Official Title: A Phase 2, Randomized, Double-Blind Study of Pembrolizumab (MK-3475) Plus Epacadostat (INCB024360) Versus Pembrolizumab Plus Placebo as First-Line Treatment in Patients With Metastatic Non-Small Cell Lung Cancer Expressing High Levels of PD-L1

NCT Number: NCT03322540

Document Date: Supplemental Statistical Analysis Plan Version 2: 04 February 2019

16.1.9 Documentation of Statistical Methods

16.1.9.1 Statistical Analysis Plan

The statistical analysis plan is found in the protocol in [16.1.1.1].

16.1.9.1.1 Supplemental Statistical Analysis Plan

(3475-ssap-654-04-amend)

Supplemental Statistical Analysis Plan (sSAP)

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. Changes to [redacted] other non-confirmatory analyses will be updated and documented in this sSAP and referenced in the Clinical Study Report (CSR) for the study. [redacted] Separate analysis plans (ie, separate documents from the sSAP) will be developed to detail PK and biomarker analyses. [redacted]

2. SUMMARY OF CHANGES

Section # and Name	Description of Change	Brief Rationale
3.1 Statistical Analysis Plan Summary (Table 1)	Updated stratification factors, primary and secondary endpoints, statistical methods for key efficacy analysis, interim analysis, multiplicity and sample size and power	To revise SAP according to Protocol Amendment 4
3.2 Responsibility for Analyses/In-House Blinding	Updated to remove information regarding interim analysis	To revise SAP according to Protocol Amendment 4
3.4.1 Efficacy Endpoints	Updated primary and secondary endpoints	To revise SAP according to Protocol Amendment 4
3.6.1 Statistical Methods for Key Efficacy Endpoints	Updated the statistical methods to align with primary and secondary endpoints	To revise SAP according to Protocol Amendment 4
3.6.1.6 Statistical Methods for Key Efficacy Endpoints (Table 4)	Updated Table 4 to align primary and secondary endpoints with Phase 2 trial design	To revise SAP according to Protocol Amendment 4

Section # and Name	Description of Change	Brief Rationale
3.7 Interim Analysis	Removed Section 3.7.1 Safety Interim Analysis and 3.7.2 Efficacy Interim Analysis	To revise SAP according to Protocol Amendment 4
3.8 Multiplicity	Removed entire section regarding interim analyses information	To revise SAP according to Protocol Amendment 4
3.9 Sample Size and Power Calculations	Updated sample size and power to align with change to Phase 2 trial design	To revise SAP according to Protocol Amendment 4
3.10 Subgroup Analyses	Updated stratification factors to remove ECOG and geographical region	To revise SAP according to Protocol Amendment 4

3. ANALYTICAL AND METHODOLOGICAL DETAILS

3.1 STATISTICAL ANALYSIS PLAN SUMMARY

Key elements of the Statistical Analysis Plan (SAP) are summarized in Table 1. The comprehensive plan is provided in Sections 3.2 through 3.12.

Table 1 Key Elements of the Statistical Analysis Plan

Study Design Overview	Phase 2 study of pembrolizumab + epacadostat vs pembrolizumab + placebo for first-line treatment of metastatic non-small cell lung cancer (NSCLC) in participants whose tumors express PD-L1 (TPS ≥50%)
Treatment Assignment	Approximately 148 participants will be randomized in a 1:1 ratio between two treatment arms: (1) pembrolizumab plus epacadostat and (2) pembrolizumab plus placebo. Stratification factor is predominant tumor histology(squamous vs non-squamous)
Analysis Populations	Efficacy: Intention to Treat (ITT) Safety: All Participants as Treated (APaT)
Primary Endpoints	• Objective response rate (ORR) per RECIST 1.1 based on BICR
Secondary Endpoints	• Progression-free survival (PFS) based on RECIST 1.1 as assessed by BICR. • Overall survival (OS). • Duration of response (DOR) based on RECIST 1.1 as assessed by BICR. • Safety and tolerability.
Statistical Methods for Key Efficacy Analyses	The primary hypothesis will be evaluated by comparing pembrolizumab plus epacadostat to pembrolizumab plus placebo in ORR using the stratified Miettinen and Nurminen method. The difference in PFS and OS will be evaluated using a stratified Log-rank test. The hazard ratio 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.
Multiplicity	For this Phase 2 trial, the overall Type I error rate is strictly controlled at 5% (one-sided) for the primary analysis of ORR. If the primary hypothesis is rejected at the α=5% level (one-sided), then testing will continue to the key secondary hypothesis of PFS. Nominal p-value for other endpoints will be reported, where applicable.
Sample Size and Power	The planned sample size is approximately 148 participants with 74 participants in each arm. For the ORR test, based on all patients randomized with minimum 18 weeks of follow-up, the study has 80.4% power to detect a 20 percentage point difference in ORR to 70% for pembrolizumab+ epacadostat vs 50% for pembrolizumab+ placebo at α=5% (one-sided).

3.2 RESPONSIBILITY FOR ANALYSES/IN-HOUSE BLINDING

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

MSD will generate the randomized allocation schedule for study treatment assignment for this protocol and the randomization will be implemented in IVRS/IWRS.

An external DMC will be convened to review accumulating safety data to provide an opportunity to terminate the study early if there are concerns regarding safety. Treatment-level results at the interim analyses will be provided by the external unblinded statistician to the eDMC. The DMC responsibilities and review schedules will be outlined in the DMC charter. The recommendation of the DMC will be communicated to the Joint Executive Oversight Committee (EOC) and, in the event of a recommendation to halt the study early due to safety concerns, to the appropriate regulatory agencies. If the DMC recommends modifications to the design of the protocol or discontinuation of the study, the EOC and possibly other limited numbers of 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

Objectives and hypotheses of the study are stated in protocol Section 4.0.

3.4 ANALYSIS ENDPOINTS

Efficacy and safety endpoints that will be evaluated for within- and/or between-treatment differences are listed below. [redacted]

3.4.1 Efficacy Endpoints

Primary

Objective Response Rate: The proportion of participants who have a confirmed CR or PR per RECIST 1.1 based on BICR.

Secondary

Progression-free Survival: The time from randomization to the first documented PD per RECIST 1.1 based on BICR or death due to any cause, whichever occurs first. See Section 10.6.1 for the definition of censoring.

Overall Survival: The time from randomization to death due to any cause.

Duration of Response: The time from first documented evidence of CR or PR until PD per RECIST 1.1 as assessed by BICR or death due to any cause, whichever occurs first, in participants who demonstrate CR or PR.

3.4.2 Safety Endpoints

Safety and tolerability will be assessed by clinical review of all relevant parameters including AEs, SAEs, fatal AEs, laboratory tests, and vital signs. Furthermore, specific events will be collected and designated as ECIs as described in protocol Section 9.3.7.

3.5 ANALYSIS POPULATIONS

3.5.1 Efficacy Analysis Populations

The analyses of primary efficacy endpoints are based on the intention-to-treat (ITT) population. All randomized participants will be included in this population. Participants will be analyzed in the treatment group to which they are randomized. Details on the approach to handling missing data are provided in Section 3.6.

3.5.2 Safety Analysis Populations

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 one dose of study treatment. Participants will be analyzed 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 medication for one 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.

[redacted]

3.6 STATISTICAL METHODS

3.6.1 Statistical Methods for Efficacy Analyses

This section describes the statistical methods that address the primary and secondary efficacy objectives. [redacted] Efficacy results that will be deemed to be statistically significant with Type I error strictly controlled at 5%. Nominal p-values will be computed for other efficacy analyses, but should be interpreted with caution due to potential issues of multiplicity.

3.6.1.1 Objective Response Rate

The stratified Miettinen and Nurminen method will be used for comparison of ORR between the treatment groups. The difference in ORR and its 95% CI from the stratified Miettinen and Nurminen method with strata weighting by sample size will be provided. The stratification factor (Section 7.3.1) based on tumor histology will be used in the analysis.

3.6.1.2 Progression-free Survival

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 (based on the stratification factors defined in Section 7.3.1). 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) 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. The stratification factor (Section 7.3.1) based on tumor histology will be applied to both the stratified log-rank test and the stratified Cox model.

Since disease progression is assessed periodically, 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. For the primary analysis, for participants who have PD, the true date of disease progression will be approximated by the date of the first assessment at which PD is objectively documented per RECIST 1.1 (based on BICR), regardless of discontinuation of study drug. Death is always considered as a confirmed PD event.

In order to evaluate the robustness of the PFS endpoint per RECIST 1.1 based on BICR, two sensitivity analyses with a different set of censoring rules will be performed. For the first sensitivity analysis, participants who miss more than one disease assessment (with or without a subsequent death or progression) are censored at the last disease assessment prior to missing visits. The second sensitivity analysis handles participants who discontinue treatment or initiate an anticancer treatment subsequent to discontinuation of study-specified treatments differently from the primary analysis. The censoring rules for primary and sensitivity analyses are summarized in Table 2. If a participant meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied.

Table 2 Censoring Rules for Primary and Sensitivity Analyses of Progression-free Survival

Situation	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2
No PD and no death; new anticancer treatment is not initiated	Censored at last disease assessment	Censored at last disease assessment	Censored at last disease assessment if still on study therapy; progressed at treatment discontinuation otherwise
No PD and no death; new anticancer treatment is initiated	Censored at last disease assessment before new anticancer treatment	Censored at last disease assessment before new anticancer treatment	Progressed at date of new anticancer treatment
No PD and no death; ≥2 consecutive missed disease assessments	Censored at last disease assessment	Censored at last disease assessment prior to ≥2 consecutive missed visits	Censored at last disease assessment
PD or death documented after ≤1 missed disease assessment	Progressed at date of documented PD or death	Progressed at date of documented PD or death	Progressed at date of documented PD or death
PD or death documented at any time after ≥2 missed disease assessments	Progressed at date of documented PD or death	Censored at last disease assessment prior to the ≥2 missed disease assessment	Progressed at date of documented PD or death
PD=progressive disease			

3.6.1.3 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 (based on the stratification factor defined in Section 7.3.1). 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 factor (Section 7.3.1) based on tumor histology 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 of last known contact. Analysis using the Restricted Mean Survival Time method may be conducted for OS to account for the possible non-proportional hazards effect.

3.6.1.4 Duration of Response

For participants who demonstrate CR or PR, DOR 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. Censoring rules for DOR are summarized in Table 3. DOR will be assessed using RECIST 1.1 by BICR.

For each DOR analysis, a corresponding summary of the reasons responding participants are censored will also be provided. Responses in participants who are alive, have not progressed, have not initiated new anti-cancer treatment, and have not been determined to be lost to follow-up are considered ongoing at the time of analysis. If a participant meets multiple criteria for censoring, the censoring criterion that occurs earliest will be applied.

Table 3 Censoring Rules for Duration of Response

Situation	Date of Progression or Censoring	Outcome
No progression or death, new anti-cancer treatment not initiated	Last adequate disease assessment	Censor (non-event)
No progression or death, new anti-cancer treatment initiated	Last adequate disease assessment before new anti-cancer therapy initiated	Censor (non-event)
Death or progression after ≥2 consecutive missed disease assessments	Last adequate disease assessment prior to the after ≥2 missed adequate disease assessments	Censor (non-event)
Death or progression after ≤1 missed adequate disease assessments	Progressive disease 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.		

3.6.2 Statistical Methods for Key Efficacy Endpoints

A summary of the primary analysis strategy for the key efficacy endpoints is provided in Table 4.

Table 4 Analysis Methods for Key Efficacy Endpoints

Endpoint/Variable	Statistical Method	Analysis Population	Missing Data Approach
Primary Analyses:			
ORR (RECIST 1.1) by BICR	Testing and Estimation: Stratified Miettinen and Nurminen method	ITT	Participants with missing data are considered non-responders
Secondary Analyses:			
PFS (RECIST 1.1) by BICR	Estimation: Stratified Cox model with Efron's tie handling method Stratified Log-rank test	ITT	Censored according to rules in Table 2
OS	Estimation: Stratified Cox model with Efron's tie handling method Stratified Log-rank test	ITT	Censored at last known alive date
DOR (RECIST 1.1) by BICR	Summary statistics using Kaplan-Meier method	All responders in ITT	Non-responders are excluded in analysis
Sensitivity analyses will be performed for PFS, ORR, and DOR based on investigator's assessment. BICR=blinded independent central review; DOR=duration of response; ITT=intent-to-treat; ORR=objective response rate; OS=overall survival; PFS=progression-free survival; RECIST 1.1=Response Evaluation Criteria in Solid Tumors.			

3.6.3 Statistical Methods for Safety Analyses

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 5). The tiers differ with respect to the analyses that will be performed. Safety parameters or AEs of special interest (AEOSI) 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% CIs provided for between-group comparisons. Other safety parameters will be considered Tier 2 or Tier 3. Tier 2 parameters will be assessed via point estimates with 95% CIs provided for between-group comparisons; only point estimates by treatment group are provided for Tier 3 safety parameters.

AEOSI 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. Similarly, the combination of pembrolizumab and epacadostat has not been associated with any new safety signals. Finally, there are no known AEs associated with participants with NSCLC for which determination of a p-value is expected to impact the safety assessment. Therefore, there are no Tier 1 events expected in this study.

Adverse experiences (specific terms as well as system organ class terms) that are not pre-specified as Tier-1 endpoints will be classified as belonging to "Tier 2" or "Tier 3", based on the number of events observed and clinical interest. For Tier 2 events, 95% CIs will be provided for between-treatment differences in the percentage of participants with events; these analyses will be performed using the Miettinen and Nurminen method.

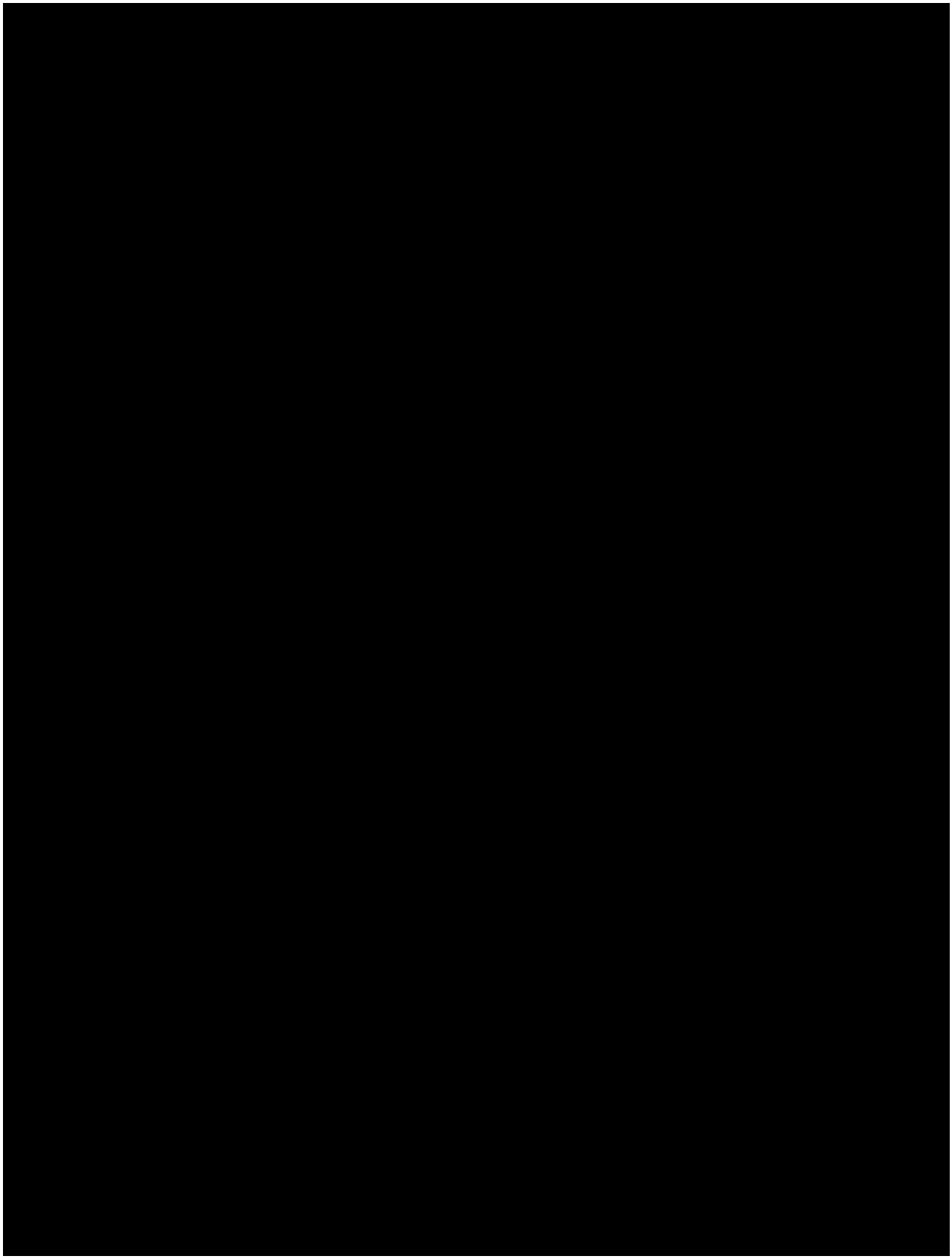
Because many 95% CIs may be provided without adjustment for multiplicity, the CIs should be regarded as a helpful descriptive measure to be used in review, not a formal method for assessing the statistical significance of the between-group differences in AEs and predefined limits of change.

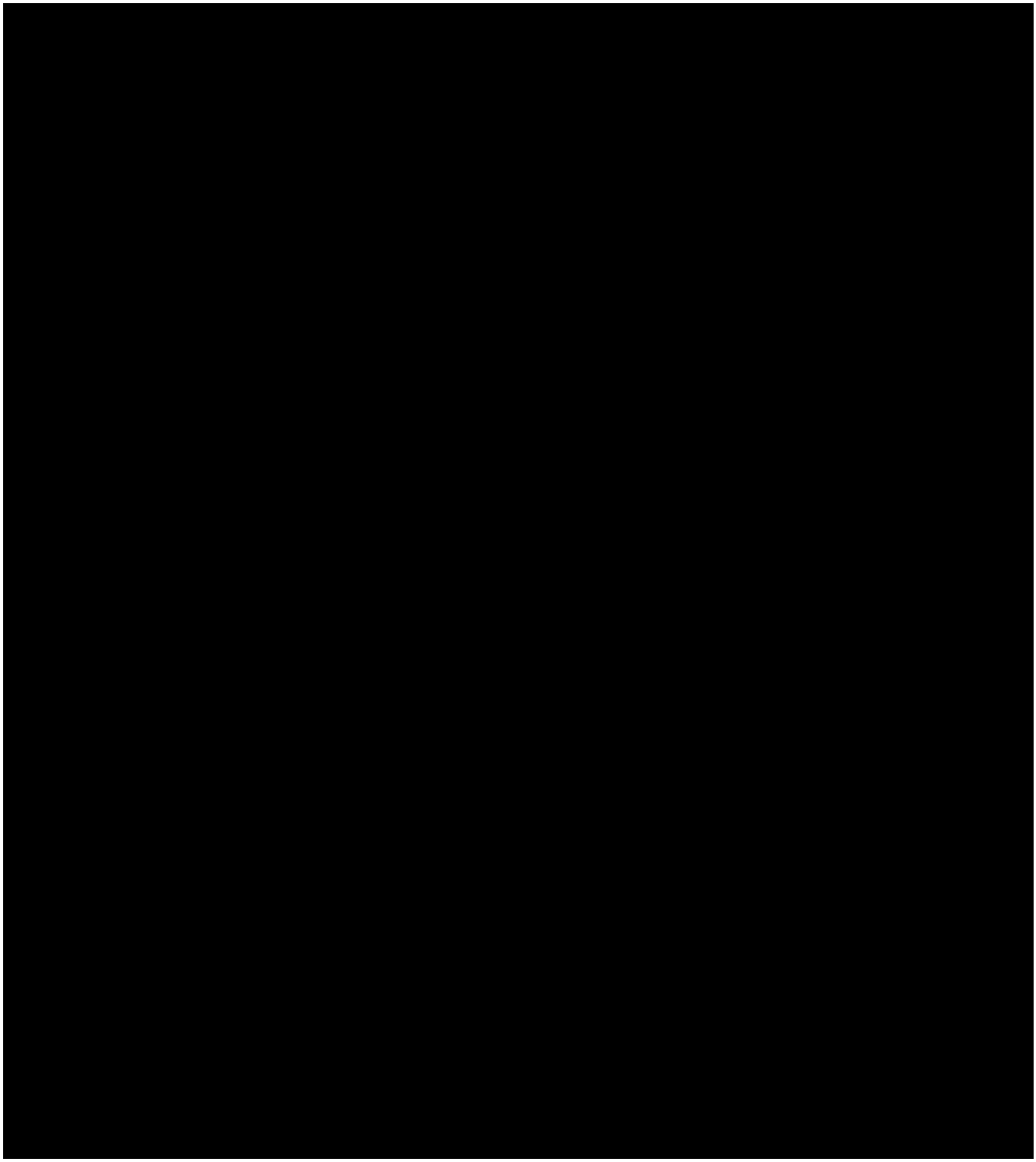
Continuous measures such as changes from baseline in laboratory, vital signs, and ECG parameters that are not pre-specified as Tier 1 endpoints will be considered Tier 3 safety parameters. Summary statistics for baseline, on-treatment, and change from baseline values will be provided by treatment group in table format.

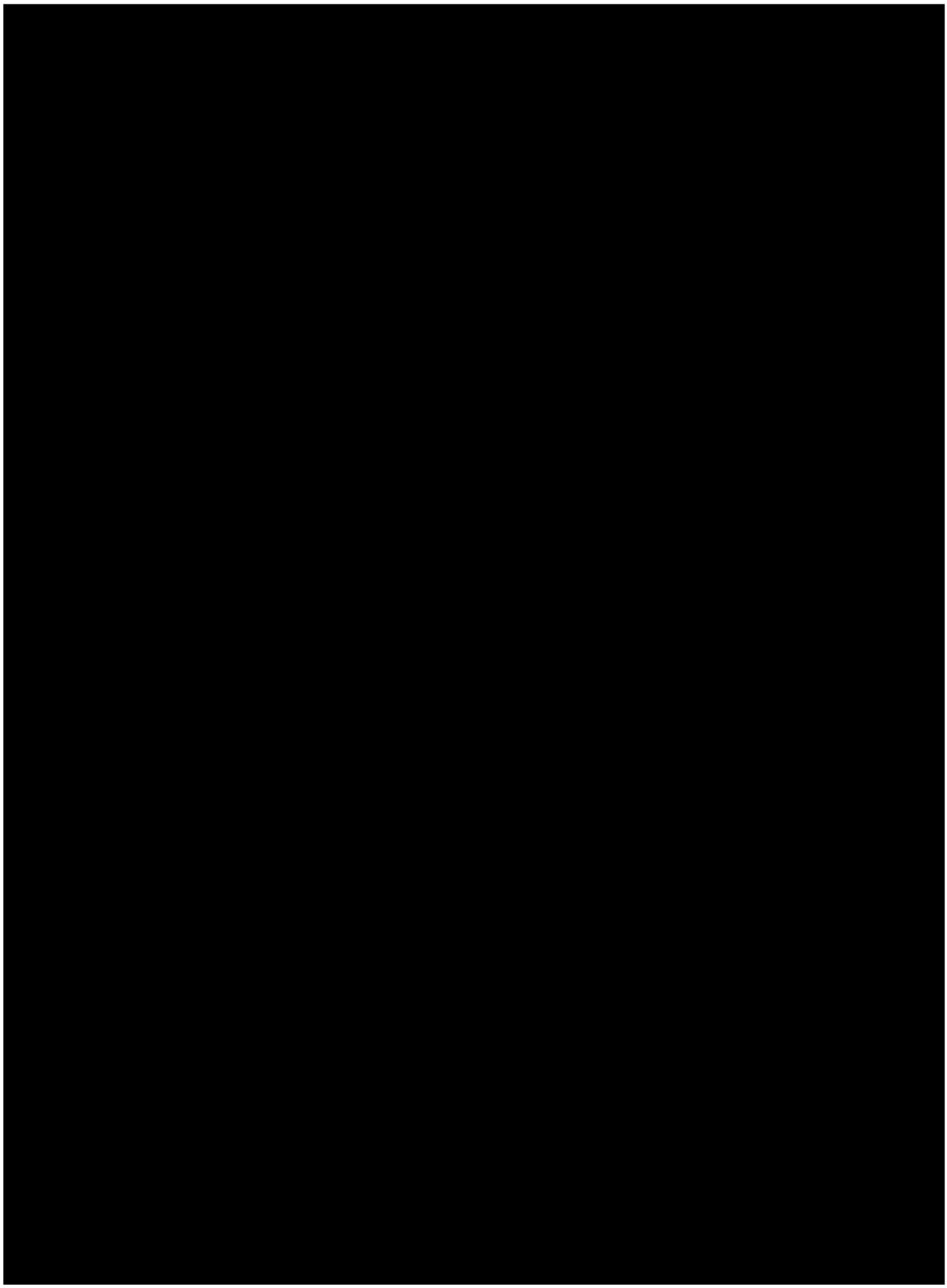
Table 5 Analysis Strategy for Safety Parameters

Safety Tier	Safety Endpoint	p-Value	95% CI for Treatment Comparison	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, SOCs (including ≥4 of participants in one of the treatment groups)		X	X
Tier 3	Specific AEs, SOCs (incidence <4 of participants in all of the treatment groups)			X
	Change from baseline results (labs, ECGs, vital signs)			X
AE=adverse event; CI=confidence interval; ECG=electrocardiogram; SOC=system organ class.				

[redacted]







3.6.5 Summaries of Baseline Characteristics and Demographics

The comparability of the treatment groups for each relevant characteristic will be assessed by the use of tables and/or graphs. No statistical hypothesis tests will be performed on these characteristics. The number and percentage of participants randomized and the primary reason for discontinuation will be displayed. Demographic variables (such as age) and baseline characteristics will be summarized by treatment either by descriptive statistics or categorical tables. The reasons for exclusion from the ITT population (if any) will be summarized.

3.7 INTERIM ANALYSES

3.7.1 Safety Interim Analyses

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

3.7.2 Efficacy Interim Analyses

No efficacy interim analysis is planned for this study. The study will be unblinded at the database lock for the primary analysis for ORR.

3.8 MULTIPLICITY

The overall Type I error rate is strictly controlled at 5% (one-sided) by fixed sequence, a closed-testing procedure. The closed testing procedure will be applied to the primary hypothesis of ORR first. If the primary hypothesis is rejected at the α=5% level (one-sided), then testing will continue to the key secondary hypothesis of PFS. Nominal p-value for each endpoint will be reported, where applicable, regardless of the outcome of the closed testing procedure dictated by the multiplicity strategy.

3.9 SAMPLE SIZE AND POWER CALCULATIONS

3.9.1 Sample Size and Power for Efficacy Analyses

The study will randomize 148 participants in a 1:1 ratio into the pembrolizumab plus epacadostat and pembrolizumab plus placebo arms. ORR is a primary endpoint for the study and PFS and OS are secondary endpoints.

Figure 1 summarizes power calculations for the primary hypothesis under various ORR difference assumptions assuming an underlying ORR of 50% in the pembrolizumab + placebo treatment group.

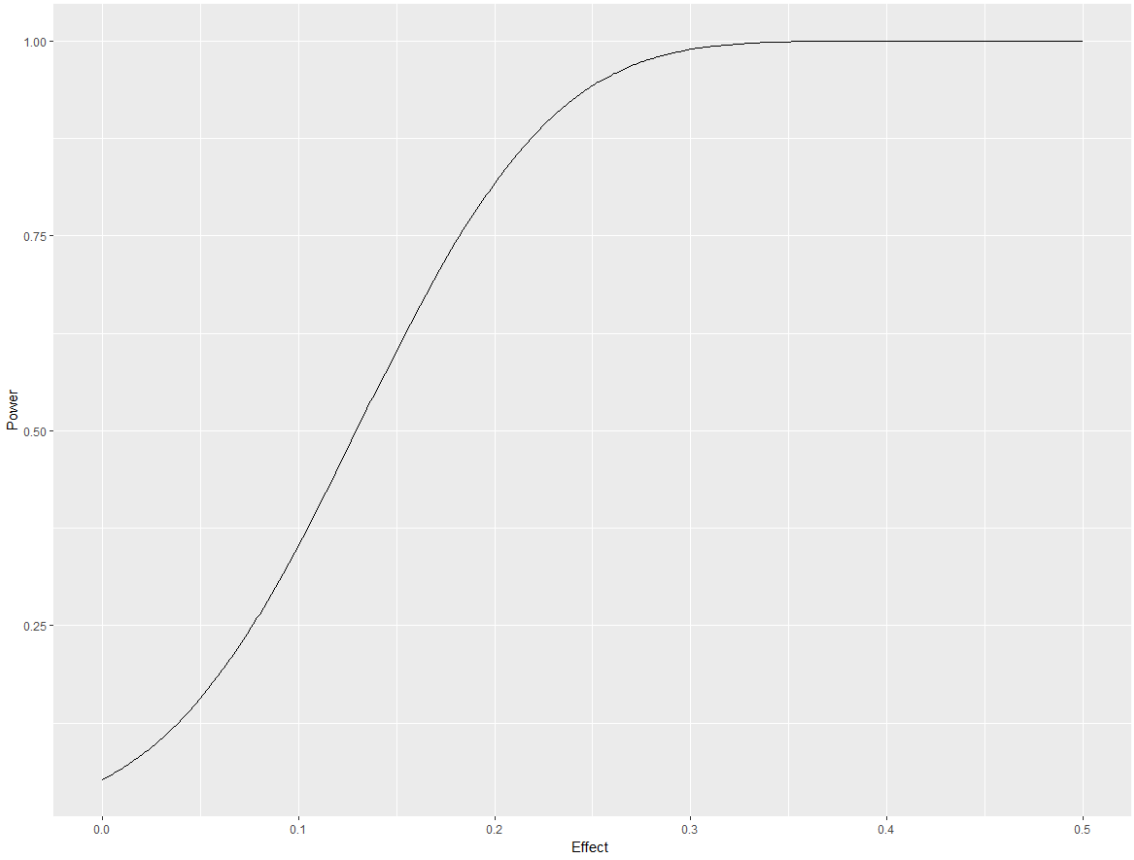


Figure 1 Power for Primary Hypothesis under Different Effect Size Assumptions

Based on the 148 participants in the control arm and the treatment arm under comparison, the power of the ORR testing at the α=0.05 (one-sided) is approximately 80.4% to detect a difference of 20 percentage points in ORR between an underlying 50% response rate in the control arm and a 70% response rate in the experimental arm.

With 95 PFS events, the study will have 80% power to detect a hazard ratio of 0.6 at an alpha level of 5% (1-sided).

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

3.10 SUBGROUP ANALYSES AND EFFECT OF BASELINE FACTORS

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

- Stratification factor
    - Predominant tumor histology (squamous vs non-squamous)
- Geographic region (East Asia vs non-East Asia)
- ECOG performance status (0 vs 1)
- Age category (<65, ≥65 years)
- Sex (female, male)
- Race (white, non-white)
- Smoking status (never vs former/current smoker)
- Brain metastasis (presence vs absence)

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

Extent of exposure for a participant is defined as number of cycles in which the participant receives the study medication. Summary statistics will be provided on extent of exposure for the APaT population.

4. REFERENCE

[Farrington, C. P. and Manning, G. 1990] Farrington CP, Manning G. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. Stat Med 1990;9:1447-54.

[Maurer, W. 2013] Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. Stat Biopharm Res 2013;5(4):311-20.

[Miettinen, O. and Nurminen, M. 1985] Miettinen O, Nurminen M. Comparative Analysis of Two Rates. Stat Med 1985;4:213-26.

5. APPENDIX