Official Protocol Title:	A Phase 3 Multicenter, Randomized, Double-blinded, Active-controlled, Clinical Study to Evaluate the Safety and Efficacy of Lenvatinib (E7080/MK-7902) with Pembrolizumab (MK-3475) in Combination with Transarterial Chemoembolization (TACE) Versus TACE in Participants with Incurable/Non-metastatic Hepatocellular Carcinoma (LEAP-012)
NCT Number:	NCT04246177
Document Date:	08-Jul-2025

Title Page

THIS SUPPLEMENTAL STATISTICAL ANALYSIS PLAN AND ALL OF THE INFORMATION RELATING TO IT ARE CONFIDENTIAL AND PROPRIETARY PROPERTY OF MERCK SHARP & DOHME LLC., A SUBSIDIARY OF MERCK & CO., INC., NJ, U.S.A (MSD).

Protocol Title: A Phase 3 Multicenter, Randomized, Double-blinded, Active-controlled, Clinical Study to Evaluate the Safety and Efficacy of Lenvatinib (E7080/MK-7902) with Pembrolizumab (MK-3475) in Combination with Transarterial Chemoembolization (TACE) Versus TACE in Participants with Incurable/Non-metastatic Hepatocellular Carcinoma (LEAP-012)

Protocol Number: 012-06

Compound Number: MK-7902

Sponsor Name:
Merck Sharp & Dohme LLC
(hereafter referred to as the Sponsor or MSD)

Legal Registered Address:
126 East Lincoln Ave.
P.O. Box 2000
Rahway, NJ 07065, USA

Confidential

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI



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 Sections 3.2-3.12.

Study Design Overview	A Phase 3 Multicenter, Randomized, Double-blinded, Active-controlled, Clinical Study to Evaluate the Safety and Efficacy of Lenvatinib (E7080/MK-7902) with Pembrolizumab (MK-3475) in Combination with Transarterial Chemoembolization (TACE) Versus TACE in Participants with Incurable/Non-metastatic Hepatocellular Carcinoma (LEAP-012)
Treatment Assignment	Approximately 450 participants will be randomized in a 1:1 ratio between 2 treatment arms: Arm A: Lenvatinib 12 mg (BW ≥60 kg) or 8 mg (BW <60 kg) orally QD plus pembrolizumab 400 mg IV Q6W plus TACE Arm B: Oral placebo QD plus IV placebo (normal saline) Q6W plus TACE. Treatment allocation/randomization will be stratified according to the following factors: • Study site • AFP ≤400 versus AFP >400 • ECOG PS 0 versus 1 • ALBI Grade (1 versus 2 or 3) • Tumor burden, utilizing the 6 and 12 rule, with 3 categories: ≤6 versus >6 but ≤12 versus >12 See Section 6.3.2 Allocation will be performed following Pocock and Simon covariate-adaptive randomization [Pocock, S. J. and Simon, R. 1975].
Analysis Populations	Efficacy: ITT Safety: APaT
Primary Endpoints/Hypotheses	1) PFS per RECIST 1.1 assessed by BICR modified to follow a maximum of 10 target lesions and a maximum of 5 target lesions per organ, with new intrahepatic lesions meeting LI-RADS 5 criteria. 2) OS
Statistical Methods for Key Efficacy Analyses	The primary hypotheses will be evaluated by comparing lenvatinib in combination with pembrolizumab plus TACE to placebo plus TACE on PFS and OS using the rerandomization tests based on the stratified log-rank test. Estimation of the HR will be performed 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	For analyses in which 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 [Miettinen, O. and Nurminen, M. 1985].

CCI

Multiplicity	The multiplicity strategy in this study will be applied to the 2 primary hypotheses (superiority of pembrolizumab in combination with lenvatinib plus TACE in PFS and in OS). The overall Type I error across the 2 hypotheses above is strongly controlled at 2.5% (one-sided). The multiplicity strategy will follow the graphical approach of Maurer and Bretz [Maurer, W. and Bretz, F. 2013] as described in Section 3.8, with initially 2.5% allocated to PFS hypotheses and 0% allocated to OS hypotheses. CCI

CCI

[redacted]

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.

This study will be conducted as a double-blind study under in-house blinding procedures. 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.

The randomization according to the Pocock and Simon covariate-adaptive procedure will be implemented in an interactive voice response system.

Blinding issues related to the planned IAs are described in Section 3.7.

3.3 Hypotheses/Estimation

Objectives and hypotheses of the study are stated in Section 3.0 of the protocol.

3.4 Analysis Endpoints

Efficacy and safety endpoints that will be evaluated are listed below.

3.4.1 Efficacy Endpoints

Primary

Secondary

- **ORR –** according to RECIST 1.1 assessed by BICR: a confirmed CR or PR.

- **DCR -** according to RECIST 1.1 assessed by BICR: CR, PR, or SD ≥6 weeks after randomization.

- **DOR -** according to RECIST 1.1 assessed by BICR: the time from first documented evidence of CR or PR until PD or death due to any cause, whichever occurs first, in participants who show confirmed CR or PR. See Section 3.6.1.6 for the definition of censoring.

- **TTP -** according to RECIST 1.1 assessed by BICR**:** the time from randomization to the first documented disease progression. See Section 3.6.1.5 for definition of censoring.

- **PFS, OR, DC, DOR and TTP -** according to mRECIST assessed by BICR.

Tertiary/Exploratory

CCI

CCI

[redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI

[redacted]

CCI

[redacted]

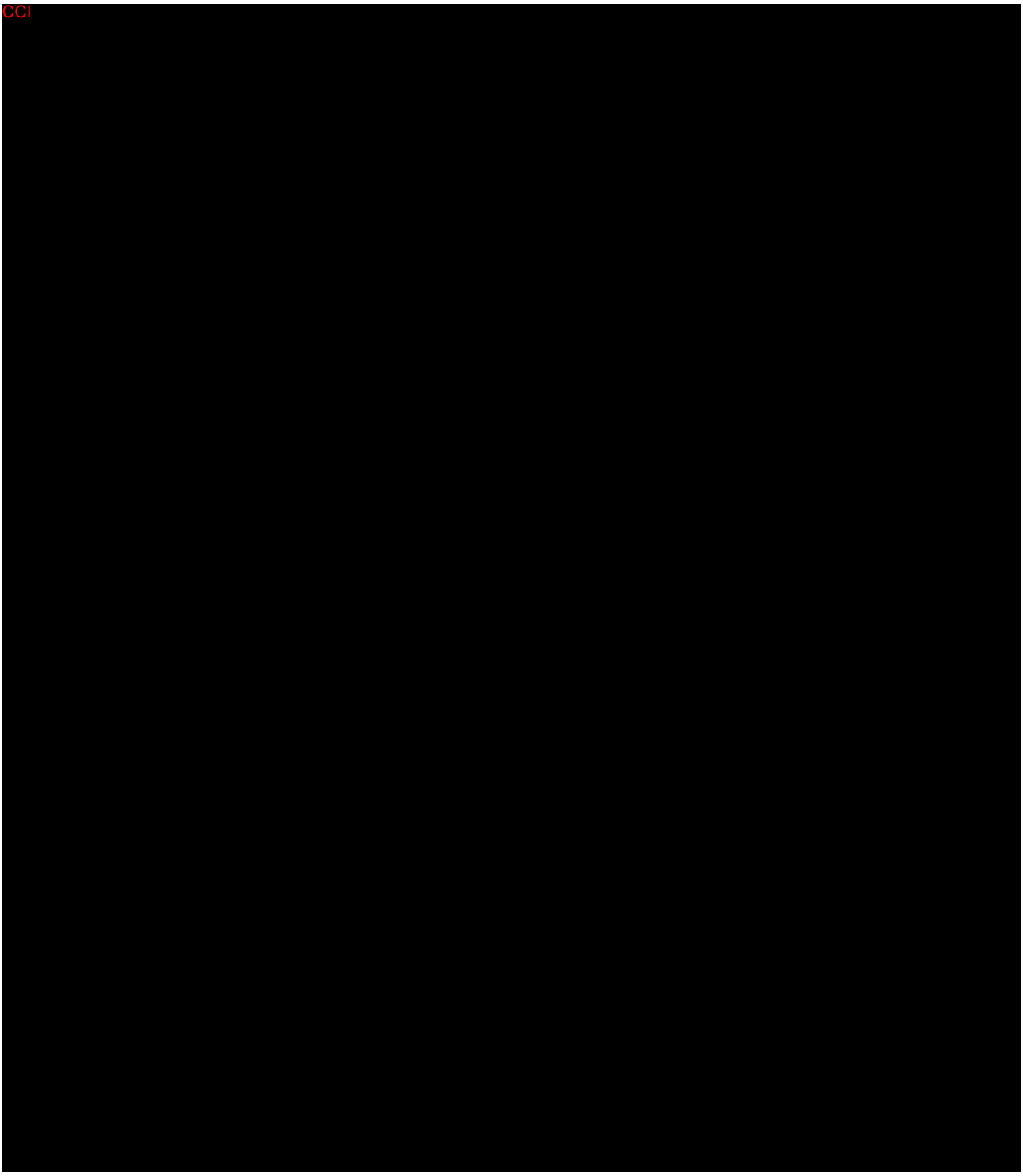
CCI

[redacted]

CCI



CCI



CCI

CCI

CCI

CCI

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

3.6.1.5 Time to Progression

The nonparametric Kaplan-Meier method will be used to estimate the TTP curve in each treatment group. The evaluation of the treatment groups in TTP will be performed using the stratified log-rank test. A stratified Cox proportional hazard model with Efron's method of tie handling will be used to estimate the magnitude of the treatment difference (ie, HR) between 2 treatment groups. The HR and its 95% CI from the stratified Cox model with a single treatment covariate will be reported.

CCI

CCI

CCI

CCI



Table 4 Analysis Strategy for Key Efficacy Endpoints

Endpoint/Variable (Description, Time Point)	Statistical Method[a]	Analysis Population	Missing Data Approach
Primary Hypothesis (H1)			
PFS per RECIST 1.1 by the BICR vendor modified to follow a maximum of 10 target lesions and a maximum of 5 target lesions per organ	Test: Rerandomization test based on the stratified log-rank test Estimation: Stratified Cox model with Efron's tie handling method	ITT	• Primary censoring rule • Sensitivity analysis 1 using log-rank test • Sensitivity analysis 2 using log-rank test (More details are in Table 1)
Primary Hypothesis (H2)			
OS	Test: Rerandomization test based on the stratified log-rank test Estimation: Stratified Cox model with Efron's tie handling method	ITT	Censored at last known alive date or data cutoff date, whichever is the earliest

CCI [redacted]

CCI [redacted]

CCI [redacted]

[redacted]

CCI [redacted]

95% CIs will be displayed for specific AEs, specific Grade 3-5 AEs and specific SAEs that meet the corresponding predefined threshold rules.

CIs for between treatment group differences will be provided using the Miettinen and Nurminen method [Miettinen, O. and Nurminen, M. 1985]. 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.

Exposure-adjusted AE

To properly account for the potential difference in follow-up time between the study arms, which is expected to be longer in the lenvatinib and pembrolizumab plus TACE arm, AE incidence adjusted for treatment exposure analyses may be performed as appropriate. Frequency of AE by time period from first dose (e.g., 0-3, 3-6, 6-12 months) may also be provided. In each time interval, the denominator is the number of participants at risk for the event during the particular time period, defined as participants who are event-free up until the start of the interval.

Time to Grade 3-5 AE

Additional exploratory analysis may be performed on the time to the first Grade 3-5 AE. The time to the first Grade 3-5 AE is defined as the time from the first day of study drug to the first event of a Grade 3-5 AE. Summary statistics will be provided.

The number and percentage of participants with laboratory toxicity grade increased from baseline will be summarized by the post-baseline maximum toxicity grade per CTCAE V5.0 for each gradable laboratory test.

For continuous safety measures, such as change from baseline in laboratory, vital signs, and ECG parameters, summary statistics for baseline, on-treatment, and change from baseline values will be provided.

3.6.2.2 Assessment of Safety Topics of Special Interest

AEs that are immune-mediated or potentially immune-mediated are considered safety topics of special interest (AEOSI) in this study. These events have been characterized consistently throughout the pembrolizumab clinical development program. Point estimates and 95% CIs for between-group difference is not expected to add value to the safety evaluation, and hence only number and percentage of participants with such pembrolizumab AEOSI will be provided, as well as the number and percentage of participants with corticosteroids administration to treat an AEOSI. Summary statistics will be provided for the analysis of time from first dose to the onset of an AEOSI.

Table 5 summarizes the analysis strategy for safety endpoints in this study.

Table 5 Analysis Strategy for Safety Parameters

Analysis Part	Safety Endpoint	Descriptive Statistics	95% between-group CI (Graphical display)
Overall Safety Assessment	Specific AEs CCI [redacted]	X	X
	Specific Grade 3-5 AE CCI [redacted]	X	X
	Specific serious A CCI [redacted]	X	X
	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 study treatment due to AE	X	
	Interruption of study treatment due to AE	X	
	AE that resulted in dose reduction	X	
	AE that resulted in death	X	
	Laboratory toxicity grade increase from baseline	X	
	Change from Baseline Results (vital signs)	X	
Assessment of safety topics of special interest	Pembrolizumab AEOSI	X	

AE = adverse event; AEOSI = adverse event of specific interest; CI = confidence interval; SOC = system organ class

CCI [redacted]

CCI



EORTC QLQ-HCC18 incorporates 6 symptom scales (fatigue, body image, jaundice, nutrition, pain, and fever), and two single items (abdominal swelling, and sex life).

EORTC QLQ-HCC18 scoring: The scoring approach for the EORTC QLQ-HCC18 is identical in principle to that for the symptom scales/single-items of the EORTC QLQ-C30. For symptom scales and single items, a high value indicates increased severity of symptoms or problems.

EQ-5D-5L scoring: The EQ-5D visual analogue scale (VAS) records the respondent's overall current health on a vertical visual analogue scale, ranging from 0 to 100, where the endpoints are labeled "The best health you can imagine (100)" and "The worst health you can imagine (0)". The EQ-5D VAS provides a quantitative measure of the patient's perception of their overall health.

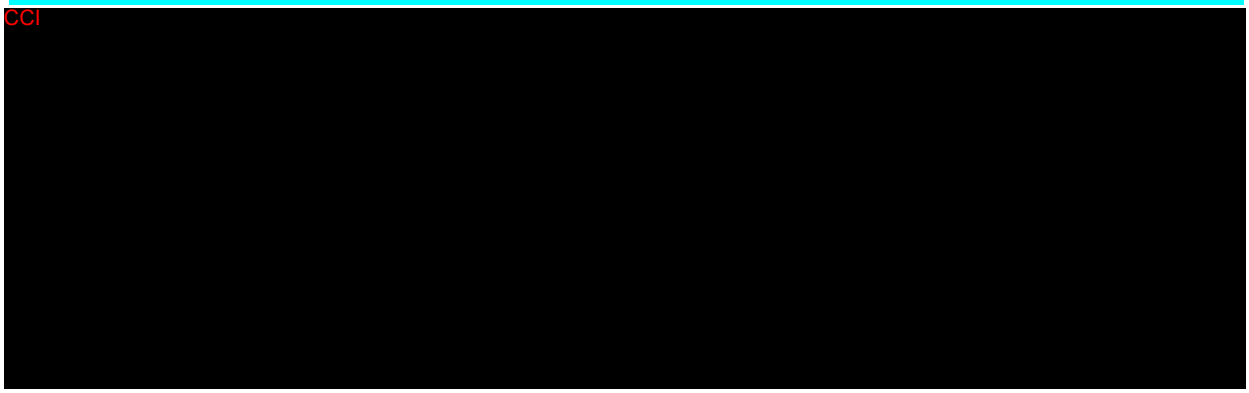
CCI

CCI




CCI


CCI


CCI

CCI

[redacted]

3.6.3.4 Time to Confirmed Deterioration (TTD)

The Kaplan-Meier method will be used to estimate the TTD curve for each treatment group. The estimate of median time to deterioration and its 95% confidence interval will be obtained from the Kaplan-Meier estimates. The treatment difference in TTD will be assessed by the stratified log-rank test, and two-sided nominal p-value will be reported. A stratified Cox proportional hazard model with Efron's method of tie handling and with a single treatment covariate will be used to report the HR and its 95% CI. The strata described for PFS in Section 3.6.1 will be used as the stratification factors in both the stratified log-rank test and the stratified Cox model.

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

3.6.3.6 Analysis Strategy for Key PRO Endpoints

CCI

[redacted]

CCI



CCI

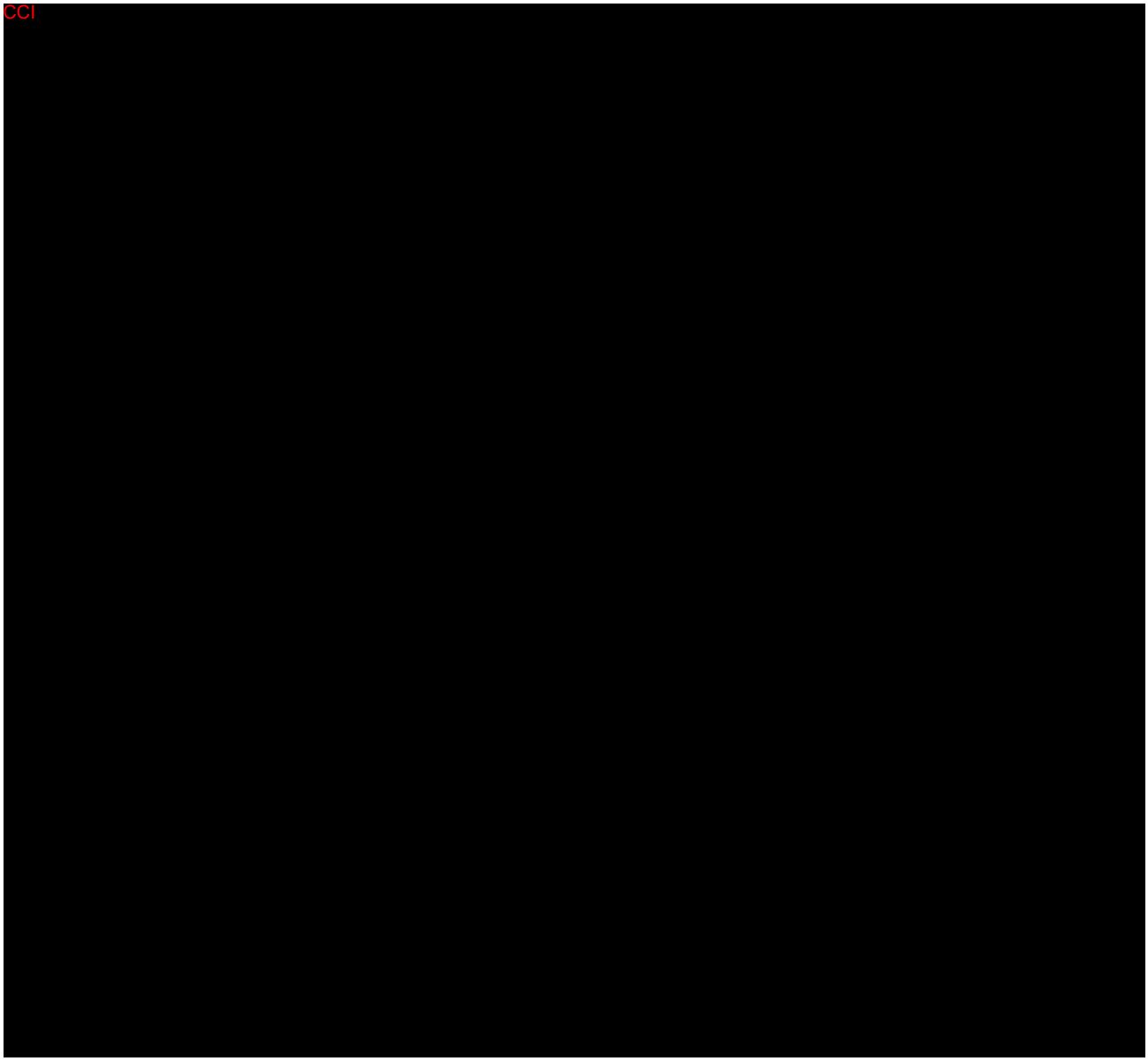


Figure 1 Multiplicity Graph for Type I Error Control



OS = overall survival; PFS = progression-free survival.

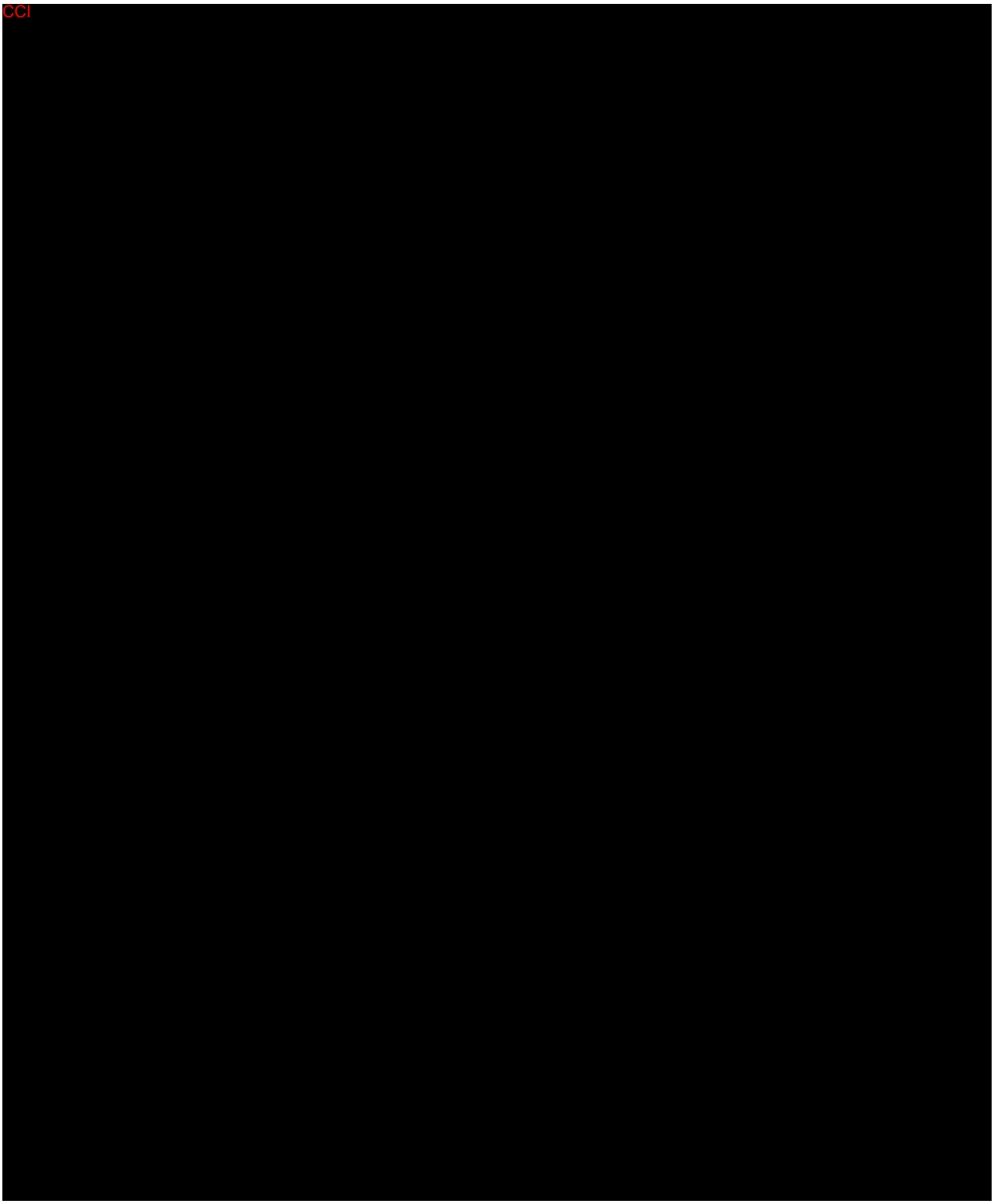
3.8.1 Progression-free Survival

The study will test PFS once at IA1 at α=0.025. Table 11shows the boundary properties for 0.025 1-sided α-level for the PFS analysis.

3.8.2 Overall Survival

The study will test OS at IA1, IA2, IA3, IA4, and final analysis. The OS hypothesis may be tested at α=0.025 if the PFS null hypothesis is rejected. Table 11 demonstrates the bounds and properties for OS hypothesis testing derived using a Lan-DeMets O'Brien-Fleming spending function with the overall α-level of 0.025. If the actual numbers of OS events at the interim and final analyses differ from those specified in the table, the bounds will be adjusted using the Lan-DeMets O'Brien-Fleming spending function accordingly.

CCI



CCI

[redacted]

CCI

[redacted]


CCI

4 REFERENCES

[Clopper, C. J. and Pearson, E. S. 1934]	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.	[03Y75Y]
[Cocks, K., et al 2012]	Cocks K, King MT, Velikova G, de Castro G, Martyn St-James M, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European organisation for the research and treatment of cancer quality of life questionnaire core 30. Eur J Cancer. 2012 Jul;48(11):1713-21.	[04SL8J]
[King, M. T. 1996]	King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Qual Life Res. 1996;5:555-67.	[052KJK]
[Liang, K.-Y. and Zeger, S. L. 2000]	Liang K-Y and Zeger SL. Longitudinal data analysis of continuous and discrete responses for pre-post designs. Sankhya: The Indian Journal of Statistics 2000;62(Series B, Pt. 1):134-48.	[00QJ0M]
[Marrero, J. A., et al 2018]	Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. Hepatology. 2018 Aug;68(2):723-50.	[056Z3B]
[Maurer, W. and Bretz, F. 2013]	Maurer W and Bretz F. Multiple testing in group sequential trials using graphical approaches. Stat Biopharm Res 2013;5(4):311-20.	[03XQVB]
[Miettinen, O. and Nurminen, M. 1985]	Miettinen O and Nurminen M. Comparative Analysis of Two Rates. Stat Med 1985;4:213-26.	[03QCDT]

[Osoba, D., et al 1998]	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.	[00TWVP]
[Plamadeala, V. and Rosenberger, W. F. 2012]	Plamadeala V and Rosenberger WF. Sequential monitoring with conditional randomization tests. The Annals of Statistics. 2012;40(1):30–44.	[08HR4V]
[Pocock, S. J. and Simon, R. 1975]	Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975;31:103-15.	[03R3BF]
[Scott, N. W., et al 2008]	Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. EORTC QLQ-C30 reference values [Internet]. Belgium: EORTC Quality of Life Group; 2008. Available from: http://groups.eortc.be/qol/sites/default/files/img/newsletter/reference_values_manual2008.pdf.	[04HN7P]

5 SUPPORTING DOCUMENTATION

5.1 Appendix 1: Technical Details on the cLDA model

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).

5.2 Appendix 2: Approval Information

The sSAP Amendment 02 of Protocol MK-7902-012-06 was approved by the BARDS TA head (or designee).

Name: Seth Thompson Date: 23-FEB-2026