Statistical Analysis Plan

CELLTRION, Inc.

A Double-Blind, Randomized, Active-Controlled, Parallel-Group, Phase 3 Study to Compare Efficacy and Safety of CT-P16 and EU-Approved Avastin as First-Line Treatment for Metastatic or Recurrent Non-Squamous Non-Small Cell Lung Cancer

Protocol Number CT-P16 3.1

SAP Version: <_____>

SAP Date: <_____>

Section 2-3: Objectives, Endpoints, Estimands, and Analysis Populations

2 Objectives, Endpoints, and Estimands

2.1 Primary Objective

- To demonstrate CT-P16 is similar to EU-Approved Avastin in terms of efficacy as determined by objective response rate (ORR) up to Cycle 6 during the Induction Study Period

2.1.1 Primary Endpoint

- Objective response rate (ORR, %) based on BOR during the Induction Study Period by RECIST v.1.1

2.2 Secondary Objectives

- To evaluate additional efficacy profiles including ORR during the Whole Study Period, response duration, time to progression (TTP), progression-free survival (PFS), and overall survival (OS)
- To evaluate the pharmacokinetics (PK) of trough serum concentration (Ctrough)
- To evaluate safety profile including immunogenicity
- To evaluate quality of life (QoL)

2.2.1 Secondary Endpoints

Efficacy Endpoints

- ORR based on BOR during the Whole Study Period by RECIST v.1.1
- Response duration: the time between initial response (CR or PR) and PD/recurrence
- TTP: the time from randomization until PD/recurrence

- PFS: the time from randomization until PD/recurrence or death due to any cause, whichever occurs first
- OS: the time from randomization until death due to any cause

Pharmacokinetic Endpoints

- Ctrough: Trough serum concentration

Quality of Life Analyses

- QLQ-C30 and QLQ-LC13, using EORTC QLQ

Safety Endpoints

- Incidence and severity of AEs, including SAEs graded according to the NCI CTCAE v.5.0.
- Incidence and severity of AESIs graded according to the NCI CTCAE v.5.0.
- Immunogenicity, as assessed by the incidence of antidrug antibody and neutralized antidrug antibody.
- Vital sign measurements
- ECG
- Physical examination findings
- ECOG
- Pregnancy testing
- Clinical laboratory analyses
- Previous and concomitant medications

2.3 Exploratory Objectives

[GAP: No exploratory objectives defined in protocol]

2.3.1 Exploratory Endpoints

[GAP: No exploratory endpoints defined in protocol]

3.3 Analysis Populations

Intent-to-treat (ITT) Population: The ITT population is defined as all patients randomly assigned to study drug, regardless of whether or not any study treatment dosing is completed. Patients will be assigned to treatment groups based on randomization.

Per-Protocol (PP) Population: The PP population is defined as all randomly assigned patients who have at least one response evaluation after receiving at least one full dose of study drug (CT-P16 or EU-Approved Avastin) in the Induction Study Period and who do not have any major protocol deviation. A major protocol deviation is one that may affect the interpretation of the primary endpoint and it will be defined in the statistical analysis plan (SAP). Final determinations of the PP population will be made at the blinded data review meeting before unblinding. Patients will be assigned to treatment groups based on randomization.

PK Population: The PK population is defined as all patients who receive at least one full dose of study drug (CT-P16 or EU-Approved Avastin) and who have at least one post treatment PK result.

If any patients are found to be non-compliant with respect to dosing, a decision will be made on a case-by-case basis at the blinded data review meeting before unblinding. Patients will be assigned to treatment groups based on treatment actually received.

Safety Population: The safety population is defined as all randomly assigned patients who receive at least one dose (full or partial) of study drug (CT-P16 or EU-Approved Avastin). Patients will be assigned to treatment groups based on treatment actually received.

11.1 Schedule of Assessments

Table 11-1 Schedule of Events

Study Period	**Screening**	**Induction Study Period (every 3 weeks)**						**Maintenance Study Period (every 3 weeks)**	**EOT Visit**	**Follow-Up Period (every 9 weeks)**
Cycle		C1	C2	C3	C4	C5	C6	C1, C2, C3, C4...		
Day	-28 to -1	D1	D1	D1	D1	D1	D1	D1	3 weeks after last dose	
Informed Consent	X									
Eligibility (Inc/Exc)	X	X								
Medical history	X									
Smoking history	X									
Diagnosis of	X									

nsNSCLC										
TNM staging (AJCC 8th)	X									
EGFR mutation / ALK rearrangement	X									
Viral assessment (HBV, HCV, HIV)	X									
Clinical laboratory test (8)	X	X	X	X	X	X	X	X	X	
Pregnancy test (9)	X							X (Cycle 1 only)	X	
12-lead ECG (10)	X	X	X	X	X	X	X	X	X	
Physical examination (11)	X	X	X	X	X	X	X	X	X	
ECOG performance status (11)	X	X	X	X	X	X	X	X	X	
Vital signs (11, 12)	X	X	X	X	X	X	X	X	X	

Height	X									
Weight	X	X	X	X	X	X	X	X	X	
Randomization		X								
CT-P16 or EU-Approved Avastin		X	X	X	X	X	X	X		
Paclitaxel and Carboplatin (13)		X	X	X	X	X	X			
Hypersensitivity monitoring		X	X	X	X	X	X	X		
PK sampling (14)		X	X	X	X	X	X	X (C1, C3, C6...)	X	
Tumor response evaluation (15, 16)	X		X		X		X	X (C3, C6, C9...)	X	X
Brain CT or MRI (17)	X									
Bone Scan (18)	X									
Immunogenicity (19)		X	X		X		X	X (C3, C6, C9...)	X	X (1st visit)
Quality of	X		X		X		X	X (C3,	X	

Life (20)								C6, C9...)		
Prior/ Concomitant medication (21)	X	X	X	X	X	X	X	X	X	
Adverse Events (22)	X	X	X	X	X	X	X	X	X	
Serious ADRs (23)										X
Survival follow-up										X

Footnotes:

11.1.1. Pathological diagnosis (nsNSCLC, predominant non-squamous) and EGFR mutation / ALK rearrangement will be confirmed by biopsy or cytology specimens before randomization.

11.1.2. Viral assessment (HBsAg, HBsAb, HBcAb, HCV antibody, and HIV antibody) will be performed at the Screening visit. If HBsAg is negative, but HBcAb is positive (regardless of HBsAb status), HBV DNA must be negative to allow participation in the study.

11.1.3. Clinical laboratory tests for Cycle 1 will be performed within 14 days prior to Day 1 of Cycle 1. Thereafter, clinical laboratory tests will be performed within 3 days prior to Day 1 of each cycle. Coagulation (prothrombin time and prothrombin time international normalized ratio) test will be performed at Screening, Day 1 of Cycle 1 and Cycle 2 of Induction Study Period and when clinically indicated for the patient who has been administered aspirin. Clinical laboratory test samples will be analyzed at the local laboratory.

11.1.4. Urine pregnancy test should be performed at the Screening visit (within 3 days before Day 1 of Cycle 1), within 3 days before Day 1 of Cycle 1 of the Maintenance Study Period and EOT visit, or at any time if pregnancy is suspected in females of childbearing potential only. If the urine pregnancy test gives equivocal results, a serum pregnancy test will be performed to exclude pregnant women in this study. For the Screening visit, only patients who have confirmed negative pregnancy test results will be included in this study. Pregnancy test samples will be analyzed at the local laboratory.

11.1.5. 12-lead ECG at the Screening visit will be performed within 14 days prior to Day 1 of Cycle 1 of the Induction Study Period. For all other cycles, 12-lead ECG will be performed within 3 days prior to Day 1 of each cycle.

11.1.6. These assessments should be performed prior to study treatment administration.

11.1.7. Vital signs (including blood pressure, heart and respiratory rates, and temperature) will be assessed prior to study drug administration and be measured after 5 minutes of rest (sitting). Height will be assessed at Screening only as a baseline measurement.

11.1.8. Paclitaxel and carboplatin will be administered on Day 1 of each cycle in Induction Study Period.

11.1.9. Pharmacokinetic samples will be collected on Day 1 of each cycle (prior to the beginning of the study drug administration [-3 days as window are allowed]) in the Induction Study Period, on Day 1 (-3 days as window are allowed) of Cycle 1 and every 3 cycles (end of Cycle 3, Cycle 6, and Cycle 9...) in the Maintenance Study Period, and EOT visit. In patients whose dose is delayed from the planned schedule, serum blood samples will be obtained on Day 22 of the last cycle (-3 days as window are allowed). Pharmacokinetic samples will be analyzed at the central laboratory.

11.1.10. Tumor evaluation will be assessed at Screening and every 2 cycles during the Induction Study Period and every 3 cycles during the Maintenance Study Period, and at the EOT visit. During the Follow-Up Period, it will be performed every 9 weeks until PD, death, withdrawal or start of new anti-cancer therapy if PD is not confirmed during the Induction or Maintenance Study Periods.

11.1.11. Tumor assessment will be performed within 4 weeks before the start of study treatment.

11.1.12. Brain CT or MRI will be assessed at Screening as mandatory to assess brain metastasis. During the study, imaging will be performed for patients with brain metastasis at baseline or for any suspected new lesions.

11.1.13. Bone scans will be assessed at Screening as mandatory. During the study, scans will be performed for patients with bone metastasis at baseline or for any suspected new lesions.

11.1.14. Immunogenicity will be assessed on Day 1 of Cycle 1 (predose), every 2 cycles during the Induction Study Period, every 3 cycles during the Maintenance Study Period and EOT visit. In the Follow-Up period, immunogenicity will be assessed once at the first visit of Follow-Up Period (9th week). A window of -3 days before Day 1 of the next cycle are allowed. In patients whose dose is delayed from the planned schedule, serum blood samples will be obtained on Day 22 of the last cycle (-3 days as window are allowed).

11.1.15. Quality of life should be assessed within 7 days before Day 1 of the next cycle, except screening assessment.

11.1.16. All medications used during the study, as well as all medications taken within 30 days of Day 1 of Cycle 1 in the Induction Study Period and until 28 days after the last dose of study treatment. Concomitant medications relevant to serious adverse drug reactions (ADR) occurred

after the EOT visit will also be collected. For patients who do not enter the Follow-Up Period, the last assessed concomitant medications will be collected.

11.1.17. Adverse events will be assessed from the date the informed consent form is signed until up to 28 days from last dose of study drug, regardless of the relationship to the study drug. Where an ADR is ongoing at the EOT visit, the ADR will be followed until one of the following: resolution or improvement from baseline, relationship reassessed as unrelated, start of new anti-cancer therapy, confirmation from the investigator that no further improvement can be expected, end of collection of clinical or safety data, or final database closure. For patients who do not enter the Follow-Up Period, the last assessed status of AEs will be collected.

11.1.18. Serious ADRs occurring during the Follow-Up Period will be reported and followed up.

3.4 Timing of Analysis

The study is designed to have multiple reporting points for the clinical study report (CSR), with a primary analysis occurring after the completion of the Induction Study Period.

3.4.1 Primary Analysis

The primary analysis of efficacy and safety is planned after all patients have completed Cycle 6 of the Induction Study Period or have terminated early from the study before reaching the end of Cycle 6. The primary objective is to demonstrate the similarity of CT-P16 to EU-Approved Avastin based on the objective response rate (ORR) up to Cycle 6.

At this time, the study will be unblinded to pre-defined personnel from the Sponsor and Contract Research Organization (CRO) for reporting purposes as described in Section 9.6 of the protocol. The study will remain blinded to the investigators, patients, and other pre-defined blinded personnel until the end of the study.

3.4.2 DSMB Analyses

An independent Data Safety Monitoring Board (DSMB) will review and evaluate accumulating safety data throughout the study to ensure the safety of study patients. The DSMB and the statistical team providing the safety analyses for the DSMB will have access to the randomization code upon request during closed sessions.

The frequency and specific timing of these safety reviews will be described in a separate document titled ‘Independent DSMB Charter’.

3.4.3 Interim Reporting

An additional analysis is planned to be performed for reporting purposes once one year of data has been collected following the enrollment of the last patient. This analysis will include updated secondary efficacy endpoints (e.g., PFS, OS) and safety data.

3.4.4 Final Analysis

The final analysis will be performed after all patients have completed the study (including the Follow-Up Period) or have discontinued from the study, and the database has been finalized for

study completion. This is expected to occur approximately 3 years after the enrollment of the last patient.

The final analysis will include the final assessment of all secondary efficacy, pharmacokinetic, quality of life, and safety endpoints. The study will be fully unblinded to all personnel at this stage.

3.7 Decision Criteria/Statistical Hypotheses

The primary objective of this study is to demonstrate that CT-P16 is similar to EU-Approved Avastin in terms of efficacy, as determined by the objective response rate (ORR) up to Cycle 6 during the Induction Study Period.

3.7.1 Primary Efficacy Hypothesis

The primary analysis will test the hypothesis of therapeutic equivalence between CT-P16 and EU-Approved Avastin. Therapeutic similarity will be concluded if the 95% confidence interval (CI) for the difference in ORR (CT-P16 minus EU-Approved Avastin) is entirely contained within the predefined equivalence margin of (-12.5%, 12.5%).

The statistical hypotheses for the primary endpoint are:

- Null Hypothesis (H_0): The difference in ORR ($\pi_{CT-P16} - \pi_{EU-Avastin}$) is $\le -12.5\%$ or $\ge 12.5\%$.
- Alternative Hypothesis (H_1): $-12.5\% < \pi_{CT-P16} - \pi_{EU-Avastin} < 12.5\%$.

Where π represents the ORR in the respective treatment groups.

3.7.2 Significance Level and Confidence Intervals

All statements of statistical significance will be based on a two-sided test with an overall level of significance (α) of 0.05, unless otherwise specified. This corresponds to two one-sided tests (TOST) each at the 0.025 significance level.

- Primary Endpoint: Therapeutic similarity will be evaluated using a 2-sided 95% CI for the difference in proportions.
- Secondary Endpoints: For other efficacy analyses (ORR for the Whole Study Period, TTP, PFS, and OS), 2-sided 95% CIs will be presented for descriptive purposes and to evaluate the precision of the estimates.

3.7.3 Decision Criteria

The primary analysis will be performed using the Intent-to-Treat (ITT) and Per-Protocol (PP) populations (as defined in Section 3.3). For the primary objective to be met, equivalence must be demonstrated in both populations.

The decision criteria for rejecting the null hypothesis of non-equivalence are:

- The primary analysis will utilize a logistic regression model considering treatment group as a fixed effect and other covariates (to be defined in Section 4.1).

- The resulting odds ratio and its 95% CI will be converted into the difference of proportions using the Delta method.
- If the entire 95% CI for the difference in ORR falls within the interval (-12.5%, 12.5%), CT-P16 will be declared similar to EU-Approved Avastin.

3.7.4 Secondary Efficacy Endpoints

No formal hypothesis testing for superiority or equivalence is planned for secondary efficacy endpoints. These endpoints (ORR during the Whole Study Period, Response Duration, TTP, PFS, and OS) will be summarized by treatment group using point estimates (proportions or medians) and their corresponding 95% CIs. Time-to-event endpoints will be estimated using the Kaplan-Meier method (see Section 6 for further details).

3.7.5 Multiplicity

As there is only one primary endpoint and one primary comparison for the demonstration of similarity, no multiplicity adjustments are required for the primary analysis. Multiplicity adjustments for secondary endpoints or subgroup analyses are not planned (refer to Section 4.3 for further details).

3.7.6 Interim Analysis Decision Criteria

No interim analyses for the purpose of early termination for efficacy or futility are planned. Safety data will be monitored by an independent Data Safety Monitoring Board (DSMB) on an ongoing basis to ensure patient safety, as described in Section 3.4.2.

4.1 General Methodology

All analyses described in this plan will be performed using SAS v9.4 or higher unless specified otherwise.

The analyses documented here are considered a priori analyses in that they are defined prior to database lock. Changes to the planned analyses, if any, designed after database lock and/or unblinding will be considered post hoc analyses and will be applied as exploratory methodology. All post hoc analyses will be identified in the clinical study report.

The term “descriptive statistics” refers to the number of patients (n), mean, median, standard deviation (SD), minimum, and maximum for continuous variables; and refers to the number and/or percentage of patients (or events) for categorical variables.

Unless specified otherwise, summaries will be presented by treatment arm (CT-P16 and EU-Approved Avastin). Where specified in Section 6 or Section 7, certain efficacy and safety endpoints may also be summarized for the overall population.

For the primary efficacy analysis (see Section 3.7.3), a logistic regression model will be utilized. Continuous variables will be summarized by treatment and visit (where applicable) using descriptive statistics as defined above. Categorical variables will be summarized using frequency tables showing the number and percentage of patients within a particular category.

All data collected will be listed by patient, treatment, and visit where applicable.

Summaries of efficacy and safety parameters that are presented by-visit will use pre-defined visit windows as described in Section 4.5. Descriptive statistics will be presented as described in this section unless justified otherwise for specific endpoints.

Baseline for all assessments is defined in Section 4.2. Analysis populations referenced throughout this document (Intent-to-Treat, Per-Protocol, Safety, and PK populations) are defined in Section 3.3.

4.2 Key Definitions

The following definitions and rules will be used to establish the temporal boundaries for analysis and to derive key variables.

- First Dose Date: The date on which a patient receives the first administration of any component of the study treatment (CT-P16/EU-Approved Avastin, paclitaxel, or carboplatin). This is typically Cycle 1 Day 1 of the Induction Study Period.
- First Dose Date for the Induction Study Period: Defined as the First Dose Date (as described above).
- First Dose Date for the Maintenance Study Period: The date of the first administration of monotherapy (CT-P16 or EU-Approved Avastin) during the Maintenance Study Period.
- First Visit Date: The date on which the patient signed the informed consent form (ICF).
- Last Dose Date: The date of the last administration of any component of the study treatment (CT-P16/EU-Approved Avastin, paclitaxel, or carboplatin) received by the patient during the study.
- Last Dose Date for the Induction Study Period: The date of the last administration of any component of the study treatment during Cycles 1 through 6 of the Induction Study Period.
- Last Dose Date for the Maintenance Study Period: The date of the last administration of monotherapy (CT-P16 or EU-Approved Avastin) during the Maintenance Study Period.
- Last Visit Date: The date of the last study-related contact with the patient, which may be the date of the End of Treatment (EOT) visit or the final Follow-Up Period visit (e.g., date of death or last survival contact).
- Last Visit Date for the Induction Study Period: For patients who enter the Maintenance Study Period, this is the date of the last assessment performed prior to the start of monotherapy. For patients who do not enter the Maintenance Study Period, this is the date of the EOT visit.
- Last Visit Date for the Maintenance Study Period: The date of the EOT visit following the completion of monotherapy or premature discontinuation from the Maintenance Study Period.
- Baseline Value: Defined as the last non-missing observation (including central laboratory results, vital signs, and tumor assessments) recorded prior to the First Dose Date. If an assessment is performed on Study Day 1, it will be considered a baseline measurement only if it is explicitly recorded as being performed prior to the first dose of study treatment. If

multiple assessments are performed on the same pre-treatment day, the assessment closest to the time of dosing (but before dosing) will be used.

- Change from Baseline Value: Calculated as (Post-Baseline Value – Baseline Value). If either the Baseline Value or the Post-Baseline Value is missing, the change from baseline will be missing.
- Study Day 1: Defined as the First Dose Date (the date of the first administration of any study treatment component).
- Study Day: Defined as the number of days from Study Day 1.
- For assessments on or after Study Day 1: Study Day = (Date of Assessment – Study Day 1 Date) + 1.
- For assessments prior to Study Day 1: Study Day = (Date of Assessment – Study Day 1 Date).
- There is no Study Day 0.
- Duration of Treatment: Calculated as (Last Dose Date – First Dose Date) + 1.
- Duration of Treatment for the Induction Study Period: Calculated as (Last Dose Date for the Induction Study Period – First Dose Date for the Induction Study Period) + 1.
- Duration of Treatment for the Maintenance Study Period: Calculated as (Last Dose Date for the Maintenance Study Period – First Dose Date for the Maintenance Study Period) + 1.
- Duration of Study: Calculated as (Last Visit Date – Date of Randomization) + 1.
- Screening Period: Defined as the period from the First Visit Date (ICF signature) to the day before the First Dose Date.
- Induction Study Period: Defined as the period from Study Day 1 through the day before the first dose of monotherapy in the Maintenance Study Period, or until the EOT visit for patients who do not proceed to maintenance therapy.
- Maintenance Study Period: Defined as the period from the First Dose Date for the Maintenance Study Period through the date of the EOT visit.
- Follow-Up Period: Defined as the period beginning the day after the EOT visit until the date of death, withdrawal of consent, or the end of the study, whichever occurs first.
- Controlled Disease: Defined as a patient achieving a Best Overall Response (BOR) of Complete Response (CR), Partial Response (PR), or Stable Disease (SD) as assessed at the end of Cycle 6 (Induction Study Period) based on RECIST v.1.1. [Note: This determines eligibility for entry into the Maintenance Study Period].
- Age: Calculated as the integer part of (Date of Informed Consent – Date of Birth) / 365.25. [GAP: If only year of birth is collected, age will be calculated as (Year of Consent – Year of Birth)].

4.3 Multiplicity Adjustment

This study is designed to demonstrate therapeutic similarity between CT-P16 and EU-Approved Avastin based on a single primary efficacy endpoint. Consequently, the potential for Type I error inflation due to multiple comparisons is limited.

4.3.1 Primary Efficacy Analysis

As described in Section 3.7.1, the primary efficacy endpoint is the Objective Response Rate (ORR) during the Induction Study Period. Similarity will be concluded if the 2-sided 95% confidence interval (CI) for the difference in ORR (CT-P16 minus EU-Approved Avastin) is entirely contained within the equivalence margin of (-12.5%, 12.5%).

Since there is only one primary endpoint and one primary comparison between two treatment groups, no multiplicity adjustment is required for the primary analysis. The alpha level is fixed at 0.05 (two-sided) for this single confirmatory test.

4.3.2 Secondary Efficacy Endpoints

Secondary efficacy endpoints include ORR during the Whole Study Period, response duration, Time to Progression (TTP), Progression-Free Survival (PFS), and Overall Survival (OS). As established in Section 3.7.4, these endpoints are considered supportive and exploratory.

No formal hypothesis testing for superiority or equivalence will be performed for these secondary endpoints. All p-values and 95% CIs presented for secondary endpoints will be for descriptive purposes only, and no adjustments for multiplicity will be applied. Statistical significance for secondary endpoints will not be used to support confirmatory claims of similarity or superiority.

4.3.3 Interim Analyses

As specified in Section 3.4 and Section 7.6.7 of the protocol, there are no interim analyses planned for the purpose of early termination of the study for efficacy or futility.

The analysis performed after the completion of the Induction Study Period (as described in Section 3.4.1) is for the purpose of primary reporting and does not involve alpha-spending or stopping boundaries that would require multiplicity adjustment. The independent Data Safety Monitoring Board (DSMB) will perform ongoing safety reviews, but these reviews are not intended for formal efficacy testing and will not affect the Type I error rate of the primary analysis.

4.3.4 Subgroup Analyses

Any subgroup analyses (as described in Section 7.5 of the protocol) are considered exploratory. No adjustments for multiplicity will be performed for subgroup comparisons, and results will be interpreted as hypothesis-generating.

4.5 Visit Windows

Visit windowing will be applied for the analyses which use visit categories. The following conventions will be used to define visit windows for assessments performed during the Induction Study Period, Maintenance Study Period, and Follow-Up Period.

Study Day and Study Periods are defined in Section 4.2. Baseline is defined in Section 4.2 as the last non-missing observation prior to the First Dose Date (Study Day 1).

4.5.1 Induction Study Period Windows

The following windows apply to assessments scheduled every cycle (e.g., Vital Signs, Weight, Physical Examination, ECOG, and Clinical Laboratory Analyses) during the Induction Study Period.

Analysis Visit	**Target Study Day**	**Study Day Range**
Baseline	1 (Pre-dose)	low < 1
Cycle 2	22	2 – < 33
Cycle 3	43	33 – < 54
Cycle 4	64	54 – < 75
Cycle 5	85	75 – < 96
Cycle 6	106	96 – < End of Induction Study Period

The following windows apply to assessments scheduled every 2 cycles (e.g., Tumor Response Evaluation, Quality of Life, and Immunogenicity) during the Induction Study Period.

Analysis Visit	**Target Study Day**	**Study Day Range**
Baseline	1 (Pre-dose)	low < 1
End of Cycle 2	42	2 – < 64
End of Cycle 4	84	64 – < 106
End of Cycle 6	126	106 – < End of Induction Study Period

4.5.2 Maintenance Study Period Windows

Visit windows for the Maintenance Study Period are calculated relative to the Maintenance Study Day 1 (the date of the first dose of monotherapy).

The following windows apply to assessments scheduled every cycle during the Maintenance Study Period (e.g., Vital Signs, Weight, Labs).

Analysis Visit	**Target Maintenance Day**	**Maintenance Study Day Range**
Maintenance Cycle 2	22	2 – < 33
Maintenance Cycle 3	43	33 – < 54
Maintenance Cycle 4	64	54 – < 75
...	...	(Incremental steps of 21 days)

The following windows apply to assessments scheduled every 3 cycles during the Maintenance Study Period (e.g., Tumor Response Evaluation, PK Sampling, Immunogenicity, and Quality of Life).

Analysis Visit	**Target Maintenance Day**	**Maintenance Study Day Range**
End of Maintenance Cycle 3	63	2 – < 95

End of Maintenance Cycle 6	126	95 – < 158
End of Maintenance Cycle 9	189	158 – < 221
...	...	(Incremental steps of 63 days)

4.5.3 Follow-Up Period Windows

Visit windows for the Follow-Up Period are calculated relative to the number of days since the End of Treatment (EOT) Visit.

Analysis Visit	Target Day Since EOT	Day Range Since EOT
Follow-Up Week 9	63	1 – < 95
Follow-Up Week 18	126	95 – < 158
Follow-Up Week 27	189	158 – < 221
...	...	(Incremental steps of 63 days)

4.5.4 General Windowing Rules

1. Period Boundaries: Assessments will be mapped to windows only within the period in which they occurred. A window will not span across the boundary between the Induction Study Period and the Maintenance Study Period, or between the Maintenance Study Period and the Follow-Up Period.
2. Multiple Assessments: If more than one assessment falls within the same visit window, the assessment closest to the target day will be used for analysis. In the event of two assessments being equidistant from the target day, the later assessment will be used.
3. End of Treatment (EOT) Visit: The EOT visit is a nominal visit occurring 3 weeks after the last dose. Data collected at the EOT visit will be mapped to the appropriate cycle window based on the Study Day of the assessment, unless otherwise specified for specific safety summaries.
4. Unscheduled Visits: Data from unscheduled visits will be included in the windowing process and may be used as the analyzed value for a visit if it is the closest record to the target day.

Section 6: Efficacy Analyses

6.1 Analyses Supporting Primary Objective(s)

6.1.1 Primary Endpoint(s)/Estimand(s)

The primary objective is to demonstrate therapeutic similarity between CT-P16 and EU-Approved Avastin in patients with metastatic or recurrent nsNSCLC.

6.1.1.1 Definition of Endpoint(s)

The primary efficacy endpoint is the Objective Response Rate (ORR), defined as the proportion of patients achieving a Best Overall Response (BOR) of either Complete Response (CR) or

Partial Response (PR) during the Induction Study Period (up to Cycle 6), as determined by an Independent Tumor Review Committee (central review) using RECIST v.1.1.

Estimand Specification:

- Population: Intent-to-Treat (ITT) population and Per-Protocol (PP) population (refer to Section 3.3).
- Variable: Objective Response (Yes/No).
- Intercurrent Events (ICE) and Handling:
- Premature discontinuation from study drug/period due to any reason (e.g., AE, withdrawal, death) prior to the first post-baseline tumor assessment: These patients will be considered non-responders (Composite strategy).
- Use of prohibited anti-cancer therapy prior to progression: Tumor assessments performed after the start of new anti-cancer therapy will be excluded. If no prior response was achieved, the patient is a non-responder (Hypothetical strategy).
- Missing Data: Patients with no post-baseline tumor assessment or whose response cannot be evaluated (NE) will be treated as non-responders for the primary analysis.
- Population-level Summary: Difference in ORR (%) between CT-P16 and EU-Approved Avastin.

6.1.1.2 Primary Analysis

The primary analysis will compare the ORR between the CT-P16 and EU-Approved Avastin groups in both the ITT and PP populations. Equivalence will be assessed using a logistic regression model.

- Model Specification: The model will include treatment group as a fixed effect. Covariates will include the stratification factors: sex (female vs. male), disease status (recurrence vs. metastatic), and ECOG performance score (0 vs. 1). Country will be included as a covariate unless there are sparse data across centers/countries, in which case it may be pooled by region.
- Estimation: The odds ratio and its 95% confidence interval (CI) will be estimated from the model.
- Conversion: The resulting odds ratio and its 95% CI will be converted into a difference in proportions (CT-P16 – EU-Approved Avastin) using the Delta method.
- Similarity Criterion: Therapeutic similarity will be concluded if the 2-sided 95% CI for the difference in ORR is entirely contained within the equivalence margin of (-12.5%, 12.5%).

Descriptive statistics for the primary endpoint will include the number and percentage of responders (CR + PR) and non-responders (SD, PD, NE) per treatment group.

6.1.1.3 Secondary Analyses of Primary Endpoint

The primary analysis methodology described in Section 6.1.1.2 will be repeated using the BOR as determined by the investigator's local review to support the findings of the central review.

6.1.1.4 Sensitivity Analysis/Analyses

To assess the robustness of the primary analysis, the following sensitivity analyses will be performed on the ITT population:

- Complete Case Analysis: Patients with missing or NE responses will be excluded from the denominator to evaluate the impact of the non-responder imputation.
- Confirmed vs. Unconfirmed Response: While the primary endpoint requires confirmation (as per RECIST 1.1), an analysis including unconfirmed CR/PR will be performed.
- Alternative Covariate Adjustment: A model including only the treatment effect and no other covariates.

6.1.1.5 Subgroup and Supplementary Analysis/Analyses

The primary efficacy endpoint (ORR up to Cycle 6, central review) will be summarized by treatment group for the following subgroups in the ITT population:

- Age (< 65 vs. ≥ 65 years)
- Sex (Male vs. Female)
- Race and Ethnicity
- Smoking History (Never smoker vs. Current/Former smoker)
- Baseline Disease Status (Metastatic vs. Recurrent)
- Baseline ECOG Score (0 vs. 1)

For each subgroup, the difference in ORR and its 95% CI will be calculated using the same methodology as the primary analysis, provided there are sufficient patients in each category (e.g., $n > 10$ per treatment group).

6.2 Secondary Efficacy Analyses

6.2.1 ORR during the Whole Study Period

The ORR based on BOR during the Whole Study Period (Induction + Maintenance + Follow-up) will be analyzed for both the ITT and PP populations. The proportion of responders and the 2-sided 95% CI for each treatment group will be calculated using the Clopper-Pearson method. The difference in proportions and its 95% CI will be calculated using the methodology described in Section 6.1.1.2.

6.2.2 Time-to-Event Endpoints (PFS, TTP, OS, Response Duration)

Time-to-event (TTE) endpoints will be analyzed using the ITT and PP populations.

- Progression-Free Survival (PFS): Time from randomization until the first documentation of PD (per RECIST v.1.1) or death due to any cause.
- Time to Progression (TTP): Time from randomization until the first documentation of PD.
- Overall Survival (OS): Time from randomization until death due to any cause.
- Response Duration: Time from the first documented response (CR or PR) until the first documentation of PD or death (among responders only).

Statistical Methodology for TTE:

- Kaplan-Meier (KM) Method: Used to estimate the survival distributions for each treatment group.
- Summary Statistics: Median TTE and survival rates at specific landmarks (e.g., 6 months, 12 months) will be calculated with 95% CIs. The 95% CI for the median will be calculated using the Brookmeyer-Crowley method.
- Hazard Ratio (HR): A stratified Cox proportional hazards model, including the same stratification factors used in the randomization, will be used to estimate the HR (CT-P16 / EU-Approved Avastin) and its 95% CI.
- Log-Rank Test: A stratified log-rank test will be used to compare the survival distributions between groups for descriptive purposes (refer to Section 4.3.2).

Censoring Rules for PFS and TTP:

- Patients alive and without PD at the time of data cutoff/study completion will be censored at the date of the last adequate tumor assessment.
- Patients starting a new anti-cancer therapy prior to PD will be censored at the date of the last adequate tumor assessment prior to the start of the new therapy.
- If PD or death occurs after missing two or more consecutive scheduled tumor assessments, the patient will be censored at the date of the last adequate tumor assessment prior to the missing visits.

6.2.3 Quality of Life (QoL)

QoL data (EORTC QLQ-C30 and QLQ-LC13) will be analyzed using the ITT population.

- Actual scores and changes from baseline (refer to Section 4.2) will be summarized for each visit using descriptive statistics as described in Section 4.1.
- Summaries will be provided for the global health status/QoL scale, functional scales, and symptom scales.
- Missing items within a scale will be handled according to the EORTC QLQ-C30 Scoring Manual.

---

7 Safety Analyses

7.1 Safety Analysis

The Safety Population (as defined in Section 3.3) will be used for all safety analyses. Descriptive statistics, as described in Section 4.1, will be presented by treatment group (CT-P16 and EU-Approved Avastin) for the safety endpoints listed in Section 2.2.1. Baseline for all safety assessments is defined in Section 4.2.

7.1.3 Clinical Laboratory Panels

Clinical laboratory assessments include hematology, clinical chemistry, coagulation, and urinalysis. As specified in Section 6.5.2.9 and the Schedule of Events (Section 11.1), laboratory tests will be performed at Screening, on Day 1 of each cycle during the Induction Study Period,

on Day 1 of Cycle 1 and every 3 cycles during the Maintenance Study Period, and at the End of Treatment (EOT) visit.

All laboratory analyses will be performed by local laboratories. Laboratory data will be summarized and listed based on the standard units provided by the laboratories.

7.1.3.1 Laboratory Parameters

The following parameters will be summarized and/or listed:

- Hematology: Hematocrit, hemoglobin, white blood cell count, absolute neutrophil count (ANC), and platelets.
- Clinical Chemistry: Albumin, alkaline phosphatase, alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen (BUN), calcium, chloride, creatine phosphokinase, creatine kinase-myocardial band isoenzyme, total cholesterol, creatinine, creatinine clearance (CrCl), gamma-glutamyl transferase, glucose, lactate dehydrogenase, triglyceride, high-density lipoprotein cholesterol, phosphate, potassium, sodium, total bilirubin, direct bilirubin, total protein, and uric acid.
- Coagulation: Prothrombin time (PT) and prothrombin time international normalized ratio (INR). [Note: Per Section 6.5.2.9, these are only required for patients on low-dose aspirin or who recently stopped aspirin].
- Urinalysis: Bilirubin, blood, glucose, ketones, leukocytes, nitrite, pH, protein, specific gravity, and urobilinogen. Microscopic examination and 24-hour urine collection will be performed if indicated.
- Pregnancy Testing: Urine or serum pregnancy tests for females of childbearing potential.

7.1.3.2 Analysis Timepoints

Quantitative laboratory data will be summarized by visit according to the visit windows defined in Section 4.5. Summaries will include the actual values at each visit and the change from baseline for all post-baseline visits during the Induction and Maintenance Study Periods.

7.1.3.3 Toxicity Grading

Hematology and clinical chemistry parameters will be graded for severity according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 5.0. Parameters that cannot be graded by CTCAE will be summarized relative to the local laboratory normal reference ranges (Low, Normal, High).

7.1.3.4 Shift Tables

Shift tables will be used to summarize the change in status from baseline to the worst post-baseline value during the treatment period (defined as the period from first dose to 28 days after the last dose).

- For CTCAE-gradable parameters, shifts will be presented from the baseline grade to the maximum (worst) post-baseline grade.
- For parameters summarized by reference range, shifts will be presented from baseline category (Low, Normal, High) to the worst post-baseline category. Analyses for "High" and "Low" directions will be performed separately where applicable.

7.1.3.5 Treatment-Emergent Abnormalities

A treatment-emergent abnormality is defined as a post-baseline value that represents a worsening from baseline (either an increase in CTCAE grade or a shift from Normal/Low to High, or Normal/High to Low). The number and percentage of patients with treatment-emergent Grade 3 or 4 abnormalities will be summarized by treatment group.

Liver function test (LFT) abnormalities of potential clinical concern, including potential Hy's Law cases, will be identified based on the following criteria:

- ALT or AST > 3 × Upper Limit of Normal (ULN)
- Total Bilirubin > 2 × ULN
- Alkaline Phosphatase < 2 × ULN

[GAP: Specific LFT outlier tables should be discussed with the medical team].

7.1.3.6 Handling of Unscheduled and Duplicate Assessments

Unscheduled laboratory assessments will be included in the determination of the worst post-baseline value and in shift tables. However, they will be excluded from by-visit summary tables unless they fall within a visit window and are the closest assessment to the target day (see Section 4.5.4).

If multiple laboratory values are collected on the same Study Day, the value collected closest to the scheduled assessment time will be used for by-visit summaries. For worst-case analyses, the most extreme value (highest grade or most distant from the reference range) will be used.

11.5 Index of Tables, Figures, and Listings

This appendix provides a complete index of all planned Tables, Figures, and Listings (TFLs) for this Statistical Analysis Plan.

11.5.1 Index of Tables

Number	Title	Population
14.2.1	Change from Baseline in EORTC QLQ-C30 Scores by Visit	Intent-to-Treat
14.2.2	Frequency of Best Overall Response during Induction Study Period (Central Review)	Intent-to-Treat
14.2.3	Frequency of Best Overall Response during Induction Study Period (Central Review)	Per-Protocol
14.2.4	Equivalence Analysis of Objective Response Rate	Intent-to-Treat

	(Delta Method) - Central Review	
14.2.5	Equivalence Analysis of Objective Response Rate (Delta Method) - Central Review	Per-Protocol
14.2.6	Objective Response Rate during Whole Study Period	Intent-to-Treat
14.2.7	Summary of Progression-Free Survival (Kaplan-Meier Estimates and Stratified Cox Model)	Intent-to-Treat
14.2.8	Summary of Time to Progression (Kaplan-Meier Estimates)	Intent-to-Treat
14.2.9	Summary of Overall Survival (Kaplan-Meier Estimates)	Intent-to-Treat
14.2.10	Summary of Response Duration	Intent-to-Treat
14.2.11	Subgroup Analysis of Primary Efficacy Endpoint (ORR)	Intent-to-Treat
14.3.1	Summary of Hematology Parameters by Visit - Actual Values and Change from Baseline	Safety
14.3.2	Summary of Clinical Chemistry Parameters by Visit - Actual Values and Change from Baseline	Safety
14.3.3	Shift Table of CTCAE Grade from Baseline to Worst Post-Baseline Value - Hematology	Safety
14.3.4	Shift Table of CTCAE Grade from Baseline to Worst Post-Baseline Value - Clinical Chemistry	Safety
14.3.5	Shift Table of Reference Range Category from Baseline to Worst Post-Baseline Value - All Parameters	Safety
14.3.6	Incidence of Treatment-Emergent Grade 3 or 4 Laboratory Abnormalities	Safety
14.3.7	Summary of Patients with Potential Clinically Important Liver Function Abnormalities	Safety

11.5.2 Index of Figures

Number	Title	Population
14.2.F1	Kaplan-Meier Plot of Progression-Free Survival	Intent-to-Treat
14.2.F2	Kaplan-Meier Plot of Overall Survival	Intent-to-Treat
14.2.F3	Forest Plot of ORR Difference by Subgroup	Intent-to-Treat
14.2.F4	Mean (+/- SE) Change from Baseline in QLQ-C30 Global Health Status over Time	Intent-to-Treat
14.3.F1	Box Plots of ALT and AST over Time by Treatment Group	Safety

11.5.3 Index of Listings

Number	Title	Population
14.2.L1	Individual Patient Tumor Response Assessments (Central Review)	Intent-to-Treat
14.2.L2	Individual Patient Tumor Response Assessments (Local Review)	Intent-to-Treat
14.2.L3	Individual Patient Quality of Life Scores	Intent-to-Treat
14.2.L4	Dates of Progression and Survival Status	Intent-to-Treat
14.3.L1	Laboratory Results - Hematology	Safety
14.3.L2	Laboratory Results - Clinical Chemistry	Safety
14.3.L3	Laboratory Results - Coagulation and Urinalysis	Safety
14.3.L4	Laboratory Results - Pregnancy Testing	Safety
14.3.L5	Patients with Treatment-Emergent Grade 3 or 4 Laboratory Abnormalities	Safety