Statistical Analysis Plan

A Phase 2, Open-label, Multicenter, Randomized, Multidrug Platform Study of Durvalumab (MEDI4736) Alone or in Combination with Novel Agents in Subjects with Locally Advanced, Unresectable, Stage III Non-small Cell Lung Cancer (COAST)

Protocol Number: D9108C00001

TABLE OF CONTENTS

LIST OF TABLES

List of Abbreviations

Abbreviation or Specialized Term	Definition
AE	adverse event
ATC	Anatomical Therapeutic Chemical
BICR	Blinded Independent Central Review
BMI	body mass index
C_{Cr}	creatinine clearance rate
CRF	case report form
CCI	CCI
cCRT	concurrent chemoradiotherapy
CR	complete response
DoR	duration of response
DA	disease assessment
DC	disease control
DCO	data cutoff
DCR	disease control rate
eCRF	electronic Case Report Form
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
GFR	glomerular filtration rate
HR	hazard ratio
CCI	[redacted]
CCI	CCI
CCI	CCI
CCI	CCI
CCI	CCI
CCI	CCI
Ig	immunoglobulin
IHC	immunohistochemistry
IP	investigational product
IPD	important protocol deviation
ITT	Intent-to-treat
IV	intravenous
IxRS	interactive voice/web response system
LLOQ	lower limit of quantification
MedDRA	Medical Dictionary for Regulatory Activities
CCI	CCI
LRV	lower reference value
NSCLC	non-small cell lung cancer
OR	objective response

Abbreviation or Specialized Term	Definition
ORR	objective response rate
OS	overall survival
OS-12 (18, 24, 36)	overall survival at 12 (18, 24, 36) months
PD	progressive disease
PFS	progression-free survival
PFS-12 (18, 24, 36)	progression-free survival at 12 (18, 24, 36) months
PK	pharmacokinetics
PR	partial response
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SC	subcutaneous
SD	stable disease
SPP	statistical programming plan
TFL	Tables, figures, listings
CCI	CCI
TV	target value

1 INTRODUCTION

This document describes the statistical analysis methodology for protocol D9108C00001, a multidrug platform study of Durvalumab (MEDI4736) alone or in combination with novel agents in subjects with locally advanced, unresectable, stage III non-small cell lung cancer (NSCLC). As background information, an overview of the study design is provided. The main portion of this document details the statistical summaries relating to each study objective and describes the general conventions and definitions that will be used. In addition, a set of table templates and specifications will be included in a statistical programming plan (SPP) to complement this document.

2 STUDY OVERVIEW

2.1 Study Objectives

2.1.1 Primary Study Objective(s)

To compare the antitumor activity of durvalumab alone vs. durvalumab in combination with novel agents.

2.1.2 Secondary Study Objectives

Safety:

- To evaluate the safety and tolerability of durvalumab alone and durvalumab in combination with novel agents.

Clinical activity:

- To further compare the efficacy of durvalumab alone vs durvalumab in combination with novel agents.

Pharmacokinetics (PK):

- To describe the PK of durvalumab alone and durvalumab in combination with novel agents.

- To describe the PK of novel agents in combination with durvalumab.

Immunogenicity:

- To assess the immunogenicity of durvalumab alone or in combination with novel agents.

- To assess the immunogenicity of novel biologic agents in combination with durvalumab.

2.1.3 Exploratory Study Objectives

CCI

[redacted]

2.2 Study Design

COAST is a Phase 2, open-label, multicenter, randomized, multidrug platform study of durvalumab alone or in combination with novel agents in subjects with locally advanced, unresectable, Stage III NSCLC.

The study will evaluate the clinical activity and safety of durvalumab alone or in combination with novel agents in subjects with Stage III NSCLC who have not progressed following definitive concurrent chemoradiotherapy (cCRT) (Figure 1 in protocol). Subjects will be

randomized into the study and initiate study treatment within 42 days from the last session of cCRT. Up to approximately 80 sites globally will participate in this study.

Subjects may be treated for up to 12 months. Study treatment will be discontinued upon confirmed disease progression, unacceptable toxicity, or other reasons, e.g., subject decision or noncompliance with study procedures. Overall, all subjects continuing in the study must adhere to the study procedures for up to 5 years (from randomization) provided they achieve disease control (DC) and remain disease free.

After completing the 5-year post-randomization study period, all subjects included into any treatment arm may be contacted annually for survival until 5 years after the final subject is randomized into the study or until the sponsor stops the study, whichever occurs first. The platform nature of the study allows an experimental arm to be permanently closed while the overall COAST study remains open.

New durvalumab combination experimental arms may be added based on emerging non-clinical and clinical data via protocol amendment. Interim futility analyses will be performed using Bayesian predictive probability.

Details about treatment regimen can be found in Section 3.1.2 of the protocol.

2.3 Treatment Assignment and Blinding

Treatment Assignment

An interactive voice/web response system (IxRS) will be used to assign investigational product (IP) kit numbers to each subject who meets the eligibility criteria. A randomization method with dynamically changing randomization ratios will be employed to account for fluctuation in the number of treatment arms open for enrollment over the course of the study. The randomization will be stratified by histology (adenocarcinoma vs non-adenocarcinoma) and will use an equal ratio to all study treatment arms open for enrollment (e.g., if experimental arms are opened sequentially, an experimental arm is added/closed, or enrollment in an experimental arm is suspended). Additionally, the following considerations regarding randomization will apply:

- If there is no experimental arm open for enrollment, then enrollment will be paused. Any subject who had consented previously and was in screening when the experimental arm(s) was closed due to safety concerns may be allowed to enter into the study but must be allocated to the control arm.

- Once the control arm enrolls 60 subjects, a subsequent randomization scheme may be initiated to optimize the number of subjects allocated to control arm.

- At any point during the study, the control arm will have no less than a 33% chance to be randomized in comparison to the active experimental arm that has the lowest randomization ratio in the study.

Blinding

The study is not blinded.

2.4 Sample Size

Sixty subjects will be enrolled in each of the combination therapy experimental arms and more than 60 subjects will be enrolled in the durvalumab control arm dependent upon the length of the study. The sample size of 60 subjects per experimental arm is not designed to make explicit power and Type I error considerations for a hypothesis test. It is primarily chosen to obtain a preliminary assessment of antitumor activity with a certain degree of precision.

Table 2.4-1 shows estimated differences in objective response rate (ORR) between an experimental arm and control arm along with 2-sided 95% exact CIs with a sample size of 60 subjects each. Sixty subjects per treatment arm provides a 95% CI with reasonable width (approximately ± 18%) for the estimated ORR difference between a durvalumab combination experimental arm and the durvalumab control arm. For a 20% increase in ORR in any durvalumab combination experimental arm over durvalumab control arm (50% vs 30%, respectively), the lower limit of the 95% CI for the difference between these arms is 1%, which is greater than 0%, suggesting that the durvalumab combination experimental arm will have a higher ORR than the durvalumab control arm with 95% confidence.

Table 2.4-1 Estimated Differences in ORR Between Experimental and Control Arm of 60 Subjects Each

Number of Responders (ORR)		Difference (%) in ORR (2-sided 95% Exact Confidence Interval)
Experimental Arm (n = 60)	Control Arm (n = 60)	
24 (40%)	18 (30%)	10% (-9%, 28%)
27 (45%)	18 (30%)	15% (-4%, 33%)
30 (50%)	18 (30%)	20% (1%, 38%)
33 (55%)	18 (30%)	25% (6%, 42%)

ORR = objective response rate.

3 STATISTICAL METHODS

3.1 General Considerations

Tabular summaries will be presented by treatment group. Categorical data will be summarized by frequency distribution (number and percentage of subjects falling within each category). Continuous variables will be summarized by descriptive statistics including N, mean, standard deviation, median, and range (minimum and maximum). All available data will be used, and thus missing data will not be imputed, unless otherwise specified. In general, subjects with missing data for a parameter will be excluded from the summary of this parameter. Tables will be supported by data listings showing the original data forming the basis for the summaries. Data listings will be sorted by treatment group, subject number and date collected where applicable.

Unless stated otherwise, two-sided confidence intervals will be produced at 95%. Baseline values will be defined as the last valid assessment prior to the first administration of IP. Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured are considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

Data analyses will be conducted using the SAS® System (SAS Institute, Inc., Cary, NC, USA) Version 9.4 or above, unless otherwise specified. All analysis outputs will be validated according to MedImmune SAS programming standards and MedImmune validation procedures.

3.2 Analysis Populations

The analysis populations are defined in Table 3.2-1.

Table 3.2-1 Analysis Populations

Population	Description
Intent-to-treat (ITT) Population	The ITT population includes subjects who are randomized. Subjects will be analyzed according to their randomized treatment group.
As-treated Population	The As-treated population includes all subjects who receive any IP. Subjects will be analyzed according to the treatment they really receive.
Response-evaluable Population	The Response-evaluable population includes subjects from the As-treated population who have a baseline disease assessment (DA), have the opportunity to be followed for at least 16 weeks at the time of the DCO (i.e., dosed at least 16 weeks prior to the time of the DCO), and either has at least one post-baseline disease assessment and/or discontinued treatment due to death or disease progression. The response-evaluable population will not be used for the final analysis.

Table 3.2-1 Analysis Populations

Population	Description
PK-evaluable Population	The PK-evaluable population includes subjects from the As-treated population who have a non-missing baseline PK concentration and at least one non-missing post-baseline PK concentration.

IP = investigational product; PK = pharmacokinetic.

All baseline and efficacy parameters will be summarized based on the ITT population as primary analysis. The efficacy parameters may be summarized based on response-evaluable population as supportive analysis. Safety parameters will be summarized based on the As-treated population as primary analysis. The analysis of pharmacokinetics will be based on the PK-evaluable population.

3.3 Study Subjects

3.3.1 Subject Disposition and Completion Status

Subject disposition and completion status will be summarized for all subjects screened, and percentages will be based on subjects randomized and treated. In addition, disposition of subjects throughout the study with respect to discontinuation of treatment and end of study status including reasons will be provided. Summaries of the number and percentage of subjects entered at each site will be provided.

The mortality summary will include subjects with end of study status of death, as well as cause of death (toxicity related to IP or disease under investigation, or other).

Duration of follow-up will also be summarized.

A listing of global/country study disruptions (impacting visits, concomitant medications and/or exposure) will be provided.

3.3.2 Demographics and Baseline Characteristics

Demographics and baseline disease characteristics will be summarized for the ITT population. Demographic information related to sex, age, race, weight, height, and body mass index (BMI) will be presented by treatment group and for all subjects combined. A summary of smoking history will also be included.Tumor history including histology, stage, and pertinent biomarker results at the time of initial diagnosis and at study entry will be summarized. Baseline tumor characteristics including sum of target lesions diameters and ECOG performance status will also be summarized.

The summary for prior anticancer treatment will include the number and percent of subjects by treatment category (systemic therapy, radiation, cancer related surgery, other), number of prior systemic regimens and best response (complete response, partial response, stable disease, progressive disease, not evaluable, not applicable, per RECIST v1.1) to the most recent line of therapy, type of platinum agent (cisplatin, carboplatin, cisplatin and carboplatin, derived using the WHO Drug dictionary coded terms), time from last radiation to study drug initiation/randomization, and total amount of radiation received.

3.3.3 Study Drug Exposure

- Treatment exposure will be summarized for the As-treated population. Duration of exposure to IP(s) in weeks and cycles will be summarized by descriptive statistics and by frequency. The duration of exposure for each IP is defined as below. last dose date plus the dosing interval (days) minus first dose date for subjects who completed treatment.
- minimum of (date of death plus 1 day, last dose date plus the dosing interval) minus first dose date for subjects who discontinued treatment.

Dose intensity and relative dose intensity of IP(s) will be summarized by descriptive statistics. The relative dose intensity is a percent of total actual dose that a subject received during corresponding study treatment period versus the total intended dose for the same study treatment period according to the study protocol. The details of the dose intensity calculation will be provided in the SPP as part of the standard exposure TFL templates.

Dosing deviations for IP(s), per the CRF, will be summarized with reasons for deviations for the following categories: delays, omissions, and interruptions. Duration of dosing delays will be derived where a delay is flagged on the CRF, based on the scheduled dosing dates and the visit window. The number of subjects with dosing delays and total dose delays will be summarized.

The use of subsequent anticancer treatment after the discontinuation of study treatment will be summarized by the type of subsequent anticancer treatment if data allow.

3.3.4 Concomitant Medications

The number and percentage of subjects who took concomitant medications will be summarized by the ATC classification and the generic name coded by WHO Drug Dictionary Global B3 version March 2022 for the ITT Population. Concomitant medications will include all concomitant medications taken on or after the date of first dose of any IP or

any concomitant medication started prior to the first dose of study treatment that continued beyond the date of first dose of any IP.

3.3.5 Protocol Deviations

The incidence of important protocol deviations will be summarized by deviation categories. The COVID-19 related IPDs will also be summarized. A listing will be provided with protocol deviation details. None of the deviations will lead to subjects being excluded from any analysis populations described in Section 3.2, unless otherwise specified. If a deviation is serious enough to have a potential impact on the primary analysis, sensitivity analyses may be performed.

3.4 Efficacy Analyses

3.4.1 Primary Efficacy Endpoint(s) and Analyses

3.4.1.1 Primary Efficacy Endpoint(s) and Analysis

The primary efficacy endpoint is objective response (OR) per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1). ORR is defined as the proportion of subjects with a best overall response of confirmed CR or confirmed PR that occur prior to disease progression, and/or prior to the initiation of subsequent anticancer treatment. Confirmation of response requirements are defined in Section 3.4.2.1 for best overall response. The primary analysis of ORR will be based on the ITT population. An estimate of the difference in ORR as well as its 2-sided 95% exact CI between each experimental arm and control arm will be reported. The rates between each experimental arm and control arm will be tested for significance with a Cochran–Mantel–Haenszel test, stratified by histology (adenocarcinoma or non-adenocarcinoma). Subjects that have missing overall response assessments will be considered as non-responders, and will therefore be counted in the denominator, but not in the numerator of ORR. Primary efficacy analysis will include control subjects that are concurrently enrolled with each experimental arm.

3.4.1.2 Additional Analyses of the Primary Efficacy Endpoint(s)

The secondary analysis of ORR may be based on the Response-evaluable population, and/or Fisher's exact test. Sensitivity analyses may be provided based on a comparison between each experimental arm and all control subjects at the time of analysis.

3.4.2 Secondary Efficacy Endpoint(s) and Analyses

The secondary efficacy endpoints include disease control (DC); duration of response (DoR); progression-free survival (PFS); progression-free survival at 12 months (PFS-12) and overall survival (OS). The efficacy endpoints will be summarized based on the ITT population and may be based on the Response-evaluable population as a supportive analysis. For analyses that involve an arm comparison between an experimental arm and control arm, only control subjects that are concurrently enrolled with the experimental arm will be included.

Efficacy analyses, except for OS, will be based on an application of Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 (Eisenhauer et al, 2009) to investigator assessed tumor measurements.

Programmatic derivation guidance used for the application of RECIST v1.1 is provided in Appendix 7.1. RECIST v1.1 guidelines will be used to determine disease response.

3.4.2.1 Best Overall Response

Best overall response (BOR) will be based on all post-baseline disease assessments that occur prior to disease progression, and/or prior to the initiation of subsequent anticancer treatment. BOR will be summarized with the number and percentage of subjects for the following categories: complete response (CR); partial response (PR); stable disease (SD); progressive disease (PD); and non-evaluable (NE). At least 7 weeks (8 weeks minus 1 week to allow for the protocol-defined disease assessment window) from randomization of IP must elapse without a subsequently radiological disease progression to assign a best overall response of SD.

Confirmation of CR and PR is required and must occur at least 4 weeks after initial documentation of CR or PR. If CR is pending confirmation and is designated at an assessment followed by 1 or more NE assessments, and/or PR assessments such that 1) the Target Lesion Response is CR and the Non-Target Lesion Response is NE, or 2) the Target Lesion Response is PR due solely to an increase in one or more lymph nodes to a size ≥ 10 mm and the Non-Target Lesion Response is either CR or NE, CR may be confirmed thereafter. Similarly, if a PR is pending confirmation and is designated at an assessment followed by 1 or more NE and/or SD assessments, PR may be confirmed thereafter.

In general, subjects not classifiable under the RECIST v1.1 response categories due to insufficient data or early death will be classified as NE for BOR but will be counted in the denominator of all response rate calculations. This generalization includes if a subject has missing lesion data at baseline. In this scenario, the subject will be classified as NE for BOR.

If a subject is missing lesion data at a disease assessment and yet PD criteria is met despite the missing data, the subject will be classified as PD.

3.4.2.2 Duration of Response

Duration of response (DoR) is defined as the time from the first documentation of a subsequently confirmed OR until the first documentation of a disease progression or death due to any cause, whichever occurs first, regardless of whether the subject receives subsequent anticancer treatment prior to progression. Only subjects who have achieved OR (confirmed CR or confirmed PR) will be evaluated for DoR. DoR is defined in months as follows:

DoR (months) = (Date of PD/death or censoring – Date of first confirmed disease response + 1) / (365.25/12)

The date of PD/death or censoring is the same as defined for PFS in Section 3.4.2.4. The median DoR and its 95% CI will be estimated using the Kaplan-Meier method.

3.4.2.3 Disease Control

Disease control rate (DCR) is defined as the proportion of subjects with a BOR of confirmed CR, confirmed PR, or SD (maintained for ≥ 16 weeks). DCR will be estimated with a 95% CI using the exact probability method. An estimate of the difference in DCR between each experimental arm and control arm will be reported and tested for significance with the Cochran–Mantel–Haenszel test.

3.4.2.4 Progression-free Survival

Progression-free survival (PFS) is defined as the time from randomization until the first documentation of a disease progression or death due to any cause, whichever occurs first, regardless of whether the subject receives subsequent anticancer treatment prior to progression. Subjects who have no documented progression and are still alive at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST v1.1 assessment. PFS is defined in months as follows:

PFS (months) = (Date of PD/death or censoring – Date of randomization + 1) / (365.25/12),

The censoring guidance and the date of PD/death or censoring are given in the table below. The number and percentage of subjects experiencing a PFS event and Kaplan-Meier plots of PFS will be presented by treatment group. The median PFS and its 95% CI will be estimated using the Kaplan-Meier method. The proportion of subjects progression-free and alive at 12

months (PFS-12) and associated 95% CI will be estimated using the Kaplan-Meier method. Additional PFS landmarks to be included in the final analysis are PFS-18, PFS-24, and PFS-36.

The difference in PFS between each experimental arm and control arm will be tested for significance by using a stratified log-rank test on histology (adenocarcinoma vs non-adenocarcinoma). The log-rank test will be carried out with the Breslow method for handling ties. The hazard ratio (HR) (experimental vs. control) with two-sided 95% CI will be estimated by stratified Cox regression model on histology with ties handled by the Efron method (Efron et al, 1977).

The comparison of PFS-X (X=12, 18, 24, 36 months) between each experimental arm and control arm will be performed by normal approximation under complementary loglog (cloglog) transformation (Klein et al, 2007). The test statistic shown below asymptotically follows a chi-square distribution with one degree of freedom under the null hypothesis $H_0: S_1(t) = S_2(t)$.

$$X^2 = \frac{\left(\log\left(-\log\left(\hat{S}_1(t)\right)\right) - \log\left(-\log\left(\hat{S}_2(t)\right)\right)\right)^2}{\hat{\sigma}_1(t)^2/\left(\log\left(\hat{S}_1(t)\right)\right)^2 + \hat{\sigma}_2(t)^2/\left(\log\left(\hat{S}_2(t)\right)\right)^2}$$

where

- $\hat{S}_k(t)$, with $k = 1, 2$ is the Kaplan-Meier estimate of the survival function for the *k*th treatment group
- $\hat{\sigma}_k(t)^2 = \sum_{t_{ki} \le t} \frac{d_{ki}}{n_{ki}(n_{ki} - d_{ki})}$ (Greenwood's formula) is from the estimated variance $\hat{V}(\hat{S}_k(t)) = \hat{S}_k(t)^2 \hat{\sigma}_k(t)^2$ of the KM estimate $\hat{S}_k(t)$ with $k = 1, 2$
- d_{ki} refers to the number of events (i.e., deaths or disease progressions) at time t_{ki}
- n_{ki} refers to the number of subjects at risk at time t_{ki}

Table 3.4.1 Summary of Censoring Guidelines for PFS

Situation	Date of PD/Death or Censoring	PFS Outcome
Documented Progressive Disease (PD) or death	Date of earliest sign of PD or death, whichever comes first	Event (unless the censoring rule specified below)
Death or PD immediately after ≥ 2 consecutive missed or non-evaluable disease assessments as per the protocol specified assessment schedule	Date of last progression-free disease assessment prior to missed or non-evaluable assessments, or randomization, whichever occurred last	Censored
No PD or death at time of analysis or lost to follow-up	Date of last evaluable progression-free disease assessment	Censored
No tumor assessment at baseline and no evidence of PD at first post-baseline disease assessment OR No tumor assessment post-first dose, and no death prior to second scheduled post-baseline disease assessment	Date of randomization	Censored

PD = progressive disease; PFS = progression-free survival.

Subjects having missing lesion data at baseline or no disease assessments post-first dose of IP will have PFS censored at the date of randomization unless the subject dies prior to the second scheduled post-baseline disease assessment in which case the death date will qualify as a PFS event.

If a subject has two or more consecutive completely missed or non-evaluable assessments followed immediately by death or an assessment showing radiologic disease progression, then the subject will be censored for PFS. PFS will be censored at the date of randomization or the last progression-free disease assessment prior to the missed or non-evaluable assessments, whichever occurred last.

If a subject has two or more consecutively missed or non-evaluable assessments followed by an assessment showing no radiologic disease progression, then the assumption will be that the subject did not progress during the missed or non-evaluable assessments.

The length of gap (in number of days) to determine two or more consecutive assessments is defined as the following:

- 126 days during the first 12 months (A 8-week disease assessment interval plus a 7-day window for each assessment).

- 182 days from month 12 to month 24 (A 12-week disease assessment interval plus a 7-day window for each assessment).

- 422 days after month 24 (A 6-month disease assessment interval plus a 4-week window for each assessment).

Subjects remaining on study without radiologic disease progression or death at the time of analysis will be censored for PFS at the date of their last evaluable disease assessment.

3.4.2.5 Overall Survival

Overall survival (OS) is defined as the time from randomization until death due to any cause. A subject alive at the end of study or lost to follow-up will be censored for OS at the last date when the subject was known to be alive. The last date for each individual subject is defined as the latest among the following dates recorded on the case report forms (CRFs):

- AE start, stop, and change dates
- Admission and discharge dates of hospitalization
- Start date of protocol deviations
- Study treatment date
- Date of last contact, withdrawal consent, refuse to be followed up, or last known alive on end of treatment, end of study, and survival status/follow-up CRFs
- Laboratory test dates including (but not limited to) hematology, chemistry, urinalysis, coagulation, tumor biopsy, immunoglobin, pharmacokinetics, pharmacodynamic biomarkers.
- Disease assessment dates on RECIST CRF
- Date of visit, vital signs, ECOG, electrocardiogram, and physical examination
- Start and stop dates of alternative anticancer treatment
- Start and end date of concomitant medication and surgical/medical procedure

OS is defined in months as follows:

OS (months) = (Date of death or censoring – Date of randomization + 1) / (365.25/12).

The number and percentage of subjects experiencing an OS event and Kaplan-Meier plots of OS will be presented by treatment group. The median OS and its 95% CI will be estimated using the Kaplan-Meier method. The proportion of subjects alive at 12 months (OS-12) and

associated 95% CI will be estimated using the Kaplan-Meier method. Additional OS landmarks to be included in the final analysis are OS-18, OS-24, and OS-36.

The difference in OS between groups will be tested for significance by using a stratified log-rank test on histology (adenocarcinoma vs non-adenocarcinoma). The log-rank test will be carried out with the Breslow method for handling ties. The hazard ratios of each experimental arm vs. control arm with two-sided 95% CI will be estimated by a stratified Cox regression model on histology with ties handled by the Efron method (Efron et al, 1977).

The comparison of OS-X (X=12, 18, 24, 36 months) between the 2 treatment arms will be performed by normal approximation under cloglog transformation (Klein et al, 2007), where the event is death. The test statistic is defined in Section 3.4.2.4.

3.4.2.6 Other Endpoints

Time to Response

Time to response (TTR) is defined as the time from randomization until the first documentation of a subsequently confirmed OR. Only subjects who have achieved OR (confirmed CR or confirmed PR) will be evaluated for TTR. TTR is defined in months as follows:

TTR (months) = (Date of first confirmed disease response – Date of randomization + 1) / (365.25/12).

The median TTR and its 95% CI will be assessed using the Kaplan-Meier method. The number of subjects with response at different disease assessment timepoints may be provided.

Change from Baseline in Tumor Sizes

The percent change from baseline in target lesion sum of diameters (longest for non-nodal lesions, short axis for nodal lesions) will be calculated at each evaluable post-baseline disease assessment (DA) # X (X = 1, 2, 3, etc.). The percent change from baseline in target lesion sum of diameters is defined as follows:

$$\%Chg\ at\ DA\ \#X = 100 \times \frac{\sum\ Diameters\ at\ DA\ \#X - \sum\ Diameters\ at\ DA\ at\ baseline}{\sum\ Diameters\ at\ DA\ at\ baseline}$$

The percent change from baseline in target lesion sum of diameters will be presented by subject using spider plots. The best percent change from baseline in target lesion sum of diameters is defined as the largest reduction or smallest increase (in the case where a reduction does not occur) from baseline observed over all post-baseline disease assessments until disease progression. The best percent change from baseline will be presented using waterfall plots. Target lesion measurements and sum of diameters will be listed by disease assessment and subject.

3.4.3 Handling of Dropouts and Missing Data

In general, missing data are not imputed for statistical analysis. Guidance regarding the handling of dropouts, missing data, and censoring will apply uniformly to all efficacy analyses resulting from an application of RECIST v1.1 to investigator assessed tumor measurements. For investigator reported outcomes, analyses will present outcomes without consideration of missing data or censoring rules.

3.4.4 Subgroup Analyses

To assess the consistency of treatment effects, the analyses of ORR and PFS will be performed by the following subgroups:

- Age (< 70, ≥ 70 years)
- Sex (Female, Male)
- Race (Asian, non-Asian)
- Histology (Squamous, non-Squamous)
- Smoking status (Smoker, never smoker)
- ECOG performance status at baseline (0, 1)
- PD-L1 status (Positive, Low/Negative, Unknown) based on 1% and 25% cuts separately
- Platinum agent (Cisplatin, Carboplatin)
- Time from last radiation to randomization (<14 days, ≥14 days)
- Disease stage at study entry (IIIA, IIIB/C)

Forest plots will be provided for difference in ORR (confirmed/unconfirmed response and confirmed response, separately) and PFS hazard ratios.

3.4.5 Other Efficacy Analyses

Confirmed/unconfirmed ORR will be reported along with the primary analysis of confirmed ORR. Analyses will be conducted in the same way as described in Section 3.4.1.

Similarly, confirmed/unconfirmed DCR, DoR and TTR will be reported, with analyses conducted in the same way as described in Section 3.4.2.

The proportion of subjects progression-free and alive at 6 months (PFS-6) and 9 months (PFS-9) and the associated 95% CIs may be estimated using the Kaplan-Meier method.

CCI [redacted]

The number and percentage of subjects with distal metastasis at 6 months based on PET scan may be reported.

3.5 Pharmacodynamic Endpoint(s) and Analyses

Exploratory analysis of pharmacodynamic endpoints may be described separately.

3.6 Safety Analyses

All safety analyses will be performed based on the As-treated population, unless otherwise specified. Only post-baseline visits up to and including 90 days following the date of last dose of study medication will be considered in derivations.

3.6.1 Adverse Events and Serious Adverse Events

Adverse events will be coded by Medical Dictionary for Regulatory Activities (MedDRA) version 26.0 and the type, incidence, CTCAE grade and relationship to study IP will be summarized. Specific adverse events will be counted once for each subject for calculating percentages. In addition, if the same adverse event occurs multiple times within a particular subject, the maximum CTCAE grade and level of relationship observed will be reported. If any associations of interest between adverse events and baseline characteristics are observed, additional stratified results may be presented.

Treatment-emergent adverse events (TEAEs) are defined as events present at baseline that worsen in intensity after administration of study IP or events absent at baseline that emerge after administration of study IP. All TEAEs that occurred on and after first dose up to 90 days after last dose of IP (any) will be summarized overall, as well as categorized by MedDRA System Organ Class and Preferred Term. In addition, all TEAEs with an onset date no more than 90 days after the last dose of IP (any) will be listed.

The AEs occurring from the signing of the informed consent and prior to the initiation of study IP will be listed.

3.6.2 Adverse Events of Special/Potential Interest

An adverse event of special/potential interest (AESI/AEPI) is one of scientific and medical interest specific to understanding of the IP and may require close monitoring and rapid communication by the investigator to the sponsor. The AESI categories for each IP can be found in protocol Section 5.3. Other categories may be added, or existing terms may be modified, as necessary. The most recent version of the AESI/AEPI preferred terms list at clinical data lock will be used to identify and categorize AESI/AEPI.

The AESIs/AEPIs, as applicable, will be summarized similarly as other AEs described in Section 3.6.1.

Immune-mediated adverse events (imAEs) for durvalumab will be automatically adjudicated per the global imAE charter. The imAEs will be summarized, including separate summaries of the number and percentage of subjects with imAEs by category, by outcome, by maximum CTCAE grade and descriptive statistics of time to onset of imAE by AESI category.

3.6.3 Deaths and Treatment Discontinuations due to Adverse Events

For those subjects who died due to one or multiple AEs, the AE(s) contributing to death will be summarized.

Summaries will be provided for TEAEs resulting in permanent discontinuation of IP. Supporting listings will be provided for AEs resulting in death and AEs resulting in permanent discontinuation of IP.

3.6.4 Clinical Laboratory Evaluation

Laboratory tests will be grouped according to chemistry, hematology, coagulation, and urinalysis. Listings will be provided for all laboratory results, including coagulation. For selected chemistry and hematology tests, the change in each laboratory parameter from baseline to each post-baseline scheduled visit will be summarized graphically. Descriptive statistics will be provided for the clinical laboratory results and changes from baseline by scheduled time of evaluation including minimum and maximum post-baseline values.

Laboratory parameters will be assessed at baseline as well as throughout the study. Frequencies of worst observed grade 0-4 toxicity, as defined by the NCI CTCAE v5.0, will be presented for each laboratory parameter. The analysis will present the worst grade

observed and the rates of subjects with grade 3-4 toxicity. Additionally, 2-grade shifts from baseline to worst toxicity grade will be summarized for chemistry and hematology laboratory parameters. Separate summaries indicating hyper- and hypo- directionality of change will be produced, where appropriate. Shifts from baseline relative to the normal range (low, normal, high) will also be presented.

Liver Function Parameters

Subjects with elevated post-baseline alanine aminotransferase (ALT), aspartate aminotransferase (AST) or total bilirubin that fall into the following categories will be identified. The number and percentage of these subjects will be summarized.

Liver Function Parameters	**Category**
ALT	• ≥ 3× – ≤ 5× ULN • > 5× – ≤ 8× ULN • > 8× – ≤ 10× ULN • > 10× – ≤ 20× ULN • > 20× ULN
AST	• ≥ 3× – ≤ 5× ULN • > 5× – ≤ 8× ULN • > 8× – ≤ 10× ULN, • > 10× – ≤ 20× ULN • > 20× ULN
Total bilirubin	• ≥ 2× – ≤ 3× ULN • > 3× – ≤ 5× ULN • > 5× ULN
ALT or AST	• ≥ 3× – ≤ 5× ULN • > 5× – ≤ 8× ULN • > 8× – ≤ 10× ULN, • > 10× – ≤ 20× ULN • > 20× ULN
Potential Hy's law	• (AST ≥ 3 × ULN or ALT ≥ 3 × ULN) and (Total Bilirubin ≥ 2×ULN)[a]

ALT = alanine aminotransferase; AST = aspartate aminotransferase; ULN = upper limit of normal range.

[a]: Total bilirubin ≥ 2× ULN is defined as at least one case of post-dose total bilirubin ≥ 2 x ULN occurred after ALT or AST ≥ 3 x ULN post treatment.

Subjects who meet the potential Hy's law criteria will be listed.

Assessment of Nephrotoxicity

Creatinine clearance rate (C_{Cr}) will be calculated using serum creatinine (S_{Cr}) in mg/dL and the Cockcroft-Gault formula to estimate glomerular filtration rate (GFR), defined as follows:

$$C_{Cr}\ (\text{mL/minute}) = \{((140 - \text{age}) \times \text{weight})/(72 \times S_{Cr})\}(\times 0.85, \text{if female})$$

A shift table from baseline to "worst-case" on treatment C_{Cr} value will be provided. Baseline and "worst-case" on treatment C_{Cr} value will be categorized as follows:

- Normal: ≥ 90 mL/min
- Mild Impairment: ≥ 60 − < 90 mL/min
- Moderate Impairment: ≥ 30 − < 60 mL/min
- Severe Impairment: ≥ 15 − < 30 mL/min
- Kidney Failure: < 15 mL/min

3.6.5 Other Safety Evaluations

3.6.5.1 Vital Signs

Vital signs will be assessed at baseline and throughout the study. Vital signs will be summarized by study visit which will include descriptive statistics for the value of the parameters and the changes from baseline to post-baseline scheduled visits. Minimum and maximum post-baseline values will also be summarized..

3.6.5.2 Electrocardiogram

No electrocardiogram analyses are planned. ECGs were only collected prior to study treatment initiation.

3.6.5.3 Eastern Cooperative Oncology Group Performance Status

Eastern Cooperative Oncology Group (ECOG) performance status will be assessed at baseline as well as throughout the study. ECOG will be summarized for baseline and the "worst" case including descriptive statistics for the value of the parameters and the shift table from baseline to the "worst" performance post-baseline.

3.6.5.4 Physical Examinations

No physical examination analyses are planned.

3.6.6 Subgroup Analyses

No subgroup analyses are planned.

3.7 Immunogenicity

Immunogenicity results will be summarized for the As-treated population. Number and percentage of subjects in the following categories will be provided.

- ADA positive at baseline and/or post-baseline visits.
- Persistent positive, defined as positive at ≥2 post-baseline assessments (with ≥16 weeks between first and last positive) or positive at last post-baseline assessment.
- Transient positive, defined as negative at last post-baseline assessment and positive at only one post-baseline assessment or at ≥2 post-baseline assessments (with <16 weeks between first and last positive).
- Treatment-boosted, defined as baseline positive ADA titer that was boosted to a 4-fold or higher-level following drug administration.

Similarly to safety analyses, only post-baseline visits up to and including 90 days following the date of last dose of study medication will be considered when deriving the categories. Study discontinuation blood samples will be summarized at the closest nominal time point that does not already have a value. All subjects with titer information will be shown in data listings.

3.8 Pharmacokinetics

For each treatment group in the PK-evaluable population, serum concentrations of the study drug will be summarized by visit/timepoint with descriptive statistics. Individual novel agent concentrations will be summarized similarly. Serum concentrations will be reported in ug/mL, and the lower limit of quantification (LLOQ) for each analyte will be defined in the SPP. The following statistics are presented: n, n below LLOQ, geometric mean, geometric coefficient of variation (%), arithmetic mean, standard deviation, coefficient of variation (%), median, minimum, and maximum.

More details about calculations can be found in the SPP. All serum concentration data will be listed for each subject, by analyte.

4 INTERIM ANALYSIS

Interim analyses comparing durvalumab in combination with novel agents to durvalumab alone will be performed in a continuous manner using a joint Bayesian predictive probability. Bayesian predictive probability has been used in the setting of single-arm trials with tumor response endpoints such as OR and allows for continuous assessments for early Go/No-Go decision making (Lee and Liu, 2008). A joint Bayesian predictive probability approach

allows for continuous assessments of the delta (Δ), or difference, of the DCRs between durvalumab/novel agent combinations and durvalumab alone (Pulkstenis et al., 2017).

Interim analyses will begin after approximately 30 subjects in control arm (durvalumab alone) have been randomized and have reached the data cutoff criterion. The data cutoff criterion is defined as the opportunity to be followed for at least 16 weeks at the time of the data cutoff (i.e., dosed at least 16 weeks prior to the time of the data cutoff). The interim analysis will be based on the subset of the ITT Population that have satisfied the data cutoff criteria. Enrollment will not pause during the interim analysis before the decision is made.

Following this initial interim analysis, subsequent interim analyses will be performed for each experimental arm regardless of whether it starts enrollment later than control arm. Specifically, an interim analysis will be performed once each experimental arm has approximately 30 ITT subjects reach the data cutoff criterion. For interim analyses that will lead to an early No-Go decision, a response is defined as either a confirmed or unconfirmed CR or PR as per RECIST v1.1. Only concurrent control subjects, i.e., enrolled during the same period as the experimental arm, will be included in interim analyses. A comparison of each experimental arm versus all control subjects may be included as a supportive analysis.

The Target Value (TV) is set to demonstrate a 15% increase in ΔDCR at 16 weeks, where ΔDCR is the difference in DCR between an experimental arm and control arm. At final analysis, No-Go criteria will be met if the probability that the true ΔDCR exceeds the prespecified TV of 15% is less than 10% (i.e., No-Go: P [ΔDCR > TV] < 10%).

Given the existing observed data during the continuous monitoring stage, the joint Bayesian predictive probability is obtained by calculating the probability of reaching a No-Go decision should the treatment groups be randomized and evaluated to the maximum planned final sample size of 60. Further randomization into both treatment groups will be terminated if No-Go criteria are met.

- Early No-Go is met if it is predicted that there is a high probability of reaching a No-Go decision upon full randomization of 60 subjects given the existing observed data (i.e., predictive probability of a No-Go decision > 95%).

Table 4.1 illustrates the sample algorithm to make an early No-Go decision based on DCR at 16 weeks at the interim analysis when control arm and an experimental arm each have 30 subjects.

Table 4.1 Criteria of Early No-Go Decisions Based on DCR at 16 Weeks

Number of Subjects with DCR at 16 weeks in Control Arm	Number of Subjects with DCR at 16 weeks in Experimental Arm
10	≤ 7
11	≤ 8
12	≤ 8
13	≤ 9
14	≤ 10
15	≤ 11
16	≤ 12
17	≤ 13
18	≤ 14
19	≤ 15
20	≤ 16
21	≤ 17
22	≤ 19
23	≤ 20
24	≤ 21
25	≤ 23
26	≤ 24
27	≤ 26
28	≤ 28
29	≤ 29
30	≤ 30

DCR = disease control rate.

Operating characteristics of this continuous monitoring method are presented in Table 4.2 based on 10000 simulations using the prespecified TV.

Table 4.2 Operating Characteristics of Continuous Monitoring for No-Go

True δ (BM DCR = 60%)	All No-Go (Interim & final)	Early No-Go (Confirmed at final)
0%	68%	19% (18%)
5%	46%	11% (10%)
10%	25%	5% (4%)
15%	11%	2% (2%)
20%	3%	1% (0)
25%	1%	0 (0)

Criteria: Target Value (TV) $\delta = 15\%$; No-Go: $Pr(\delta > TV) < 10\%$; Early No-Go: $PP(\text{No-Go}) > 95\%$);
BM DCR: Benchmark Disease Control Rate.

5 REFERENCES

Efron B. The efficiency of Cox's likelihood function for censored data. J Am Stat Assoc. 1977:72(359):557-65.

Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Statistics In Medicine. 1991;10:1665-77.

Eisenhauer EA, Therasseb P, Bogaertsc J, Schwartzd LH, Sargente D, Fordf R, Danceyg J, Arbuckh S, Gwytheri S, Mooneyg M, Rubinsteing L, Shankarg L, Doddg L, Kaplanj R, Lacombec D, Verweijk J. European Journal of Cancer. 2009;45:228-247.

Klein JP, Logan B, Harhoff M, Anderson PK. Analyzing survival curves at a fixed point in time. Statistics in Medicine. 2007; 26:4505-19.

Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clin Trials. 2008; 5:93-106.

Pulkstenis E, Patra K, Zhang J. A Bayesian paradigm for decision-making in proof-of-concept trials. Journal of Biopharmaceutical Statistics. 2017; 27(3):442-456.

Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, Lin NU, Litière S, Dancey J, Chen A, Hodi FS, Therasse P, Hoekstra OS, Shankar LK, Wolchok JD, Ballinger M, Caramella C, de Vries EG; RECIST working group. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncology. 2017 Mar;18(3): e143-e152.

6 VERSION HISTORY

Version	Date	Summary of Changes	Reason for Change
1.0	01FEB2019	Initial document	Initial document
2.0	11DEC2019	• Revised the ITT definition • Clarifications added about including concurrently enrolled control subjects in efficacy analyses.	• To be aligned with protocol AM3. • To address comments received from the format review meeting. • Administrative changes.

		• Modified/Added subgroups for efficacy analyses. • Added a missing superscript in the formula of PFS-12 testing. • Updated signature page to include new team members. • Post cCRT window adjusted to 42 days to match Protocol AM3. • Added language about coagulation and thyroid analyses.	
3.0	14DEC2020	• Additional summaries added to Section 3.3.2. • New subgroups added to Section 3.4.4. • Added PFS-6 and PFS-9 in Section 3.4.5.	• Layout update as per review comments from IA#1.
4.0	07FEB2023	Updates made for final CSR delivery: 1. Section 3.3.1 updated to include listing of global/country situation disruptions. 2. Section 3.3.2 updated to describe categorization of platins. 3. Section 3.6.2 updated with minor clarifications as well as description of durvalumab imAE analyses. 4. Additional subgroup added to Section 3.4.4 for ORR. 5. Subgroup analyses included for PFS. 6. Section 3.4.2.4 (PFS) updated to include additional landmarks at 18 and 24 months. 7. Section 3.4.2.5 (OS) updated to include additional landmarks at 18 and 24 months. 8. Section 3.4.2.5 (OS) updated to include additional field for deriving last known alive date. 9. Section 3.4.4 updated to include forest plots for subgroup analyses of ORR and PFS. 10. Section 3.4.5 updated with clarification that unconfirmed ORR will also be reported. 11. iRECIST confirmed not to be reported in CSR (Section 3.4.5). Other minor clarifications were made throughout.	1. Required listing to describe impact of COVID-19 on the study. 2. Previously this was not well-defined and had been subject to manual review. The team agrees all terms should be coded and therefore WHO Drug information can be used to derive categories. 3. Automatic adjudication of durvalumab imAEs is a required process. Outputs have been added to comply with this. 4. Disease stage at study entry (IIIA vs. IIIB/C) is considered a clinically important subgroup. 5. Though confirmed ORR is the primary endpoint, PFS is also clinically relevant and therefore subgroups will be explored. 6. Data maturity permits further landmarks to be presented. 7. Data maturity permits further landmarks to be presented. 8. Update provides further coverage of potential last known alive date and increased accuracy. 9. Forest plots provide a clear way to present subgroup results. 10. This is per the CSP but was not previously clarified in the SAP. 11. Confirmed with clinical team that iRECIST is not required for the CSR.

5.0	04SEP2023	1. Added a few terms in the list of abbreviations 2. Section 3.1 clarified for the definition of baseline 3. Updated SAS version to 9.4 4. Table 3.2-1 Added PK-evaluable population 5. Section 3.3.3 removed data cut-off from duration of exposure definition 6. Section 3.3.4 add ATC and WHO Drug version 7. Section 3.3.5 Added COVID-19 related IPDs 8. Section 3.4.2.4 Added PFS-36 month landmark 9. Section 3.4.2.4 Updated and clarified the Greenwood's formula, and added more details 10. Section 3.4.2.5 Added OS-36 month landmark 11. Section 3.5 Updated section to clarify that no pharmacodynamic analyses are planned at this stage 12. Section 3.6 Added clarification regarding any derivation timeframe (in Labs, VS, ECOG) 13. Section 3.6.4 Updated wording to increase readiness and match what the expectations are, for the laboratory tables 14. Section 3.6.4 Creatinine clearance rate formula added 15. Section 3.6.5.1 Removed 10% subjects wording and added minimum and maximum post-baseline summaries of vital signs 16. Section 3.6.5.2 Added information as to why no ECG analyses are planned for the CSR 17. Section 3.6.5.4 Added a section for physical examinations (none planned) 18. Section 3.8 Added PK analysis plan 19. Section 3.8 Mentioned that all results to be reported in ug/mL Other minor clarifications were made throughout.	1. Previously missing some abbreviations, and new ones corresponding to new updates 2. Lack of information in defining baseline when no time indicator captured 3. Latest SAS version available 4. SAP now includes PK analysis 5. No DCO defined in final CSR 6. Clarify that the summary of concomitant medications is by ATC and PT 7. COVID-19 related IPDs might be required for the CSR 8. Agreed with the team in prior communications to include PFS-18, PFS-24 and PFS-36 landmarks for the CSR 9. Lack of precision in the formula details (n_i refers to deaths or disease progressions in the PFS section, difference between i and k) 10. Agreed with the team in prior communications to include OS-18, OS-24, and OS-36 landmarks for the CSR 11. Previous version outdated (MedImmune process) - Wording confirmed with PK/PD team 12. Discussed during the Dry Run review with the team to consider only post-baseline visits up to and including 90 days post last dose of study drug 13. Wording was unclear and disorganized, clarification was needed 14. Information more complete with additional details provided 15. CSP doesn't mention a selection criterion (10% of subjects) so not needed here, but mentions min/max summaries of vital signs 16. CSP mentions ECG analyses while ECGs were only collected at screening 17. CSP mentions physical examination statistical analysis,

			but none are planned for the CSR (data do not allow) 18. After discussion with the study team and the PK/PD team, PK analysis plan needed to be included in the SAP. Previous version outdated (MedImmune process). 19. After discussing with the PK/PD team, Durvalumab serum concentrations should be converted from ng/mL to ug/mL

7 APPENDIX

7.1 Derivation of RECIST v1.1 Disease Assessment Overall Response

Guidance regarding the handling of dropouts and missing data will apply uniformly to all efficacy analyses resulting from an application of RECIST v1.1 to investigator assessed tumor measurements. For investigator reported outcomes, analyses will present outcomes reported by the investigator without consideration of missing data or censoring rules.

7.1.1 Target Lesion Response

Target lesion response will be programmatically derived on the data collection instrument once RECIST v1.1 criteria are applied to the site personnel recorded target lesion measurements.

Possible values include:

- CR – Complete Response
- PR – Partial Response
- SD – Stable Disease
- PD – Progressive Disease
- NE – Non-evaluable
- NA – Not Applicable (*set value for all post-baseline disease assessments only if no target lesions are identified at baseline*)

The derivation for target lesion response is as follows *(please note the order of the algorithm below is important)*:

1. If "Any Target Lesions Present" equals "No" on the *Target Lesions – Baseline* CRF, then all post-baseline "Target Lesion Response" equals "NA".

2. Else, if "Percent Change from Nadir Sum of Diameters" is greater than or equal to 20% and the absolute increase from the nadir (defined as the "Total" for each post-baseline disease assessment minus the "Nadir Sum of Diameters") is greater than or equal to 5 mm, then "Target Lesion Response" equals "PD".

3. Else, if "Not Done" is selected, or "Measurement" is left blank, or "Lesion no longer Measurable" is selected and equal to "NE", or "Lesion Intervention" is selected for any Target Lesion identified at Baseline, then "Target Lesion Response" equals "NE".

4. Else, if "Total Non-Lymph Node" equals "0" and all Lymph Node Target Lesion "Measurements" are less than "10" individually, then "Target Lesion Response" equals "CR".

   Note: This step requires examining "Measurements" separately for Target Lesions with "Lymph Node" equal to "Yes" and "No".

5. Else, if "Percent Change from Baseline Sum of Diameters" is less than or equal to -30%, then "Target Lesion Response" equals "PR".

6. Else, "Target Lesion Response" equals "SD".

If a subject has a missing tumor measurement at a disease assessment for 1 or more target lesions, the sum of diameters (longest for non-nodal lesions, short axis for nodal lesions) will be reported for the remaining target lesions. These data will be used to indicate radiologic disease progression if the sum of diameters for the observed lesions increases at least 20% from the nadir sum of diameters of all target lesions and demonstrates at least a 5 mm absolute increase from the nadir sum of diameters of all target lesions, in spite of the missing data (or if other criteria for PD are met).

7.1.2 Non-Target Lesion Response

Non-target lesion response will be assigned by site personnel following a qualitative overall assessment of all non-target lesions.

Possible values include:

- CR – Complete Response

  - Disappearance of all non-target lesions and normalization of tumor marker level. All lymph nodes must be non-pathological in size (<10mm short axis)

- Non-CR/Non-PD – Non-Complete Response / Non-Progressive Disease
  - Persistence of one or more non-target lesion(s) and/or maintenance of tumor marker level above the normal limits

- PD – Progressive Disease
  - Unequivocal progression of existing non-target lesions.

- NE – Non-evaluable

- NA – Not Applicable (*set value for all post-baseline disease assessments only if no non-target lesions are identified at baseline*)

Though non-target lesion responses are a subjective decision made by the site personnel, certain responses may be limited depending on the non-target lesion statuses recorded. An algorithm is provided below highlighting appropriate possible non-target lesion responses based on recorded data. Reaching a red box (■) signifies having reached the only allowable non-target lesion responses based on non-target lesion statuses. Reaching a green box (■) signifies having reached the end of the algorithm and more than one possible non-target lesion response is possible from which the Investigator may choose.

1. a) If no non-target lesions are identified at baseline, all post-baseline non-target lesion responses should equal NA. ■
   b) Else, if any non-target lesions are identified at baseline, responses may be limited to CR, Non-CR/Non-PD, PD, NE (i.e., responses of NA are not permitted).
   Go to Rule 2.
2. a) If all non-target lesions have a status are “Absent”, the responses may be limited to CR. ■
   b) Else, if at least one non-target lesion status is NOT “Absent”, the responses may be limited to Non-CR/Non-PD, PD, NE (i.e., responses of CR, NA are not permitted).
   Go to Rule 3.
3. a) If all non-target lesions have a status of “Unequivocal Progression”, responses may be limited to PD. ■
   b) Else, if no non-target lesions have a status of “Unequivocal Progression”, responses may be limited to Non-CR/Non-PD, NE (i.e., responses of CR, PD, NA are

not permitted).
Go to Rule 4.
c) Else, if at least one (but not all) non-target lesion has a status of "Unequivocal Progression", the responses may be limited to Non-CR/Non-PD, PD, NE (i.e., responses of CR, NA are not permitted). (*Note: No response has been eliminated as an option here.*)
Go to Rule 5.

4. a) If all non-target lesions have a status of "Non-evaluable" and/or "Not Done" is selected, responses may be limited to NE. ■
   b) Else, if no non-target lesions have a status of "Non-evaluable" and "Not Done" is not selected, responses may be limited to Non-CR/Non-PD (i.e., responses of CR, PD, NE, NA are not permitted). ■
   c) Else, if at least one (but not all) non-target lesion has a status of "Non-evaluable" and/or "Not Done" is selected, the responses may be limited to Non-CR/Non-PD, NE (i.e., responses of CR, PD, NA are not permitted). ■
   (*Note: No response has been eliminated as an option here.*)
5. a) If all non-target lesions have a status of "Non-evaluable" and/or "Not Done" is selected, responses may be limited to NE. ■
   b) Else, if no non-target lesions have a status of "Non-evaluable" and "Not Done" is not selected, responses may be limited to Non-CR/Non-PD, PD (i.e., responses of CR, NE, NA are not permitted). ■
   c) Else, if at least one (but not all) non-target lesion has a status of "Non-evaluable" and/or "Not Done" is selected, the responses may be limited to Non-CR/Non-PD, PD, NE (i.e., responses of CR, NA are not permitted). ■
   (*Note: No response has been eliminated as an option here.*)

If a subject has a missing tumor status at a disease assessment for 1 or more non-target lesions, radiologic disease progression will be determined if the remaining non-target lesions qualitatively demonstrate unequivocal progression (or if other criteria for PD are met).

7.1.3 Disease Assessment Overall Response per RECIST v1.1

Investigator visit disease response will be programmatically derived on the data collection instrument using RECIST v1.1 criteria based upon target lesion response, non-target lesion response, and new lesion data. Missing values in any of target lesion response, non-target lesion response, and new lesion data will result in the disease response not being derived.

Possible values include:

- CR – Complete Response

- PR – Partial Response
- SD – Stable Disease
- PD – Progressive Disease
- NE – Non-evaluable

Target Lesion Response	Non-Target Lesion Response	New Lesion	Derived RECIST Disease Response
CR	CR or NA	No	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	CR or Non-CR/Non-PD or NE or NA	No (or NE)	PR
SD	CR or Non-CR/Non-PD or NE or NA	No (or NE)	SD
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NE	CR or Non-CR/Non-PD or NE or NA	No	NE
NA	CR	No	CR
NA	Non-CR/Non-PD	No	SD (Non-CR/Non-PD)[a]
NA	NE or NA	No (or NE)	NE
NA	CR or Non-CR/Non-PD	NE	SD (Non-CR/Non-PD)[a]

[a] Per RECIST v1.1, "SD (Non-CR/Non-PD)" is preferred over "SD" for non-target disease since SD is increasingly used as endpoint for assessment of efficacy in some trials so to assign this category when no lesions can be measured is not advised.

(*Note: "(or NE)" values under New Lesion will only be included in confirmation of progression or confirmation of new lesions are required per protocol. The last 4 rows may be eliminated from any study that requires identification of at least one measurable lesion at Baseline. One may choose to allow such cells to remain if an independent central review is included in the trial.)*

If a subject has a missing tumor measurement at some assessment(s) for 1 or more target lesions and criteria for PD are not otherwise met, an overall response of NE will be assigned for the assessment(s).

7.1.4 Locoregional therapy

Any subject receiving locoregional therapy, including surgery, while on study that directly affects one or more of the target lesions selected at baseline will be identified. A subject with a subsequent response or SD will be considered to be non-evaluable at all disease assessments that occur on or after the date of locoregional therapy. Otherwise, the subject will be assessed ignoring the locoregional therapy.

7.1.5 Assignment of Dates of Disease Progression or Disease Response

For all analyses of endpoints resulting from an application of RECIST v1.1 to investigator assessed tumor measurements, there may be cases in which disease assessments span a series of dates. For establishing the start date of a subsequently confirmed response in which the disease assessment spans multiple days, the date of response assigned will be the latest date of evaluations corresponding to the disease assessment. The date of latest evaluation will also be assigned for a mid-study assessment showing SD as the date assigned for the purposes of censoring duration of response, TTP and PFS.

The date of PD will be the first date at which any objective diagnostic test provides data indicating PD. Specifically, for RECIST v1.1 the date of PD will be the earliest of the following 3 evaluation dates:

- Date of PD as indicated by target lesions: If PD is triggered by a change in sum of diameters of target lesions, and the dates of evaluation of the target lesions vary for the same assessment, assign the first evaluation date among target lesions.

- Date of PD as indicated by non-target lesions: If the dates of evaluation of the non-target lesions vary for the same assessment, assign the first evaluation date for which any non-target lesion exhibits a status of Unequivocal Progression.

- Date of PD as indicated by new lesions: If multiple new lesions are identified and the dates of evaluation for the new lesions vary for the same assessment, assign the first evaluation date for which any new lesion is detected.

In scenarios where the Investigator disease response is either a response or PD and differs from that of the application of RECIST v1.1 to investigator assessed tumor measurements separate response and/or progression dates will be required. Determination of the start date of a subsequently confirmed response in which the disease assessment spans multiple days remains the same as described previously. Specifically, the date of response assigned will be the latest date of evaluations corresponding to the disease assessment. The date of PD will be the earliest date of evaluations corresponding to the disease assessment.

7.2 Efficacy Criteria of Making an Internal Go Decision at an Interim Analysis

At an interim analysis, an internal Go decision may be made based on the delta (Δ), or difference, of the ORRs between durvalumab/novel agent combinations and durvalumab

alone. The internal Go decision may initiate further development of the combinations. The Go decision will not only be based on meeting the pre-specified efficacy criteria but also the totality of data including safety profiles, data from other studies, and internal portfolio prioritization.

For interim analyses that will lead to an internal Go decision, a response is defined as a confirmed CR or PR as per RECIST v1.1. The Lower Reference Value (LRV) is set to an 8% increase in ΔORR to demonstrate clinically meaningful increase over durvalumab alone. At final analysis, Go criteria will be met if the probability that the true ΔORR exceeds the prespecified LRV of 8% is more than 80% (i.e., Go: P[ΔORR > LRV] > 80%).

Early Go criteria are met if it is predicted that there is a high probability of reaching a Go decision upon full randomization of 60 subjects given the existing observed data (i.e., predictive probability of a Go decision > 95%). If Go criteria are met, randomization may continue, and planning activities for further development of the durvalumab/novel agent combination may be triggered.

Table 7.2-1 illustrates the sample algorithm to make an early Go decision based on ORR at an interim analysis when control arm and an experimental arm each have 30 subjects.

Table 7.2-1 Criteria of Early Go Decisions Based on ORR

Number of Subjects with OR in Control Arm	Number of Subjects with OR in Experimental Arm
3	≥ 12
4	≥ 13
5	≥ 14
6	≥ 16
7	≥ 17
8	≥ 18
9	≥ 19
10	≥ 20
11	≥ 21
12	≥ 22
13	≥ 23
14	≥ 24
15	≥ 25
16	≥ 25
17	≥ 26
18	≥ 27
19	≥ 28
20	≥ 28
21	≥ 29

Table 7.2-1 Criteria of Early Go Decisions Based on ORR

Number of Subjects with OR in Control Arm	Number of Subjects with OR in Experimental Arm
22	30

ORR = objective response rate.

Operating characteristics of this continuous monitoring method are presented in Table 7.2-2 based on 10000 simulations using the prespecified LRV.

Table 7.2-2 Operating Characteristics of Continuous Monitoring for Go

True δ (BM ORR = 30%)	All Go (Interim & final)	Early Go (Confirmed at final)
0%	3%	1% (0)
8%	19%	3% (3%)
15%	47%	10% (10%)
20%	69%	18% (17%)
25%	85%	31% (30%)

Criteria: Lower Reference Value (LRV) δ = 8%; Go: Pr(δ > LRV) > 80%; Early Go: PP(Go) > 95%);
BM ORR: Benchmark Objective Response Rate.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d9108c00001-sap-ed-5		
Document Title:	Statistical Analysis Plan Edition 5	
Document ID:	Doc ID-005095573	
Version Label:	5.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
12-Sep-2023 13:27 UTC	PPD [redacted]	Author Approval
11-Sep-2023 18:32 UTC	PPD [redacted]	Content Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.