16.1.9 Documentation of Statistical Methods

16.1.9.1 Statistical Analysis Plan

The Statistical Analysis Plan is found in the protocol in [16.1.1.4].

16.1.9.1.1 Supplemental Statistical Analysis Plan

(3475-ssap-p775-01)

Supplemental Statistical Analysis Plan (sSAP)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
AE	Adverse event
AEOSI	Adverse event of special interest
APaT	All Participants as Treated
AUC	Area under the concentration-time curve
BICR	Blinded independent central review
CCI [redacted]	[redacted]
CI	Confidence interval
CR	Complete response
CRO	Contract research organization
CTCAE	Common Terminology Criteria for Adverse Events
CCI [redacted]	[redacted]
DMC	Data monitoring committee
dMMR	Mismatch repair deficient
CCI [redacted]	[redacted]
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOC	Executive Oversight Committee
EORTC	European Organisation for the Research and Treatment of Cancer
FAS	Full Analysis Set
FDA	Food and Drug Administration
HRQoL	Health-Related Quality-of-Life
CCI [redacted]	[redacted]
IRT	Interactive response technology
ITT	Intention-to-treat
IV	Intravenous(ly)
MMR	Mismatch repair
MSD	Merck Sharp & Dohme Corp.
ORR	Objective response rate

OS	Overall survival
PD	Progressive disease
PFS	Progression-free survival
CCI	[redacted]
PK	Pharmacokinetic
pMMR	Mismatch repair proficient
PR	Partial response
Q3W	Every 3 weeks
QD	Once daily
QLQ-C30	Quality of life Questionnaire C30
QoL	Quality of life
RECIST	Response Evaluation Criteria In Solid Tumors
RMST	Restricted Mean Survival Time
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Stable disease
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TPC	Treatment of physician's choice

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. Analysis of pharmacokinetic data will be described in a separate SAP. In addition, analysis of blood and tumor biomarkers will be descried in a separate document.

2 SUMMARY OF CHANGES

The following changes made to the sSAP were not directly related to changes required due to a protocol amendment (MK-3475-775-07)

- Details of analysis of PRO data included.
- Details of analysis of all exploratory endpoints included in the sSAP
- Details of the alpha spending boundaries for the primary hypothesis evaluating the effect of treatment on OS in the All comer participants included in the sSAP

3 ANALYTICAL AND METHODOLOGICAL DETAILS

3.1 Statistical Analysis Plan Summary

Key elements of the statistical analysis plan are summarized below; the comprehensive plan is provided in Section 3.2 through Section 3.13.

Study Design Overview	A Multicenter, Open-label, Randomized, Phase 3 Trial to Compare the Efficacy and Safety of Lenvatinib in Combination with Pembrolizumab Versus Treatment of Physician’s Choice in Participants with Advanced Endometrial Cancer
Treatment Assignment	Approximately 780 eligible participants (660 mismatch repair proficient [pMMR] participants and 120 MMR deficient [dMMR] participants) will be randomized to one of the following 2 treatment arms in a 1:1 ratio: • Arm A: lenvatinib 20 mg (orally, QD) plus pembrolizumab 200 mg (IV Q3W) • Arm B: TPC consisting of either doxorubicin 60 mg/m^2 Q3W, or paclitaxel 80 mg/m^2 given weekly, 3 weeks on/1 week off Randomization will follow a predefined randomization scheme based on the following stratification factors: MMR status (pMMR or dMMR), ECOG performance status (0 or 1), geographic region (Region 1: Europe, USA, Canada, Australia, New Zealand, and Israel or Region 2: rest of the world), and prior history of pelvic radiation (yes or no). First, participants will be stratified according to MMR status. Then, only within the pMMR stratum, participants will be further stratified according to ECOG performance status, geographic region, and prior history of pelvic radiation. A total of 9 strata will be utilized for the study.
Analysis Populations	Efficacy: Intention to Treat (ITT) Safety: All Participants as Treated (APaT)
Primary Endpoints	• Progression-free survival (PFS) based on RECIST 1.1 as assessed by BICR. • Overall survival (OS).

Secondary Endpoints	• Objective response rate (ORR) by BICR using RECIST 1.1. • Health-Related Quality of Life using the EORTC QLQ-C30. • Safety and tolerability of the two treatment groups. • Plasma concentration of lenvatinib versus time. • Model-predicted clearance and AUC for lenvatinib.
Statistical Methods for Key Efficacy Analyses	The primary hypotheses will be evaluated by comparing in PFS and OS using a stratified Log-rank test. The hazard ratio (HR) will be estimated using a stratified Cox regression 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 results 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-group comparisons; 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	Two interim analyses are planned in this study and will be performed by an independent unblinded statistician and programmer. Results of these analyses will be reviewed by the DMC. Details are provided in Section 3.7. • Interim Analysis 1 (IA1) ○ Timing: to be performed after both ~368 OS events have been observed in the pMMR participants and at least 6 months after last participant randomized ○ Primary purpose: final efficacy analysis for PFS and interim efficacy analysis for OS • Interim Analysis 2 (IA2) ○ Timing: to be performed after both ~463 OS events have been observed in the pMMR participants and at least 12 months after last participant randomized ○ Primary purpose: interim efficacy analysis for OS • Final Analysis (FA) ○ Timing: to be performed after both ~526 OS events have been observed in the pMMR participants and at least 18 months after last participant randomized ○ Primary purpose: final efficacy analysis for OS
Multiplicity	The total family-wise error rate (Type-I error) among the dual-primary PFS and OS and the secondary ORR endpoints is strongly controlled at one-sided 0.025 level. A 0.0005 Type I error rate is initially allocated to test PFS and 0.0245 Type I error rate is initially allocated to test OS between two treatment arms in pMMR participants. Details of alpha allocation strategy among hypotheses of PFS, OS, and ORR are provided in Section 3.8 Multiplicity. The study will be considered positive if either testing of PFS or testing of OS is significant in pMMR participants.
Sample Size and Power	The planned sample size is approximately 780 participants (660 pMMR participants and 120 dMMR participants) with 330 pMMR participants and 60 dMMR participants in each arm. For the pMMR participants: With approximately 564 PFS events at the planned PFS analysis, the study will have at least 99% of power to detect a hazard ratio of 0.55 at the one-sided 0.0005 significance level. With approximately 368, 463, and 526 OS events in the pMMR participants at the planned IA1, IA2, and final OS analysis (FA), respectively, the study will have 90% power to detect a hazard ratio of 0.75 at the one-sided 0.0245 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 MSD.

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

Although the study is open label, analyses or summaries generated by randomized treatment assignment, and actual treatment received status will be limited and documented.

The external DMC will serve as the primary reviewer of the unblinded results of the interim analyses and will make recommendations for discontinuation of the study or modification to a Joint Executive Oversight Committee (EOC). Depending on the recommendation of the external DMC, the Sponsor or MSD may prepare a regulatory submission. If the external DMC recommends modifications to the design of the protocol or discontinuation of the study, this EOC and limited additional Sponsor/MSD personnel may be unblinded to results at the treatment level in order to act on these recommendations. Additional logistical details, revisions to the above plan and data monitoring guidance will be provided in the DMC charter.

3.3 Hypotheses/Estimation

In all randomized participants with advanced endometrial cancer:

Objective/Hypothesis	Endpoint
Primary	
• Objective: To demonstrate that lenvatinib in combination with pembrolizumab is superior to Treatment of Physician's Choice (TPC) in improving progression-free survival (PFS). Hypothesis (H1): The combination of lenvatinib and pembrolizumab is superior to TPC as assessed by PFS in pMMR participants. Hypothesis (H4): The combination of lenvatinib and pembrolizumab is superior to TPC as assessed by PFS in all-comer participants.	• PFS, defined as the time from date of randomization to the date of the first documentation of disease progression, as determined by blinded independent central review (BICR) per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1), or death from any cause, whichever occurs first.
• Objective: To demonstrate that lenvatinib in combination with pembrolizumab is superior to TPC in improving overall survival (OS). Hypothesis (H2): The combination of lenvatinib and pembrolizumab is superior to TPC as assessed by OS in pMMR participants. Hypothesis (H5): The combination of lenvatinib and pembrolizumab is superior to TPC as assessed by OS in all-comer participants.	• OS, defined as the time from date of randomization to date of death from any cause.
Secondary	
• Objective: To compare the objective response rate (ORR) of participants treated with lenvatinib in combination with pembrolizumab versus TPC by BICR Hypothesis (H3): The combination of lenvatinib and pembrolizumab is superior to TPC as assessed by ORR in pMMR participants.	• ORR, defined as the proportion of participants who have best overall response of either complete response (CR) or partial response (PR), as determined by BICR per RECIST 1.1.

Objective/Hypothesis	**Endpoint**
Hypothesis (H6): The combination of lenvatinib and pembrolizumab is superior to TPC as assessed by ORR in all-comer participants.	
• Objective: To evaluate the impact of treatment on Health-Related Quality of Life (HRQoL) as assessed by using the global score of the European Organisation for the Research and Treatment of Cancer (EORTC) QLQ-C30 for participants treated with lenvatinib in combination with pembrolizumab versus TPC in pMMR participants and in all-comer participants.	• HRQoL will be assessed using the global score of the EORTC QLQ-C30.
• Objective: To assess safety and tolerability of treatment with lenvatinib in combination with pembrolizumab versus TPC in pMMR participants and in all-comer participants.	• Incidence of treatment-emergent adverse events (TEAEs), serious AEs (SAEs), and immune-related AEs. • Proportion of participants discontinuing study treatment due to TEAEs. • Time to treatment failure due to toxicity, defined as the time from the date of randomization to the date that a participant discontinues study treatment due to TEAEs.
• Objective: To characterize the population pharmacokinetics (PK) of lenvatinib when co-administered with pembrolizumab in pMMR participants and in all-comer participants.	• Plasma concentration of lenvatinib versus time.
• Objective: To assess the relationship between exposure to lenvatinib and safety events related to lenvatinib in pMMR participants and in all-comer participants.	• Clearance and area under the concentration-time curve (AUC) for lenvatinib.
Exploratory	
CCI [redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

Objective/Hypothesis	Endpoint
CCI [redacted]	[redacted]

3.4 Analysis Endpoints

3.4.1 Efficacy Endpoints

3.4.1.1 Primary

- PFS by BICR - defined as the time from the date of randomization to the date of the first documentation of disease progression, as determined by blinded BICR of objective radiographic disease progression per RECIST 1.1 or death due to any cause (whichever occurs first). See Section 3.6.1 – Statistical Methods for Efficacy Analyses for definition of censoring.
- OS - defined as the time from the date of randomization to the date of death due to any cause. Participants who are lost to follow-up and those who are alive at the date of data cut-off will be censored at the date the participant was last known alive, or date of data cut-off, whichever occurs first.

3.4.1.2 Secondary

- ORR - defined as the proportion of participants who have best overall response of either CR or PR as determined by BICR per RECIST 1.1.
- HRQoL will be assessed using the global health status score of the EORTC QLQ-C30.
- Safety will be assessed summarizing the incidence of TEAEs, SAEs, and irAEs; proportion of participants who discontinued treatment due to TEAEs; and time to treatment failure due to toxicity (defined as the time from the date of randomization to the date that a participant discontinues study treatment due to TEAEs).
- Plasma concentration of lenvatinib versus time.
- Model-predicted clearance and AUC for lenvatinib

3.4.1.3 Exploratory

- CCI

[redacted]

[redacted]

- CCI [redacted]

[redacted]

3.4.2 Safety Endpoints

Safety endpoints are described in Section 3.6.2.

3.5 Analysis Populations

3.5.1 Efficacy Analysis Populations

The Intention-to-Treat (ITT) population will serve as the population for the primary efficacy analyses. All randomized participants will be included in this population. Participants will be analyzed in the treatment group to which they are randomized.

3.5.2 Safety Analysis Population

The All Participants as Treated (APaT) population will be used for the analysis of safety data in this study. The APaT population consists of all randomized participants who received at least 1 dose of study treatment. Participants will be included in the treatment group corresponding to the study treatment they actually received. 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 1 cycle, but receives the correct treatment for all other cycles, will be analyzed according to the correct treatment group and a narrative will be provided for any events that occur during the cycle for which the participant is incorrectly dosed.

At least 1 laboratory or vital sign measurement obtained subsequent to at least 1 dose of study treatment is required for inclusion in the analysis of each specific parameter. To assess change from baseline, a baseline measurement is also required.

3.5.3 Health Related Quality of Life Analysis Population

The HRQoL analyses are based on the HRQoL Full Analysis Set (PRO FAS) population, defined as all randomized participants who have at least one HRQoL assessment available for the specific endpoint and have received at least one dose of the study intervention. Participants will be analyzed in the treatment group to which they are randomized. There will be an all comers FAS and a pMMR FAS. Further details of each population are in [Sec. 3.13.5].

3.6 Statistical Methods

3.6.1 Statistical Methods for Efficacy Analyses

This section describes the statistical methods that address the primary, secondary and exploratory objectives. Efficacy results for pMMR participants and all-comer participants 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.

The stratification factors used for randomization (see [Sec. 3.6.1.1]) will be applied to all stratified analyses, in particular, the stratified log-rank test, stratified Cox model, and stratified Miettinen and Nurminen method [1]. 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.

Analyses will be performed in two subsets of subjects: All-comer participants and pMMR participants. In addition, select analyses may be performed for dMMR participants. All analyses performed in dMMR participants will be based on unstratified models for each endpoint. Although MMR status is a stratification factor in the trial, summary of pMMR and dMMR participants will be based on actual MMR status defined by immunohistochemistry (IHC) performed by a central vendor on tumor tissue provided by sites. If a participant is stratified as dMMR, but is determined to be pMMR by IHC, then stratification factors for the participant will be imputed based on clinical data.

3.6.1.1 Primary Efficacy Analysis

Progression-free Survival by BICR

The non-parametric Kaplan-Meier method will be used to estimate the PFS curve in each treatment group. The treatment difference in PFS 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 (i.e., HR) between the treatment arms. 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 Section <use section where randomization is described>) will be applied to both the stratified log-rank test and the stratified Cox model. Stratification factors are defined as follows:

Stratification

Treatment allocation/randomization will be stratified according to the following factors:

1. MMR status (pMMR or dMMR)
2. ECOG performance status (0 or 1)
3. Geographic region (Region 1 [Europe, USA, Canada, Australia, New Zealand, and Israel] or Region 2 [rest of the world])
4. Prior history of pelvic radiation (yes or no)

First, participants will be stratified according to MMR status. Then, only within the pMMR stratum, participants will be further stratified according to ECOG performance status, geographic region, and prior history of pelvic radiation. A total of 9 strata will be utilized for the study.

Since, stratification is layered in this study, first according to the MMR status for all subjects and then by ECOG, region and pelvic radiation history only within the pMMR stratum, the stratification will be different for the pMMR and all-comer analyses. All stratified analyses based on the all-comer population will include all 4 stratification variables in the model (9 strata), while the model for the pMMR population will include stratification variables for ECOG, region and pelvic radiation history (8 strata).

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 the date of the first assessment at which PD is objectively documented per RECIST 1.1 by BICR and the date of death. Death is always considered a PD event.

For the primary analysis, 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 participant meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied.

In order to evaluate the robustness of the PFS endpoint per RECIST 1.1 by BICR, two sensitivity analyses with different sets of censoring rules will be performed. The first sensitivity analysis follows the intention-to-treat principle. That is, PDs/deaths are counted as events regardless of missed study visits or initiation of new anti-cancer therapy. The second sensitivity analysis considers initiation of new anticancer treatment or discontinuation of treatment due to reasons other than complete response, whichever occurs later, to be a PD event for participants without documented PD or death. If a participant meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied. The censoring rules for the primary and sensitivity analyses are summarized in [Table 1].

Table 1 Censoring Rules for Primary Analysis of Progression-Free Survival Based on RECIST 1.1

Situation	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2
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	Progressed at date of documented PD or death	Progressed at date of documented PD or death
Death or progression immediately after ≥ 2 consecutive missed 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	Progressed at date of documented PD or death	Progressed at date of documented PD or death
No PD and no death; and new anticancer treatment is not initiated	Censored at last disease assessment	Censored at last disease assessment	Progressed at treatment discontinuation due to reasons other than complete response; otherwise censored at last disease assessment if still on study treatment or completed study treatment.
No PD and no death; new anticancer treatment is initiated	Censored at last disease assessment before new anticancer treatment	Censored at last disease assessment	Progressed at date of new anticancer treatment if new anti-cancer treatment is initiated; otherwise progressed at treatment discontinuation if treatment is discontinued due to reasons other than complete response; otherwise censored at last disease assessment if still on study therapy or completed the study therapy

Abbreviations: PD = progressive disease; RECIST = Response Evaluation Criteria in Solid Tumors.

PFS by investigator per RECIST 1.1 CCI [redacted] will be analyzed using the approach specified for the primary endpoint PFS BIRC above. Results based on the primary censoring rules for PFS summarized in [Table 1] above will be provided.

Overall Survival

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 (i.e., 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 [Sec. 3.6.1.1]) will be applied to both the stratified log-rank test and the stratified Cox model. Participants without documented death at the time of analysis will be censored at the date the participant was last known to be alive.

Evaluation of Proportional Hazards Assumption>

The proportional hazards assumption on OS may be examined using both graphical and analytical methods if warranted. The log [-log] of the survival function vs. time for OS will be plotted for the comparison between lenvatinib plus pembrolizumab and the TPC arm. If the curves are not parallel, indicating that hazards are not proportional, supportive analyses may be conducted to account for the possible non-proportional hazards effect associated with immunotherapies: for example, using the Restricted Mean Survival Time method [2], parametric method [3], etc.

The RMST is the population average of the amount of event-free survival time experienced during a fixed study follow-up time. This quantity can be estimated by the area under the Kaplan-Meier curve up to the follow-up time. The clinical relevance and feasibility should be taken into account in the choice of follow-up time to define RMST (e.g., near the last observed event time assuming that the period of clinical interest in the survival experience is the whole observed follow-up time for the study, but avoiding the very end of the tail where variability may be high); a description of the RMST as a function of the cutoff time may be of interest. The difference between two RMSTs for the two treatment groups will be estimated and 95% CI will be provided.

Evaluation of Crossover

Adjustment for the effect of crossover on OS may be performed based on recognized methods, e.g. the Rank Preserving Structural Failure Time (RPSFT) model proposed by Robins and Tsiatis (1989) [4], two stage model proposed by Latimer et al. (2019) [5], or other methods based on an examination of the appropriateness of the data to the assumptions required by the methods.

3.6.1.2 Secondary Efficacy Analysis

Objective Response Rate (ORR) by BICR

The stratified Miettinen and Nurminen's method will be used for the comparison of ORR between two treatment groups. The difference in ORR and its 95% CI from the stratified Miettinen and Nurminen's method with strata weighting by sample size will be reported. The stratification factors used for randomization (See [Sec. 3.6.1.1]) will be applied to the analysis. The point estimate of ORR will be provided by treatment group, together with 95% CI using exact binomial method proposed by Clopper and Pearson (1934) [6].

CCI [redacted]

3.6.1.3 Exploratory Efficacy Analysis

3.6.1.3.1 CCI

CCI

[redacted]

[redacted]

CCI

3.6.1.3.2 CCI

CCI

CCI

[redacted]

3.6.2 Safety

Safety and tolerability will be assessed by clinical review of all relevant parameters including AEs, laboratory tests, vital signs, and ECG measurements. The analysis of safety results will follow a tiered approach ([Table 4]). The tiers differ with respect to the analyses that will be performed. AEs (specific terms as well as system organ class terms) and events that meet predefined limits of change (PDLCs) in laboratory values, vital signs and ECG parameters are either prespecified as “Tier 1” endpoints or will be classified as belonging to “Tier 2” or “Tier 3” based on the observed proportions of participants with an event.

Tier 1 Events

Safety parameters that are identified a priori constitute “Tier 1” safety endpoints that will be subject to inferential testing for statistical significance with p-values and 95% confidence intervals provided for between-group comparisons. AEs of special interest (AEOSIs) 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. Further, the combination of pembrolizumab and lenvatinib has not been associated with any new safety signals. Additionally, there are no known AEs associated with participants for which determination of a p-value is expected to impact the safety assessment. Therefore, there are no Tier 1 events in this study.

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 [1].

Membership in Tier 2 requires that at least 10% of participants in any treatment group exhibit the event; all other AEs and predefined limits of change will belong to Tier 3. The threshold of at least 10% of participants was chosen for Tier 2 events because the population enrolled in this study is in critical condition and usually experiences various AEs of similar types regardless of treatment; events reported less frequently than 10% of participants would obscure the assessment of the overall safety profile and add little to the interpretation of potentially meaningful treatment differences. In addition, Grade 3 to 5 AEs (≥5% of participants in 1 of the treatment groups) and SAEs (≥1% of participants in 1 of the treatment groups) will be considered Tier 2 endpoints. Because many 95% CIs may be provided without adjustment for multiplicity, the CIs should be regarded as a helpful descriptive measure to be used in safety review, not as 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. The broad AE categories consisting of the proportion of participants with any AE, a drug related AE, a serious AE, an AE which is both drug-related and serious, a Grade 3-5 AE, a drug-related Grade 3-5 AE, and discontinuation due to an AE will be considered Tier 3 endpoints. Only point estimates by treatment group are provided for Tier 3 safety parameters.

Table 4 Analysis Strategy for Safety Parameters

Safety Tier	Safety Endpoint	p-Value	95% CI for Treatment Comparison	Descriptive Statistics
Tier 2	Grade 3-5 AE (incidence ≥5% of participants in one of the treatment groups)		X	X
	Serious AE (incidence ≥1% of participants in one of the treatment groups)		X	X
	AEs (incidence ≥10% of participants in one of the treatment groups)		X	X
Tier 3	Any AE			X
	Any Grade 3-5 AE			X
	Any Serious AE			X
	Any Drug-Related AE			X
	Any Serious and Drug-Related AE			X
	Any Grade 3-5 and Drug-Related AE			X
	Discontinuation due to AE			X
	Death			X
	Specific AEs, SOCs (incidence >0% of participants in all of the treatment groups)			X
	Change from Baseline Results (lab toxicity shift, vital signs)			X

Abbreviations: AE = adverse event; CI = confidence interval; SOC = system organ class.

Exposure-adjusted rate of AE by time period from first dose (e.g., 0-3, 3-6, 6-12 mos) may also be provided. In each time interval, the denominator person years of exposure is calculated based on the number of participants at risk for the event during the particular time period, where at risk is defined as participants who are exposed to drug at the start of indicated time interval.The numerator is the number of events occurring in the interval.

In addition, to properly account for the potential difference in follow-up time between the study arms, which is expected to be longer in the lenvatinib plus pembrolizumab arm, AE incidence adjusted for treatment exposure analyses will be performed. For exposure adjusted analyses, events count as the numerator, and person-months of exposure as the denominator.

3.6.3 Demographic and Baseline Characteristics

The comparability of the treatment groups for each relevant demographic and baseline characteristic will be assessed by the use of tables. No statistical hypothesis tests will be performed on these characteristics. The number and percentage of participants screened and randomized and the primary reasons for screening failure and discontinuation will be displayed. Demographic variables, baseline characteristics, primary and secondary diagnoses, and prior and concomitant therapies will be summarized by treatment either by descriptive statistics or categorical tables. Select tables and/or figures summarizing demographic and baseline characteristics may be generated based on the pMMR and dMMR participants.

3.7 Interim Analyses

3.7.1 Safety Interim Analyses

The safety monitoring and efficacy interim analyses will be conducted by the external DMC. Minutes from the open meetings of the DMC will be provided if requested by regulatory agencies. The recommendation whether to stop the study will be reached by the DMC based on their review of data with treatment information. The function and membership of the DMC will be described in the DMC charter.

Access to the allocation schedule for summaries or analyses for presentation to the eDMC will be restricted to an unblinded internal statistician, and, as needed, an internal scientific programmer performing the analysis, who will have no other responsibilities associated with the study.

3.7.2 Efficacy Interim Analyses

Two interim analyses are planned in addition to the final analysis for this study. The interim analyses will be performed by an independent unblinded statistician and programmer. Results of these analyses will be reviewed by the external DMC.

The PFS analysis will be performed only at the time of the first interim analysis of the study and this will be the final PFS analysis. Since the timing of the first interim analysis is driven by the required number of OS event, the observed number of PFS events may be different from the expected counts. OS analyses will be performed at the first interim, second interim, and final analysis. The Lan-DeMets spending function with O'Brien-Fleming boundary will be used for alpha allocation among the interim and final analyses of OS. Details of the boundaries for establishing statistical significance with regard to efficacy are discussed in Section 3.8. The analyses planned, endpoints evaluated, and drivers of timing are summarized in [Table 5].

Table 5 Summary of Interim and Final Analysis Strategy for the pMMR Participants

Analyses	Key Endpoints	Timing	Estimated Time after First Participant Randomized	Primary Purpose of Analysis
IA1	PFS OS	Both ~368 OS events and at least 6 months after last participant randomized	~27 months	Final PFS analysis Interim OS analysis
IA2	OS	Both ~463 OS events and at least 12 months after last participant randomized	~35 months	Interim OS analysis

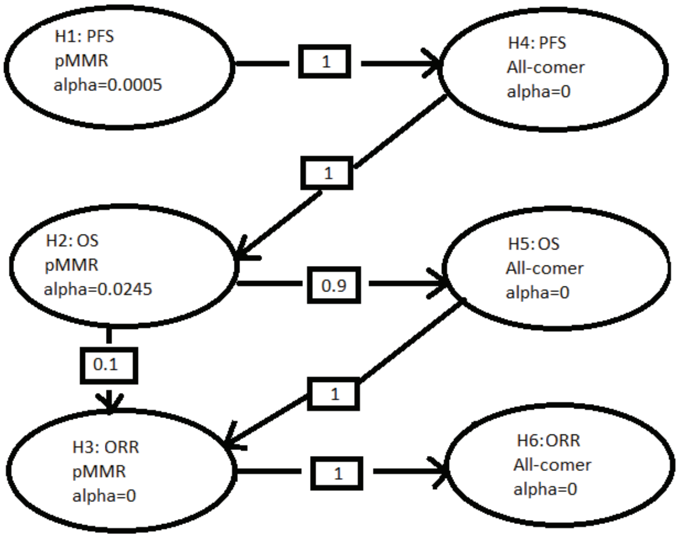
Analyses	Key Endpoints	Timing	Estimated Time after First Participant Randomized	Primary Purpose of Analysis
FA	OS	Both ~526 OS events and at least 18 months after last participant randomized [†]	~43 months [†]	Final OS analysis
Abbreviations: FA = final analysis; IA1 = interim analysis 1; IA2 = interim analysis 2; OS = overall survival; PFS = progression-free survival; pMMR = mismatch repair proficient. [†] Note that if events accrue slower than expected for the FA, the Sponsor may conduct the analysis up to 3 months after the estimated timing of the FA (ie., ~46 months after first participant randomized).				

As described in Section 3.8, the PFS and OS analyses for the all-comer participants will be performed as listed in [Table 5] if the respective analyses are successful in the pMMR participants.

3.8 Multiplicity

The total family-wise error rate (Type-I error) among the dual-primary PFS and OS and the secondary ORR endpoints is strongly controlled at one-sided 0.025 level. The multiplicity strategy will follow the graphical approach of Maurer and Bretz [7]. [Figure 1] shows the initial one-sided α-allocation for each hypothesis in the ellipse representing the hypothesis. The initial weights for reallocation from each hypothesis to the others are represented in the boxes on the lines connecting hypotheses.

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



Abbreviations: ORR = objective response rate; OS = overall survival; PFS = progression-free survival; pMMR = mismatch repair proficient.

The study initially allocates α = 0.0005, one-sided, to test PFS for pMMR participants and initially allocates α =0.0245, one-sided to test OS for pMMR participants between the two treatment arms. As shown in [Figure 1], if the null hypothesis for PFS for pMMR is rejected, α = 0.0005 will be passed to the test for PFS for all-comer participants. And if the null hypothesis for PFS for all-comer participants is rejected, α = 0.0005 will be passed to the test for OS for pMMR, therefore OS for pMMR will be tested at α =0.025. The study will be considered positive if either testing of PFS or testing of OS is significant in pMMR participants.

[Table 6] shows the bounds and boundary properties for OS hypothesis testing derived using a Lan-DeMets spending function approximating O'Brien-Fleming bounds. The bounds provided in the table assume that the expected number of OS events at IA1, IA2, and FA are 368, 463, and 526, respectively. At the time of an analysis, the observed number of events may differ from the expected. To avoid overspending at an interim analysis and leave reasonable alpha for the final analysis, the minimum alpha spending strategy will be adopted. At an IA, the information fraction used in Lan-DeMets spending function to determine the alpha spending at the IA will be based on the minimum of the expected information fraction and the actual information fraction at each analysis. Specifically:

- In the scenario that the events accrue slower than expected and the observed number of events is less than the expected number of events at a given analysis, the information fraction will be calculated as the observed number of events at the interim analysis over the target number of events at FA.

- In the scenario that the events accrue faster than expected and the observed number of events exceeds the expected number of events at a given analysis, the information fraction will be calculated as the expected number of events at the interim analysis over the target number of events at FA.

The final analysis will use the remaining Type I error that has not been spent at the earlier analyses. The observed event counts for all analyses will be used to compute correlations.

Of note, while the information fraction used for the alpha spending calculation will be the minimum of the actual information fraction and the expected information fraction, the correlations required for deriving the bounds will still be computed using the actual information fraction based on the observed number of events at each analysis over the target number of events at FA.

The minimum spending approach assumes timing is not based on any observed Z-value and thus the Z test statistics used for testing conditioned on timing are multivariate normal. Given the probabilities derived with the proposed spending method, the correlations based on actual event counts are used to compute bounds that control the Type I error at the specified alpha level for a given hypothesis conditioned on the interim analysis timing. Since this is true regardless of what is conditioned on, the overall Type I error for a given hypothesis unconditionally is controlled at the specified level. By using more conservative spending early in the study, power can be retained to detect situations where the treatment effect may be delayed.

Table 6 Boundary Properties for Planned Analyses of OS Based on Potential Alpha-Levels to be Used for Testing in the pMMR Participants

Analysis	Value	α =0.0245	α =0.025
IA1 N: 660 OS events: 368 (70%*) Month: 27	Z	2.448	2.440
	p (1-sided) [†]	0.0072	0.0073
	HR at bound [‡]	0.7747	0.7753
	P(Cross) if HR=1 [§]	0.0072	0.0073
	P(Cross) if HR=0.75 [ǀ]	0.6234	0.6259
IA2 N: 660 OS Events: 463 (88%*) Month: 35	Z	2.187	2.178
	p (1-sided) [†]	0.0144	0.0147
	HR at bound [‡]	0.8160	0.8167
	P(Cross) if HR=1 [§]	0.0165	0.0169
	P(Cross) if HR=0.75 [ǀ]	0.8260	0.8285
FA N: 660 OS Events: 526 Month: 43	Z	2.069	2.061
	p (1-sided) [†]	0.0193	0.0196
	HR at bound [‡]	0.8348	0.8355
	P(Cross) if HR=1 [§]	0.0245	0.0250
	P(Cross) if HR=0.75 [ǀ]	0.9009	0.9025

Abbreviation: HR = hazard ratio; IA= interim analysis; FA= final analysis.

The number of events and timings are estimated.

[*] Percentage of total planned events at the interim analysis.

[†] *p* (1-sided) is the nominal α for group sequential testing.

[‡] HR at bound is the approximate observed HR required to reach an efficacy bound.

[§] P(Cross) if HR=1 is the probability of crossing a bound under the null hypothesis.

[ǀ] P(Cross) if HR=0.75 is the probability of crossing a bound under the alternative hypothesis.

[Table 7] shows the bounds and boundary properties for OS hypothesis (H5 all comer participants) testing derived using a Lan-DeMets spending function approximating O'Brien-Fleming bounds. The bounds provided in the table assume that the expected number of OS events at IA1, IA2, and FA are 433, 544, and 618, respectively. At the time of an analysis, the observed number of events may differ from the expected. To avoid overspending at an interim analysis and leave reasonable alpha for the final analysis, the minimum alpha spending strategy will be adopted. At an IA, the information fraction used in Lan-DeMets spending function to determine the alpha spending at the IA will be based on the minimum of the expected information fraction and the actual information fraction at each analysis.

Table 7 Boundary Properties for Planned Analyses of OS Based on Potential Alpha-Levels to be Used for Testing in the All-comer Participants at pMMR Participant Analysis Time Points

Analysis	Value	α =0.02205	α =0.0225
IA1 N: 780 OS events: 433 (70%*) Month: 27	Z	2.5000	2.4901
	p (1-sided) [†]	0.0062	0.0064
	HR at bound [‡]	0.7862	0.7870
	P(Cross) if HR=1 [§]	0.0062	0.0064
	P(Cross) if HR=0.75 [‖]	0.6890	0.6927
IA2 N: 780 OS Events: 544 (88%*) Month: 35	Z	2.2318	2.2222
	p (1-sided) [†]	0.0128	0.0131
	HR at bound [‡]	0.8257	0.8263
	P(Cross) if HR=1 [§]	0.0147	0.0150
	P(Cross) if HR=0.75 [‖]	0.8750	0.8769
FA N: 780 OS Events: 618 Month: 43	Z	2.1109	2.1030
	p (1-sided) [†]	0.0174	0.0177
	HR at bound [‡]	0.8437	0.8443
	P(Cross) if HR=1 [§]	0.0221	0.0225
	P(Cross) if HR=0.75 [‖]	0.9354	0.9365

Abbreviation: HR = hazard ratio; IA= interim analysis; FA= final analysis.

The number of events and timings are estimated.

[*] Percentage of total planned events at the interim analysis.

[†] *p* (1-sided) is the nominal α for group sequential testing.

[‡] HR at bound is the approximate observed HR required to reach an efficacy bound.

[§] P(Cross) if HR=1 is the probability of crossing a bound under the null hypothesis.

[‖] P(Cross) if HR=0.75 is the probability of crossing a bound under the alternative hypothesis.

The study will continue if the interim OS analyses are not statistically significant. ORR will be formally tested based on IA1 only. If the IA2 or final OS analysis is statistically significant, ORR at IA1 will be tested at the time of IA2 or final OS analysis to protect family-wise type I error rate.

3.9 Sample Size and Power Calculations

The sample size is estimated based on the primary endpoints PFS and OS. A total of approximately 780 participants (including 660 participants from pMMR and 120 participants from dMMR participants) will be randomized in a 1:1 ratio (approximately 330 participants from pMMR and 60 participants from dMMR participants in each treatment arm).

The study will have been considered to have completed enrollment when 660 pMMR participants have enrolled. Enrollment of dMMR participants will be capped at 120.

Sample size and power calculations are based on pMMR participants:

The study is designed to have 90% power to detect a statistically significant difference in OS at one-sided α=0.0245 and as a result, the study will also have at least 99% power to detect a statistical significant difference in PFS at one-sided α=0.0005.

Assuming an accrual period of 19 months and a follow-up period of 24 months, a total of 660 participants are required to observe 526 death events by the time of 43 months after the first participant is randomized (19 months enrollment plus 24 months follow-up period).

For OS, a total of 526 OS events are required to detect a statistically significant difference at 0.0245 level with 90% power, under the following assumptions that: 1) the hazard ratio is 0.75 (median OS is 16.4 months in Arm A and 12.3 months in Arm B), 2) the first interim analysis is performed when approximately 368 OS events are observed (i.e. 70% of the total target death events), 3) the second interim analysis is performed when approximately 463 OS events are observed (i.e. 88% of the total target death events), and 4) Lan-DeMets spending function with O'Brien-Fleming boundary is used.

The final PFS analysis is planned to be performed at the time of the first OS interim analysis (IA1) at 27 months after the first participant is randomized. A total of 564 PFS events are estimated to be observed to detect a statistically significant difference at 0.0005 level with >99% power under the assumption that the hazard ratio is 0.55 (median PFS is 7.3 months in Arm A and 4 months in Arm B).

Power calculations are based on pMMR and dMMR participants combined (all comer):

Assuming an accrual period of 19 months and a follow-up period of 24 months, a total of 780 participants are required in the all comer population to observe 618 death events by the time of 43 months after the first participant is randomized (19 months enrollment plus 24 months follow-up period).

For OS, a total of 618 OS events are required to detect a statistically significant difference at 0.02205 level with 93.5% power, under the following assumptions that: 1) the hazard ratio is

0.75 (median OS is 16.4 months in Arm A and 12.3 months in Arm B), 2) the first interim analysis is performed when approximately 433 OS events are observed (i.e. 70% of the total target death events), 3) the second interim analysis is performed when approximately 544 OS events are observed (i.e. 88% of the total target death events), and 4) Lan-DeMets spending function with O'Brien-Fleming boundary is used.

3.10 Subgroup Analyses and Effect of Baseline Factors

To determine whether the treatment effect is consistent across various subgroups, both efficacy and safety may be analyzed for the following subgroups as appropriate. For efficacy endpoints, the hazard ratio and two-sided 95% confidence interval (CI) for comparing PFS as assessed by BICR and OS of Arm A vs. Arm B will be presented in forest plots for the subgroups. Median PFS and 95% CIs will be presented for all subgroups. Similar plots will be provided for OS. The difference in ORR and 2-sided 95% CI for comparing ORR as assessed by BICR between Arm A vs Arm B will be presented in forest plots for the subgroup. For PFS, OS, and ORR, the following subgroups will be summarized:

- Age (<65,≥65)
- Age (<65, ≥65 to <75, ≥75 to <85, ≥85)
- Race (White, Asian, Other)
- ECOG Status (0, 1)
- Region (Region 1, Region2)
- Prior History of Pelvic Radiation (Yes, No)
- Histology (Endometrioid, Non-endometrioid)
- Prior Lines of Therapy (1, 2, ≥3)
- MMR Status (pMMR, dMMR)

For safety endpoints, all TEAEs, TEAEs of CTCAE Grades 3-5, and treatment-emergent SAEs will be summarized by the following subgroups.

- Age (<65,≥65)
- Age (<65, ≥65 to <75, ≥75 to <85, ≥85)
- Race (White, Asian, Other)
- ECOG Status (0, 1)
- Region (Region 1, Region2)
- Region (US, ex-US)
- Region (EU, ex-EU)
- Renal Function Category (CrCl < 60 mL/min, >= 60mL/min)
- Hepatic Function Category (Normal, Abnormal)
- MMR Status (pMMR, dMMR)

All efficacy subgroup analyses will be performed in the pMMR and all comer subjects. Safety subgroup analysis will be provided based on the all comer participants.

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 for lenvatinib will be summarized as duration of treatment in days. The extent of exposure for pembrolizumab will be summarized as duration of treatment in cycles. Dose interruption for each drug, dose reduction for lenvatinib will be summarized. Summary statistics will be provided on Extent of Exposure for the APaT population.

3.13 Statistical Consideration for Patient-Reported Outcomes (PRO)

This section provides the statistical consideration for evaluating PRO data that will be included in the CSR.

The patient-reported outcomes are secondary and exploratory objectives in the trial. No formal hypotheses were formulated. Nominal p-value to compare lenvatinib in combination with pembrolizumab versus TPC in pMMR participants and in all-comer participants may be provided as appropriate.

The PRO instruments are EORTC QLQ-C30 (global health status and physical function), CCI [redacted]

3.13.1 Completion and Compliance Rate Summary for PROs

Completion and compliance of QLQ-C30, CCI [redacted] by visit and by treatment will be described. Numbers and percentages of complete and missing data at each visit will be summarized for each of the treatment groups. 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 is defined as the percentage of number of subjects who complete at least one item over the number of subjects in the FAS population.

$$Completion\ Rate = \frac{Number\ of\ Subjects\ who\ Complete}{Number\ of\ Randomized\ Subjects\ who\ are\ in\ FAS\ population}$$

The completion rate is expected to shrink in the later visit during study period due to the subjects who discontinued. Therefore, another measurement, Compliance Rate, defined as the percentage of observed visit over number of eligible subjects who are expected to complete the PRO assessment (not including the subjects missing by design (such as death, discontinuation, translation not available) will be employed as the support for completion rate.

$$Compliance\ Rate = \frac{Number\ of\ Subjects\ who\ Complete}{Number\ of\ Eligible\ Subjects\ who\ are\ Expected\ to\ Complete}$$

Reasons for non-completion will be summarized.

3.13.2 PRO Endpoints (Additional details on analyses methods and population included in Table 8 and the section below):

Unless otherwise specified below, the primary timepoint for assessment for the analyses of PRO is the latest timepoint at which PRO data for both groups was collected, and the overall completion is at least 60% and the overall compliance is at least 80%. Because the cycle length of the treatments assessed in this study can be 21 days (3 weeks) or 28 days (4 weeks), the primary assessment timepoint will be a multiple of 12 weeks (i.e., either 12, 24 or 36 weeks). A blinded data review prior to the database lock will be conducted to assess the completion rate and compliance rate and to determine the primary assessment time point.

- The mean score changes from baseline to the latest time point where the completion rate and compliance rates is still high enough (e.g. close to 60 and 80%, respectively) based on blinded data review as measured by the EORTC QLQ-C30 global health status/quality of life scale.
- The mean score change from baseline to the latest time point where the completion and compliance rates is still high enough (e.g. close to 60 and 80%, respectively) based on blinded data review for the following QLQ-C30 sub-scales/item : QLQ-C30 physical functioning.
- The mean score changes from baseline to the latest time point where the completion rate and compliance rates is still high enough (e.g. close to 60 and 80%, respectively) based on blinded data review for the Urological symptoms score of the CCI [redacted]
- The mean score change from baseline to the latest time point where the completion rate and compliance rates is still high enough (e.g. close to 60 and 80%, respectively) based on blinded data review for CCI [redacted]

3.13.3 Analyses Methods

Mean score change from baseline

To assess the treatment effects on the PRO score change from baseline in in EORTC QLQ-C30 Global Health Status Scale, Physical Functioning Scale; CCI [redacted] 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 [Sec. 3.6.1.1]) 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. These timepoints are consistent with the information included in the [Table 9] and [Table 10] below.

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:

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

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).

PRO Empirical Mean Change from Baseline

Line plots for the empirical mean change from baseline in EORTC QLQ-C30 Global Health Status Scale, Physical Functioning Scale; CCI [redacted] will be provided across all time points (as indicated in [Table 9] and [Table 10] below) up to the final assessment timepoint 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 and physical function scale.

[Table 8] gives an overview of the analyses planned for all PRO endpoints.

Table 8 Planned Statistical Analysis

Endpoint	Analysis	Primary Statistical Method	Report
Score change from baseline	Treatment effect estimation/comparison	Mixed effect model based on the missing at random (MAR) assumption.	lsmean score (95% C.I.) by treatment group and visit, lsmean score change (95% C.I.) from baseline by treatment group and visit, between-group difference in score change from baseline (95% C.I., p-value).

3.13.4 Analysis Population

The All-comer Full Analysis Set (FAS) consists of all randomized participants who have received at least one dose of study medication, and have completed at least one PRO assessment beyond baseline.

Participants in the All-comer Full Analysis Set who have pMMR status will be included in the pMMR Full Analysis Set.

Unless otherwise specified, all the analyses will be performed for the pMMR Full Analysis Set as well as the All-comer Full Analysis Set.

3.13.5 The schedule for PRO data collection:

Table 9 PRO Data Collection Schedule

Treatment	**Week**									**EOT Discontinuation Visit**	**Follow-up Visit**
	0	3	6	9	12	15	18	21	24		
Lenvatinib plus pembrolizumab or TPC of Doxorubicin	C1	C2	C3	C4	C5	C6	C7	C8	C9	X[1]	X

C: Cycle; D: Day

Each cycle is 3 weeks for Lenvatinib plus pembrolizumab or TPC of Doxorubicin.

[1]: Continue every 3 weeks to EOT discontinuation visit.

Treatment	Week									EOT Discontinuation Visit	Follow-up Visit
	0	4	8	12	16	20	24	28	32		
TPC of Paclitaxel	C1	C2	C3	C4	C5	C6	C7	C8	C9	X[2]	X

C: Cycle; D: Day

Each cycle is 4 weeks for TPC of Paclitaxel.

[2]: Continue every 4 weeks to EOT discontinuation visit.

The general rule of mapping relative day to analysis visit is provided in [Table 10].

Table 10 Mapping Relative Day to Analysis Visit

	Week									
	0	3	6	9	12	15	18	21	24	Week X
Day	1	22	43	64	85	106	127	148	169	Week number *7+1
Range	-7 to 1	2-32	33-53	54- 74	75-95	96-116	117-137	138-158	159-179	Week number *7 -9 to Week number *7 +11

At each scheduled visit, three instruments, EORTC QLQ-C30, CCI will be collected. If a patient does not complete the PRO instruments, the site staff will record the reason for the missing from pre-defined choices. If there are multiple PRO collections within any of the stated time windows, the closest collection to the target day will be used.

3.14 References

1. Miettinen O, Nurminen M. Comparative analysis of two rates. Stat Med 1985;4:213-26.
2. Uno, H., et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. J Clin Oncol. 2014; 32 (22): 2380-2385.
3. Anderson KM. A nonproportional Hazards Weibull Accelerated Failure Time Regression Model. Biometrics 1991; 47: 281-288.
4. Robins, J.M., Tsiatis, A.A. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics-Theory and Methods, 1991; 20 (8): 2609-2631.
5. Latimer N.R., Abrams K.R., Siebert U. Two-stage estimation to adjust for treatment switching in randomised trials: a simulation study investigating the use of inverse probability weighting instead of re-censoring. BMC Medical Research Methodology. 2019; 19:69
6. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26(4):404-13.
7. Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. Stat Biopharm Res 2013;5(4):311-20.
8. The EORTC QLQ-C30 Manuals. Reference Values and Bibliography.

CCI

CCI

Appendix 1: Technical Details

3.15 PRO Instruments:

Scoring Algorithm:

QLQ-C30 Scoring: For each scale or item, a linear transformation will be applied to standardize the score as between 0 and 100, according to the corresponding scoring standard. For functioning and global health status/quality-of-life scales, a higher value indicates a better level of function; for symptom scales and items, a higher value indicates increased severity of symptoms.

According to the QLQ-C30 Manuals, if items $I_1, I_2, \ldots, I_n$ are included in a scale, the linear transformation procedure is as follows:

1. Compute the raw score: $RS = (I_1 + I_2 + \ldots + I_n) / n$

2. Linear transformation to obtain the score S:

$$\text{Function scales: } S = \left(1 - \frac{RS - 1}{Range}\right) \times 100$$

$$\text{Symptom scales/items: } S = \frac{RS - 1}{Range} \times 100$$

$$\text{Global health status/QoL: } S = \frac{RS - 1}{Range} \times 100$$

Range is the difference between the maximum possible value of RS and the minimum possible value. If more than half of the items within one scale are missing, then the scale is considered missing, otherwise, the score will be calculated as the average score of those available items [8].

CCI [redacted]

A linear transformation will be applied to standardize the scores between 0 (worst) and 100 (best) as described above for the EORTC QLQ-C30 Scoring.