not receive subsequent therapy will be censored on the date of last disease assessments. Pembrolizumab treatment in Part 2 will be included as a subsequent therapy for participants who enter part 2.

3.6.1.5 Progression-free survival 2 (PRFS2)

The non-parametric Kaplan-Meier method will be used to estimate the distribution for PRFS2. 95% CIs for the median and point estimates at various follow-up times from randomization date will be calculated. Participants who remain alive and whose disease has not recurred, or disease has recurred, but subsequent disease progression or recurrence has not occurred, will be censored on the date of last disease assessments.

3.6.1.6 Progression-Free Survival (PFS)

For Part 2, the non-parametric Kaplan-Meier method will be used to estimate the PFS distribution. 95% CIs for the median PFS and PFS point estimates at various follow-up times from first day of study treatment in Part 2 will be calculated.

Since disease progression is assessed periodically, progressive disease (PD) can occur any time in the time interval between the last assessment where PD was not documented and the assessment when PD is documented. The true date of disease progression will be approximated by the earlier of date of the first assessment at which PD is objectively documented per RECIST 1.1 by investigators and the date of death. Death is always considered as a PD event. Sensitivity analyses will be performed based on iRECIST by investigator's assessment.

For the analysis of PFS, any participant who experiences an event (PD or death) immediately after 2 or more missed disease assessments will be censored at the last disease assessment prior to the missed visits. In addition, any participant who initiates new anti-cancer therapy prior to documented progression will be censored at the last disease assessment prior to the initiation of new anti-cancer therapy. Participants who do not start new anti-cancer therapy and who do not experience an event will be censored at the last disease assessment. If a subject meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied. The censoring rules for PFS analyses are summarized in Table 2.

Table 2 Censoring Rules for Analyses of PFS

Situation	Date of Progression or Censoring
PD or death documented after ≤ 1 missed disease assessment, and before new anti-cancer therapy, if any	Progressed at date of documented PD or death
Death or progression after ≥ 2 consecutive missed disease assessments without further valid non-PD disease assessments or after new anti-cancer therapy	Censored at last disease assessment prior to the earlier date of ≥ 2 consecutive missed disease assessment and new anti-cancer therapy, if any
No PD and no death; and new anticancer treatment is not initiated	Censored at last disease assessment
No PD and no death; new anticancer treatment is initiated	Censored at last disease assessment before new anticancer treatment
PD = progressive disease	

3.6.1.7 Objective Response Rate (ORR)

For Part 2, ORR will be calculated as the ratio of the number of participants reported to have achieved a confirmed CR or PR, divided by the number of participants in Part 2. Part 2 participants without ORR assessments will be counted as non-responders.

The point estimate of ORR will be provided, together with 95% CI using exact binomial method proposed by Clopper and Pearson (1934) [5].

3.6.1.8 Duration of Response (DOR)

If sample size permits, DOR will be summarized descriptively using Kaplan-Meier medians and quartiles. Only the subset of Part 2 participants who show a confirmed complete response or partial response will be included in this analysis. Censoring rules for DOR are summarized in Table 3.

For each DOR analysis, a corresponding summary of the reasons responding participants are censored will also be provided. Responding participants who are alive, have not progressed, have not initiated new anti-cancer treatment, have not been determined to be lost to follow-up, and have had a disease assessment within ~6 months of the data cutoff date are considered ongoing responders at the time of analysis. If a subject meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied.

Table 3 Censoring Rules for DOR

Situation	**Date of Progression or Censoring**	**Outcome**
No progression nor death, no new anticancer therapy initiated	Last adequate disease assessment	Censor (non-event)
No progression nor death, new anticancer therapy initiated	Last adequate disease assessment before new anti-cancer therapy initiated	Censor (non-event)
Death or progression immediately after ≥ 2 consecutive missed disease assessments or after new anti-cancer therapy, if any	Earlier date of last adequate disease assessment prior to ≥ 2 missed adequate disease assessments and new anti-cancer therapy, if any	Censor (non-event)
Death or progression after ≤1 missed disease assessments and before new anti-cancer therapy, if any	PD or death	End of response (Event)
Note: A missed disease assessment includes any assessment that is not obtained or is considered inadequate for evaluation of response.		

3.6.1.9 Analysis Strategy for Key Efficacy Endpoints

Table 4 summarizes the primary analysis approach for primary and key secondary efficacy endpoints. Sensitivity analysis methods are described above for each endpoint as applicable.

Table 4 Efficacy Analysis Methods for Key Efficacy Endpoints

Endpoint/Variable (Description, Time Point)	Statistical Method†	Analysis Population	Missing Data Approach
Primary Hypothesis 1			
RFS	Test: Stratified log-rank test Estimation: Stratified Cox model with Efron's tie handling method	ITT	See Table 1 for censoring rules
Secondary Hypothesis 2			
DMFS	Test: Stratified log-rank test Estimation: Stratified Cox model with Efron's tie handling method	ITT	Censored at last disease evaluation
Secondary Hypothesis 3			
OS	Test: Stratified log-rank test Estimation: Stratified Cox model with Efron's tie handling method	ITT	Censored at last known alive date

† Statistical models are described in further details in the text. For stratified analyses, the stratification factors used for randomization will be used as stratification factors for analysis.

3.6.2 Statistical Methods for Safety Analyses

Safety and tolerability will be assessed by clinical review of all relevant parameters including adverse experiences (AEs), laboratory tests and vital signs.

In Part 1, the analysis of safety results will follow a tiered approach as shown in Table 5. The tiers differ with respect to the analyses that will be performed. Adverse experiences (specific terms as well as system organ class terms) and events are either pre-specified as "Tier 1" endpoints, or will be classified as belonging to "Tier 2" or "Tier 3" based on observed proportions of participants with an event.

Tier 1 Events

Safety parameters or AEs of special interest that are identified a priori constitute "Tier 1" safety endpoints that will be subject to inferential testing for statistical significance. AEs that are immune-mediated or potentially immune-mediated are well documented and will be evaluated separately; however, these events have been characterized consistently throughout the pembrolizumab clinical development program, and determination of statistical

significance is not expected to add value to the safety evaluation. Based on a review of historic chemotherapy data and data from ongoing pembrolizumab clinical studies in gastric cancer, there are no AEs of interest that warrant inferential testing for comparison between treatment arms in this study. Therefore, there are no Tier 1 events for this protocol.

Tier 2 Events

Tier 2 parameters will be assessed via point estimates with 95% CIs provided for differences in the proportion of participants with events using the Miettinen and Nurminen method, an unconditional, asymptotic method [3].

Membership in Tier 2 requires that at least 10% of participants in any treatment group exhibit the event; all other adverse experiences and predefined limits of change will belong to Tier 3. The threshold of at least 10% of participants was chosen for Tier 2 event because the population enrolled in this study are in critical conditions and usually experience various adverse events of similar types regardless of treatment, events reported less frequent than 10% of participants would obscure the assessment of overall safety profile and add little to the interpretation of potentially meaningful treatment differences. In addition, any AE, any drug related AE, any Grade 3-5 AE, any serious AE, any AE which is both drug-related and Grade 3-5, any AE which is both serious and drug-related, a discontinuation due to an AE, and death will be considered Tier 2 endpoints. Because many 95% confidence intervals may be provided without adjustment for multiplicity, the CIs should be regarded as a helpful descriptive measure to be used in review, not a formal method for assessing the statistical significance of the between-group differences.

Tier 3 Events

Safety endpoints that are not Tier 1 or 2 events are considered Tier 3 events. Only point estimates by treatment group are provided for Tier 3 safety parameters (Table 5).

Continuous Safety Measures

For continuous measures such as changes from baseline in laboratory tests and vital sign parameters, summary statistics for baseline, on-treatment, and change from baseline values will be provided by treatment group in table format.

Table 5 Analysis Strategy for Safety Parameters

Safety Tier	Safety Endpoints†	95% CI per Cohort and Combined Treatment Group if Needed	Descriptive Statistics
Tier 2	Any AE	X	X
	Any Grade 3 to 5 AE	X	X
	Any serious AE	X	X
	Any drug-related AE	X	X
	Any serious and drug-related AE	X	X
	Any Grade 3 to 5 and drug-related AE	X	X
	Dose modification due to AE	X	X
	Discontinuation due to AE	X	X
	Death	X	X
	Specific AEs, SOC or PDLCs‡ (incidence ≥ 10% in one of the treatment groups)	X	X
Tier 3	Specific AEs, SOCs or PDLCs‡ (incidence <10% in both treatment groups)		X
	Change from baseline results (labs, ECGs, vital signs)		X

† Adverse event references apply to both clinical and laboratory AEs.

‡ Includes only those endpoints not pre-specified as Tier 1 or not already pre-specified as Tier 2 endpoints.

AE = adverse event; CI = confidence interval; ECG = electrocardiogram; PDLC = Pre-defined limits of change; SOC = system organ class; X = results will be provided.

3.6.3 Statistical Methods for Patient-Reported Outcome Analyses

This section describes the planned analyses for the PRO endpoints. The patient-reported outcomes are exploratory objectives in this study, and thus no formal hypotheses were formulated. Nominal p-value to compare the pembrolizumab arm to the control arm may be provided as appropriate.

3.6.3.1 PRO compliance summary

Completion and compliance of EORTC QLQ-C30 and EQ-5D by visit and by treatment will be described. Numbers and percentages of complete and missing data at each visit will be summarized. An instrument is considered complete if at least one valid score is available according to the missing item rules outlined in the scoring manual for the instrument.

Completion rate of treated participants (CR-T) at a specific time point is defined as the number of treated participants who complete at least one item over the number of treated participants in the PRO analysis population.

$$\text{CR-T} = \frac{Number\ of\ treated\ participants\ who\ complete\ at\ least\ one\ item}{Number\ of\ treated\ participants\ in\ the\ PRO\ analysis\ population}$$

The completion rate is expected to shrink in the later visit during study period due to the participants who discontinued early. Therefore, another measurement, compliance rate of eligible participants (CR-E) will also be employed as the support for completion rate. CR-E is defined as the number of treated participants who complete at least one item over number of eligible participants who are expected to complete the PRO assessment, not including the participants missing by design such as death, discontinuation, translation not available.

$$\text{CR-E} = \frac{Number\ of\ treated\ participants\ who\ complete\ at\ least\ one\ item}{Number\ of\ eligible\ participants\ who\ are\ expected\ to\ complete}$$

The reasons of non-completion and non-compliance will be provided in supplementary table:

- Completed as scheduled
- Not completed as scheduled
- Off-study: not scheduled to be completed.

In addition, reasons for non-completion as scheduled of these measures will be collected using "miss_mode" forms filled by site personnel and will be summarized in table format.

In study Part 1 Adjuvant Treatment, EORTC QLQ-C30 and EuroQoL EQ-5D-5L should be completed at baseline (Cycle 1), during treatment in year one (at Cycle 5, 9, 13, 17), every 12 weeks during year 2 (week 60, 72, 84, and 96 from baseline), and every 6 months during year 3 (month 30 and 36 from baseline). The schedule (study visits and estimated study times) and mapping of study visit to analysis visit for PRO data collection is provided in Table 6.

Table 6 PRO Data Collection Schedule and Mapping of Study visit to Analysis Visit in Part 1

	Treatment Week					**Follow-Up Visit** (every 12 weeks for 1st year from EOT, and every 6 months for 2nd year from EOT)					
Week	**0**	**12**	**24**	**36**	**48**	**60**	**72**	**84**	**96**	**Month 30**	**Month 36**
Cycle	C1	C5	C9	C13	C17						
Day	1	85	169	253	337	421	505	589	673	914	1096
Range	C1D1 of Part1	2 to 126	127 to 210	211 to 294	295 to 378	379 to 462	463 to 546	547 to 630	631 to 793	794 to 1005	1006 to 1187
C: Cycle; D: Day											

If the participant does not complete the PRO instruments at a scheduled time point, the site staff will record the reason the assessment was not performed. If there are multiple PRO collections within any of the stated time windows, the assessment completed closest to the target collection day will be used in the analyses.

3.6.3.2 Mean change from baseline

The prespecified time point for the mean change from baseline analysis is defined as the latest time point at which CR-T ≥ 60% and CR-E ≥ 80% based on blinded data review prior to the database lock for any PRO analysis and will be documented in a future sSAP amendment.

To assess the treatment effects on the PRO score change from baseline in the global health status/QoL, physical functioning and EQ-5D VAS outcome, a constrained longitudinal data analysis (cLDA) model proposed by Liang and Zeger [6] will be applied, with the PRO score as the response variable, and treatment, time, the treatment by time interaction, and stratification factors used for randomization (see protocol Section 6.3.1.1 – Stratification) as covariates. The treatment difference in terms of least square (LS) mean change from baseline will be estimated from this model together with 95% CI. Model-based LS mean with 95% CI will be provided by treatment group for PRO scores at baseline and post-baseline time point.

The cLDA model assumes a common mean across treatment groups at baseline and a different mean for each treatment at each of the post-baseline time points. In this model, the response vector consists of baseline and the values observed at each post-baseline time point. Time is treated as a categorical variable so that no restriction is imposed on the trajectory of the means over time. The cLDA model is specified as follows:

$$E(Y_{ijt}) = \gamma_0 + \gamma_{jt} I(t > 0) + \beta X_i, j = 1,2,,3,..,n;\ t = 0,1,2,3,..k$$

where Y_{ijt} is the PRO score for participant i, with treatment assignment j at visit t; γ_0 is the baseline mean for all treatment groups, γ_{jt} is the mean change from baseline for treatment group j at time t; X_i is the stratification factor (binary) vector for this participant, and β is the coefficient vector for stratification factors. An unstructured covariance matrix will be used to model the correlation among repeated measurements. If the unstructured covariance model fails to converge with the default algorithm, then Fisher scoring algorithm or other appropriate methods can be used to provide initial values of the covariance parameters. In the rare event that none of the above methods yield convergence, a structured covariance such as Toeplitz can be used to model the correlation among repeated measurements. In this case, the asymptotically unbiased sandwich variance estimator will be used. The cLDA model implicitly treats missing data as missing at random (MAR).

Line plots for the empirical mean change from baseline in EORTC QLQ-C30 global health status/QoL and physical functioning score will be provided across all time points as a supportive analysis.

In addition, the model-based LS mean change from baseline to the specified post-baseline time point together with 95% CI will be plotted in bar charts for EORTC QLQ-C30 global health status/quality of life scores, all functioning scores and all symptom scores.

3.6.3.3 Overall Improvement, Overall Improvement/Stability, Deterioration

Overall improvement rate will be analyzed, which is defined as the proportion of participants who have achieved an improvement as defined in Section 3.4.3 PRO Endpoints. Stratified Miettinen and Nurminen's method will be used for comparison of the overall improvement rate between the treatment groups. The difference in overall improvement rate and its 95% CI from the stratified Miettinen and Nurminen's method with strata weighting by sample size will be provided. The stratification factors used for randomization (See protocol Section 6.3.1.1 – Stratification) will be applied to the analysis.

The same method will be used to analyze overall improvement/stability rate, which is defined as the proportion of participants who have achieved improvement/stability as defined in Section 3.4.3 PRO Endpoints.

The point estimate of overall improvement rate, overall improvement/stability rate and deterioration rate will be provided by treatment group, together with 95% CI using exact binomial method by Clopper and Pearson (1934) [5].

3.6.3.4 Analysis Strategy for Key PRO Endpoints

Table 7 gives an overview of the analyses planned for key PRO endpoints.

Table 7 Analysis Strategy for Key PRO Endpoints

Endpoint/Variable	Statistical Method	Analysis Population	Missing Data Approach
Mean change from baseline in EORTC QLQ-C30 • Global health status/QoL • Physical functioning And EQ-5D VAS	cLDA model	FAS	Model-based.
Overall improvement and overall improvement/stability in EORTC QLQ-C30 • Global health status/QoL • Physical functioning	Stratified Miettinen and Nurminen method	FAS	Participants with missing data are considered not achieving improvement/stability.
Abbreviations: cLDA = constrained longitudinal data analysis, FAS = full analysis set, QoL = quality of life.			

3.6.4 Demographics and Baseline Characteristics

The comparability of the treatment groups for each relevant characteristic will be assessed by the use of tables and/or graphs. No statistical hypothesis testing will be performed on these characteristics. For Part 1, the number and percentage of participants randomized, and the primary reasons for discontinuation will be displayed. Demographic variables (e.g., age, gender) and baseline characteristics will be summarized by treatment either by descriptive statistics or categorical tables.

Impacts of COVID-19 will be evaluated or summarized if applicable. Discontinuations, protocol deviations, adverse events, and deaths associated with COVID-19 may be either summarized or listed as appropriate.

Baseline of Part 1 analysis is defined as the last non-missing assessment on or before the date of the 1st dose in Part 1.

3.7 Interim Analysis

The eDMC will serve as the primary reviewer of the results of the interim analyses and will make recommendations for discontinuation of the study or modification to an Executive Oversight Committee (EOC) of the Sponsor. Depending on the recommendation of the eDMC, the Sponsor may prepare a regulatory submission. Participant-level unblinding to support regulatory filing will be restricted to a designate team at the Sponsor, who will have no other responsibilities associated with the study.

If the eDMC recommends modifications to the design of the protocol or discontinuation of the study, this EOC may be unblinded to study results at the treatment level in order to act on these recommendations or facilitate regulatory filing. Limited additional Sponsor personnel may also be unblinded to the treatment level results of the IA(s), if required, in order to act on the recommendations of the eDMC or facilitate regulatory filing. The extent to which individuals are unblinded with respect to results of interim analyses will be documented. Additional logistical details, revisions to the above plan and data monitoring guidance will be provided in the eDMC Charter.

Treatment-level results of the efficacy interim analyses will be provided by an external unblinded statistician to the eDMC. Prior to final study unblinding, the external unblinded statistician will not be involved in any discussions regarding modifications to the protocol, statistical methods, identification of protocol deviations, or data validation efforts after the interim analyses.

3.7.1 Efficacy Interim Analyses

Five efficacy interim analyses and a final analysis are planned for this study. For all analyses, all randomized participants will be included. Results of the interim analyses will be reviewed by the DMC. Details of the boundaries for establishing statistical significance with regard to efficacy are discussed further in Section 3.8, Multiplicity.

The analyses planned, endpoints evaluated, and drivers of the timing are summarized in Table 8.

Table 8 Analyses Planned, Endpoints Evaluated, and Drivers of Timing

Analysis	Endpoint	Timing	Estimated Time after First Participant Randomized	Primary Purpose of Analysis
IA 1: Interim RFS analysis	RFS	(1) enrollment is completed, and (2) ~ 128 RFS events observed	~33 months	RFS IA
IA 2: Final RFS analysis	RFS	~179 RFS events observed	~48 months	RFS FA
IA 3: Interim DMFS analysis	DMFS	~146 DMFS events observed	~60 months	DMFS IA
IA 4: Final DMFS analysis;	DMFS	~195 DMFS events	~108 months	DMFS FA
IA 5: Interim OS analysis	OS	~154 OS events	~120 months	OS IA
FA: Final OS analysis	OS	~204 OS events	~180 months	OS FA
DMFS = distant metastatic-free survival; FA = final analysis; IA = interim analysis; OS = overall survival; RFS = recurrence-free survival.				

3.7.2 Safety Interim Analyses

The eDMC will conduct regular safety monitoring. The timing of the safety monitoring will be specified in the DMC charter.

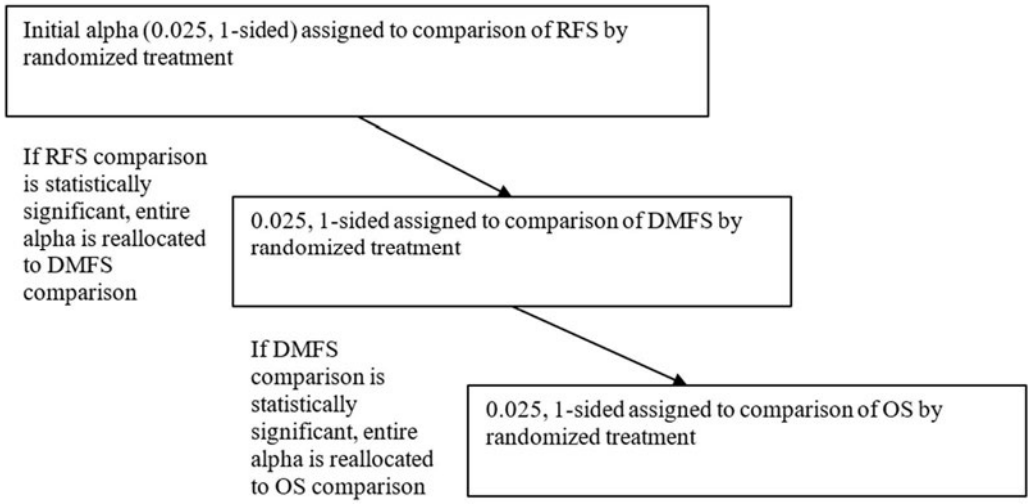
3.8 Multiplicity

The study uses the graphical method of Maurer and Bretz [4] to control multiplicity for multiple hypotheses as well as interim analyses. According to this approach, study hypotheses may be tested more than once, and when a particular null hypothesis is rejected, the alpha allocated to that hypothesis can be reallocated to other hypothesis tests.

The multiplicity strategy will be applied to the primary hypothesis and 2 secondary hypotheses. The primary hypothesis tests the superiority of pembrolizumab to placebo with respect to RFS. The 2 secondary hypotheses test the superiority of pembrolizumab to placebo with respect to DMFS and OS. The overall Type-I error among the 3 hypotheses is strongly controlled at 2.5% (one-sided), with 2.5% initially allocated to the RFS hypothesis. The study will be considered a success if RFS is demonstrated to be statistically significant at either an interim analysis or the final analysis under multiplicity control.

Figure 1 shows that the initial one-sided α allocation is assigned to the RFS hypothesis. Should the RFS comparison be statistically significant, the 2.5% alpha will be reallocated to the DMFS comparison. Should the DMFS comparison be statistically significant, the 2.5% alpha will be reallocated to the OS comparison.

Figure 1 Multiplicity Graph for Type I Error Control of Study Hypotheses



3.8.1 Recurrence-free Survival

The trial initially allocates α = 2.5%, one-sided to test RFS.

Table 9 shows the boundary properties for the interim analysis and final analysis, which were derived using a Lan-DeMets O'Brien-Fleming approximation spending function. Note that the final row indicates the total power to reject the null hypothesis for RFS. If the actual number of RFS events differs from that specified in the table, the bounds will be adjusted using the O'Brien-Fleming alpha-spending function accordingly.

Table 9 Boundary Properties for Planned Analyses of the RFS Analyses Based on α = 0.025

Analysis	Value	Efficacy
IA 1: 71%[1]	Z[2]	2.4115
N: 954	p (1-sided)[2]	0.0079
Events: 128	HR[3] at bound	0.6522
Month: 33	P(Cross)[4] if HR=1	0.0079
	P(Cross)[4] if HR=0.6	0.6717
Final (IA2)	Z	2.0029
N: 954	p (1-sided)	0.0226
Events: 179	HR at bound	0.7410
Month: 48	P(Cross) if HR=1	0.0250
	P(Cross) if HR=0.6	0.9190

[1] Percentage of total number of events expected at final analysis
[2] Boundary values for statistical significance
[3] HR= hazard ratio
[4] Probability of crossing boundary for statistical significance

3.8.2 Distant Metastases-free Survival

The trial initially allocates α=0.0, one-sided to test DMFS. If the null hypothesis for RFS is rejected, α=0.025 is fully reallocated to DMFS hypothesis testing.

Table 10 shows the boundary properties for the interim analysis and final analysis, which were derived using a Lan-DeMets O'Brien-Fleming approximation spending function. Note that the final row indicates the total power to reject the null hypothesis for DMFS. If the actual number of events at the DMFS analyses differ from those specified in the table, the bounds will be adjusted using the Lan-DeMets O'Brien-Fleming approximation spending function accordingly.

Table 10 Efficacy Boundaries and Properties for the DMFS Analyses

Analysis	Value	Efficacy
IA 3: 75%	Z	2.3401
N: 954	p (1-sided)	0.0096
Events: 146	HR at bound	0.6788
Month: 60	P(Cross) if HR=1	0.0096
	P(Cross) if HR=0.65	0.5987
Final (IA4)	Z	2.0117
N: 954	p (1-sided)	0.0221
Events: 195	HR at bound	0.7494
Month: 108	P(Cross) if HR=1	0.0250
	P(Cross) if HR=0.65	0.8425

3.8.3 Overall Survival

The study initially allocates α=0, one-sided to test OS. If the null hypothesis for DMFS is rejected, then α=0.025 is fully reallocated to OS hypothesis testing. Table 11 shows the boundary properties for the planned interim analysis (at 120 months from first participant enrolled) and final analysis (at 180 months), which were derived using a Lan-DeMets O'Brien-Fleming approximation spending function. Note that the final row indicates the total power to reject the null hypothesis for OS. If the actual number of events at the OS analyses differs from that specified in the table, the bounds will be adjusted using the Lan-DeMets O'Brien-Fleming approximation spending function accordingly.

Table 11 Efficacy Boundaries and Properties for the OS Analyses

Analysis	Value	Efficacy
IA 5: 76%	Z	2.3249
N: 954	p (1-sided)	0.0100
Events: 154	HR at bound	0.6875
Month: 120	P(Cross) if HR=1	0.0100
	P(Cross) if HR=0.67	0.5607
Final	Z	2.0138
N: 954	p (1-sided)	0.0220
Events: 204	HR at bound	0.7538
Month: 180	P(Cross) if HR=1	0.0250
	P(Cross) if HR=0.67	0.8050

3.8.4 Safety Analyses

The eDMC has responsibility for assessment of overall risk/benefit. When prompted by safety concerns, the eDMC can request corresponding efficacy data. External DMC review of efficacy data to assess the overall risk/benefit to trial participants will not require a multiplicity adjustment typically associated with a planned interim efficacy analysis; however, to account for any multiplicity concerns raised by the eDMC review of unplanned efficacy data prompted by safety concerns, a sensitivity analysis for RFS adopting a conservative multiplicity adjustment will be pre-specified in the sSAP. This analysis will be performed if efficacy data is requested by the eDMC during a safety evaluation to assess risk / benefit.

3.9 Sample Size and Power Calculations

The study will randomize approximately 954 participants in a 1:1 ratio into the pembrolizumab and placebo adjuvant treatment arms. RFS is the primary endpoint for the study, with DMFS and OS as the key secondary endpoints.

For RFS endpoint, the final analysis is event-driven and will be conducted after approximately 179 events have been observed, unless the study is terminated early. It may occur at ~ 48 months after the first participant is randomized (depending on enrollment rate and event accumulation rate). Based on a target number of 179 events at the final analysis and 1 interim analysis at approximately 71% of the target number of events, the study has ~92% power for detecting a hazard ratio of 0.6 at 2.5% (1-sided) significance level.

For DMFS endpoint, the final analysis is event-driven and will be conducted after approximately 195 events have been observed, unless the study is terminated early. It may occur at ~ 108 months after the first participant is randomized (depending on enrollment rate and event accumulation rate). Based on a target number of 195 events at the final DFMS analysis and 1 interim analysis at approximately 75% of the target number of events, the study has ~84% power for detecting a hazard ratio of 0.65 at 2.5% (1-sided) significance level.

For OS endpoint, the final analysis is calendar driven and will be conducted at approximately 15 years after the first participant is randomized, when we expect to have observed ~204 OS events. Based on the expected number of 204 events at the FA timing and 1 interim analysis at approximately 76% of the estimated number of events, the study has ~80% power for detecting a hazard ratio of 0.67 at 2.5% (1-sided) significance level.

The above sample size and power calculations are based on the following assumptions:

- RFS follows an “cure” model with a long-term RFS of 50% and the 60-month RFS estimated to be 68% for the control group.
- DMFS follows an “cure” model with a long-term DMFS of 65% and the 60-month DFMS estimated to be 78% for the control group.

- OS follows an exponential distribution with the 120-month OS estimated to be 75% for the control group.
- Enrollment period of 16 months
- An annual drop-out rate of 4.7%
- A follow-up period of 32, 92 and 164 months for RFS, DMFS and OS respectively, after the last participant is randomized.

The sample size and power calculations were performed in the software R (package "gsDesign").

3.10 Subgroup Analyses and Effect of Baseline Factors

To determine whether the treatment effect is consistent across various subgroups, the estimate of the between-group treatment effect (with a nominal 95% CI) for the primary endpoints will be estimated and plotted within each category of the following classification variables:

- T-Stage (T3b versus T4a versus T4b)
- Age (<65 years versus ≥65 years)
- Sex (male versus female)
- Race (White versus non-White)
- ECOG performance status (0 versus 1) or equivalent LPS status
- Region (US vs. Ex-US)

The consistency of the treatment effect will be assessed descriptively via summary statistics by category for the classification variables listed above. If the number of participants in a category of a subgroup variable is less than 10% of the ITT population, the subgroup analysis may not be performed for this category of the subgroup variable, and this subgroup variable may not be displayed in the forest plot. The subgroup analyses will be conducted using an unstratified Cox model.

3.11 Compliance (Medication Adherence)

Drug accountability data for study treatment will be collected during the study. Any deviation from protocol-directed administration will be reported.

3.12 Extent of Exposure

The extent of exposure will be summarized as duration of treatment in number of cycles or administrations as appropriate.

4 REFERENCES

1. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998; 16:139-44.

2. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Quality of Life Research 1996;5:555-67.

3. Miettinen O, Nurminen M. Comparative analysis of two rates. Statistics in Medicine 1985; 4:213-226.

4. Maurer W, Bretz F. Multiple Testing in Group Sequential Trials using Graphical Approaches. Statistics in Biopharmaceutical Research 2013; 5(4): 311-320.

5. Clopper CJ and Pearson ES. The Use of Confidence or Fiducial Limits illustrated in the Case of the Binomial. Biometrika, 1934: 26(4): 404-13

6. Liang K, Zeger, S (2000). Longitudinal data analysis of continuous and discrete responses for pre-post designs. Sankhyā: The Indian Journal of Statistics, 62 (Series B), 134-148.