Statistical Analysis Plan

A Phase 2 Open-label, Multicenter, Randomized, Multidrug Platform Study of Neoadjuvant Durvalumab Alone or in Combination with Novel Agents in Subjects with Resectable, Early-stage (I [>2cm] to IIIA) Non-small Cell Lung Cancer (NeoCOAST)

Protocol Number: D9108C00002

TABLE OF CONTENTS

LIST OF TABLES

List of Abbreviations

Abbreviation or Specialized Term	Definition
AE	adverse event
AEPI	adverse event of possible interest
AESI	adverse event of special interest
BMI	body mass index
CMH test	Cochran–Mantel–Haenszel test
ctDNA	circulating tumor DNA
CR	complete response
DA	disease assessment
DBL	database lock
DCO	data cutoff
DMPK	Clinical Pharmacology & Drug Metabolism and Pharmacokinetics
eCRF	electronic Case Report Form
ECOG	Eastern Cooperative Oncology Group
GFR	glomerular filtration rate
imAE	immune-mediated adverse event
IP	investigational product
ITT	intent-to-treat
IxRS	interactive voice/web response system
mRNA	messenger RNA
LRV	lower reference value
MPR	major pathological response
NSCLC	non-small cell lung cancer
ORR	objective response rate
pCR	pathological complete response
PD	progressive disease
PK	pharmacokinetics
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SD	stable disease
SPP	statistical programming plan
TV	target value

1 INTRODUCTION

This document describes the statistical analysis methodology for protocol D9108C00002, a multidrug platform study of neoadjuvant Durvalumab alone or in combination with novel agents in subjects with resectable, early-stage (I [>2cm] to IIIA) 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 assess the antitumor activity of durvalumab alone and/or in combination with novel agents, in terms of major pathological response (MPR) rate.

2.1.2 Secondary Study Objectives

Safety:

- To assess the feasibility of receiving the planned surgical resection;
- To assess the safety and tolerability of durvalumab alone and/or in combination with novel agents.

Efficacy:

- To assess the antitumor activity of durvalumab alone and/or in combination with novel agents, in terms of pathological complete response (pCR) rate.

Pharmacokinetics (PK):

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

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

[redacted]

- [redacted]

2.2 Study Design

NeoCOAST is a Phase 2, open-label, multicenter, randomized, multidrug platform study of durvalumab alone or in combination with novel agents in subjects with resectable, early-stage (Stage I [>2cm] to IIIA) NSCLC.

The study will evaluate the clinical activity and safety of neoadjuvant durvalumab alone or in combination with novel agents in subjects with resectable, early-stage NSCLC (Stage I [>2 cm] to IIIA).

Treatment arms may be opened sequentially or in parallel. Subjects will be treated with durvalumab alone or in combination with novel agents for up to 28 days, followed by surgical resection. After surgical resection, subjects will be followed up to Day 105 (Figure 1 in the protocol). Up to approximately 25 sites globally will participate in this study.

New durvalumab combination therapy arms may be added based on emerging nonclinical 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 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 lymph node involvement (Yes vs. No) and will use an equal ratio to all study treatment arms open for enrollment (e.g., if treatment arms are opened sequentially, a treatment arm is added/closed, or enrollment in a treatment arm is suspended). At the onset, the randomization scheme will use an equal ratio in all treatment arms. After 20 subjects have been enrolled into any treatment arm, the randomization ratio can be adjusted.

Blinding

The study is not blinded.

2.4 Sample Size

This study is designed to obtain preliminary clinical efficacy, safety, PK and immunogenicity data on durvalumab in combination with novel agents compared to durvalumab monotherapy. It is not designed to make explicit power and Type I error considerations for a hypothesis test.

At the onset, subjects will be randomized with equal ratios into each treatment arm. The initial arms included are: durvalumab monotherapy, durvalumab plus oleclumab, durvalumab plus monalizumab, and durvalumab plus danvatirsen. A total of up to 40 subjects per arm may be enrolled. An interim analysis will be performed in any given arm, once a respective arm enrolls at least 20 efficacy evaluable subjects.

Table 2.4-1 shows estimated differences in MPR rate between durvalumab combination therapy arm compared to the durvalumab monotherapy arm, along with 95% Wald confidence intervals (CIs), with different assumptions and a sample size of 40 subjects each assuming an asymptotic normality. The assumed 30% MPR rate with the durvalumab monotherapy arm is based on a pooled analysis of data from two recently published clinical trials evaluating anti-PD-1/PD-L1 monotherapies in a similar patient population. One study (Forde et al, 2018, see references in the protocol) reported an MPR rate of 45% (9/20) for subjects who received nivolumab monotherapy; while another study (Ruschet et al, 2018, see references in the protocol) reported an MPR rate of 20% (10/50) for subjects who received atezolizumab monotherapy. An arithmetic average of these two MPR rates is rounded to 30% as the assumed MPR rate for durvalumab. Forty subjects per treatment arm provides a 95% CI with reasonable width (~± [redacted]%) for the estimated difference between a durvalumab combination therapy arm and the durvalumab monotherapy arm. When the MPR rate for a durvalumab combination therapy arm is [redacted]%, the lower limit of the 95% CI for the difference between the durvalumab combination therapy arm and the durvalumab monotherapy arm is [redacted], which is greater than [redacted] suggesting that [redacted]

Table 2.4-1 Estimated Differences in MPR Rate Between Durvalumab Monotherapy Arm and Each Durvalumab Combination Therapy Arm (40 Subjects Each)

Number (%) of Responders		Difference (%) in MPR (2-sided 95% Wald CI)
Durvalumab Combination Therapy Arm (n = 40)	Durvalumab Monotherapy Arm (n = 40)	
16 (40%)	12 (30%)	10% (-11%, 31%)
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

CI = confidence interval; MPR = major pathological response.

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% confidence level. Baseline values will be defined as the last valid assessment prior to the first administration of investigational product (IP).

Data analyses will be conducted using the SAS® System (SAS Institute, Inc., Cary, NC, USA) Version 9.3 or above, unless otherwise specified. All analysis outputs will be validated according to AstraZeneca/MedImmune SAS programming standards and AstraZeneca/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 investigational product. Subjects will be analyzed according to the treatment they actually receive.
Efficacy Evaluable Population	The Efficacy Evaluable Population includes subjects from the As-treated Population who have the opportunity to be followed for the planned surgery by the time of the data cutoff, and either have pathological tumor response data post-surgery, or did not undergo surgical resection due to death or disease progression

All baseline and efficacy parameters will be summarized based on the ITT population as primary analysis. Safety parameters will be summarized based on the As-treated Population as primary analysis. The efficacy parameters may be summarized based on efficacy evaluable population as supportive analysis and interim analyses of efficacy will be performed on this population. Interim analysis will be performed on ITT population as sensitivity analysis if deemed appropriate.

3.3 Study Subjects

3.3.1 Subject Disposition and Completion Status

Subject disposition and completion status will be summarized for all subjects screened. Summaries of the number and percentage of subjects entered at each site will be provided. In addition, disposition of subjects throughout the study with respect to discontinuation of treatment and end of study status including reasons 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 investigational product or disease under investigation, or other).

3.3.2 Demographics and Baseline Characteristics

Demographics and baseline disease characteristics will be summarized for the ITT population. Demographic information at study entry related to sex, age, race, ethnicity, ECOG performance status, weight, height, and body mass index (BMI) will be presented by treatment group and for all subjects combined.

Tumor history including histology, stage, and pertinent biomarker results (if data is available) at the time of initial diagnosis and at study entry will be summarized. A summary of smoking history will also be included.

3.3.3 Study Drug Exposure

Treatment exposure will be summarized for the As-treated population. Duration of exposure to IP(s) in days, except for single-dose IP(s), will be summarized by descriptive statistics and by frequency.

The duration of exposure for each IP is defined as below. More details, if necessary, will be provided in the Statistical Programming Plan (SPP).

- last dose date plus the dosing interval (days) minus first dose date for subjects on treatment.
- minimum of (date of death plus 1 day, data cutoff date plus 1 day, last dose date plus the dosing interval) minus first dose date for subjects who are no longer on treatment.

Dose intensity will be listed, and relative dose intensity to IP(s) will be summarized by descriptive statistics. The relative dose intensity is a percentage 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) will be summarized with reasons for deviations for the following categories: delays, omissions, and interruptions. Dosing delays will be derived based on the scheduled dosing dates or the date of the most recent dose. 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.

3.3.4 Concomitant Medications

The number and percentage of subjects who took concomitant medications will be summarized by the generic name coded by World Health Organization (WHO) Drug Dictionary 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

Incidence of important protocol deviations will be summarized by deviation categories. 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. Disruptions caused by the COVID-19 pandemic will be listed and tabulated. Tabulation of important protocol deviations will be done including and excluding COVID-19 related deviations.

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 MPR rate. MPR rate is defined as the proportion of subjects with ≤ 10% residual viable tumor cells by investigator assessment. The primary analysis of MPR rate will be based on the ITT population. The MPR rate and the corresponding 95% CI will be reported for each arm. In addition, an estimate of the difference in MPR rate as well as its 2-sided 95% exact CI between each combination therapy arm and monotherapy arm will be reported. The rates between any durvalumab combination therapy arm and the durvalumab monotherapy arm will be evaluated and compared from the Cochran–Mantel–Haenszel (CMH) test stratified by the baseline lymph node involvement (yes or no). In case the randomization ratio between monotherapy arm and other combination therapy arms changes during the study, the CMH test may be additionally stratified by the randomization ratio of monotherapy arm vs other combination therapy arms according to randomization scheme (if allowed by strata size). The Fisher's exact test will also be conducted. Subjects that do not have the planned surgery will be considered as failures/non-responders, and will therefore be counted in the denominator, but not in the numerator of MPR rate. The primary efficacy analysis will only include subjects in monotherapy arm who are concurrently enrolled with each combination therapy arm in order to conduct rigorous between-arm comparisons. Concurrently enrolled subjects from monotherapy arm (to a respective combination therapy arm) are subjects randomized to monotherapy arm who also had opportunity to be randomized to that respective combination therapy arm. The number of subjects in monotherapy arm who are concurrently enrolled with the respective combination therapy arm will be tabulated along with the statistical inferences.

MPR assessed by independent central reviewer may be analyzed by the same approach when data is available.

3.4.1.2 Additional Analyses of the Primary Efficacy Endpoint(s)

The first additional analysis of MPR rate may be based on the efficacy evaluable population.

Additional analysis of MPR may be a Bayesian logistic regression using power prior (Ibrahim et al, 2015), performed separately for each combination therapy arm to monotherapy arm comparison. The logistic regression will include treatment arm and baseline lymph node involvement as explanatory variables. In case the randomization ratio between monotherapy arm and other combination therapy arms changes during the study, the model may be additionally adjusted for the randomization ratio of monotherapy arm vs other combination therapy arms according to the subject randomization scheme by adding a corresponding covariate (if allowed by size of covariate categories).

The power prior is intended to implement informative knowledge into prior distribution constructed from data of non-concurrent subjects in monotherapy arm. The weight of the non-concurrent data in monotherapy arm relative to the concurrent data in monotherapy arm will be optimized by Deviance Information Criterion (DIC). The odds ratio between each combination therapy arm and monotherapy arm will be presented, as well as its 95% credible interval and the posterior probability of the odds ratio being greater than 1 (i.e. the probability that the subjects in combination therapy arm has higher odds to achieve MPR than those in the monotherapy arm).

Another additional analysis of MPR may also be conducted by including all subjects from monotherapy arm for treatment comparisons if deemed appropriate, regardless of whether a monotherapy arm subject was enrolled concurrently or not with the respective experimental arms.

3.4.2 Secondary Efficacy Endpoint(s) and Analyses

The secondary efficacy endpoint is pCR rate. pCR rate is defined as the proportion of subjects with no viable tumor cells by investigator assessment, and the corresponding 95% CI will be reported for each arm. The primary analysis of pCR rate will be based on the ITT population and may be based on the efficacy evaluable population as a supportive analysis. In addition, the difference between any durvalumab combination therapy arm and the durvalumab monotherapy arm will be evaluated. Comparison of each of the combination arms with the monotherapy arm will be obtained from a stratified Cochran–Mantel–Haenszel test as described for the primary efficacy endpoint. The Fisher's exact test will be also conducted. Subjects who do not have the planned surgery will be considered as failures/non-responders, and will therefore be counted in the denominator, but not in the numerator of pCR rate. In the foregoing analysis, only subjects in the monotherapy arm who are

concurrently enrolled with combination therapy arm will be included. The number of subjects in monotherapy arm who are concurrently enrolled with the respective combination therapy arm will be tabulated along with the statistical inferences.

pCR assessed by independent central reviewer may be analyzed by the same approach when data is available.

An additional Bayesian logistic regression of the secondary efficacy endpoints using power prior and involving non-concurrent subjects data may also be performed, as described for the primary efficacy endpoint. Moreover, additional analysis may also be conducted by including all subjects from monotherapy arm for treatment comparisons if deemed appropriate, regardless of whether a monotherapy arm subject was enrolled concurrently or not with the respective experimental arms.

3.4.3 Exploratory Efficacy Endpoint(s) and Analyses

[redacted]

Efficacy analyses pertaining to objective response will be based on an application of RECIST 1.1 (Eisenhauer et al, 2009) to investigator assessed tumor measurements.

Programmatic derivation guidance used for the application of RECIST 1.1 are provided in Appendix 8.1. RECIST 1.1 guidelines will be used to determine disease response.

3.4.3.1 Best Overall Response

Best overall response (BOR) will be based on all post-baseline disease assessments that occur prior to surgery. 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). Radiographical confirmation of a response is not feasible due to the length of the study (e.g., a confirmatory scan will not be performed). The primary analysis of BOR will be based on the ITT population.

In general, subjects not classifiable under the RECIST 1.1 response categories due to insufficient data or early death will be classified as non-evaluable (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 progressive disease criteria is met despite the missing data, the subject will be classified as PD.

3.4.3.2 Objective Response Rate

Objective response rate (ORR) is defined as the proportion of subjects with a best overall response of CR or PR that occur prior to surgery and the corresponding 95% CI will be reported for each arm. Radiographical confirmation of a response is not feasible due to the length of the study (e.g., a confirmatory scan will not be performed). The primary analysis of ORR will be based on the ITT population. In addition, the difference between any durvalumab combination therapy arm and the durvalumab monotherapy arm will be evaluated. Comparison of each of the combination therapy arms with the monotherapy arm will be obtained from a stratified Cochran–Mantel–Haenszel test as described for the primary efficacy endpoint. The Fisher's exact test will be also conducted. 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. In the foregoing analysis, only subjects in the monotherapy arm who are concurrently enrolled with combination therapy arm will be included. The number of subjects in monotherapy arm who are concurrently enrolled with the respective combination therapy arm will be tabulated along with the statistical inferences.

An additional Bayesian logistic regression for the ORR using power prior and involving non-concurrent subjects data may also be performed, as described for the primary efficacy endpoint. Moreover, additional analysis may also be conducted by including all subjects from monotherapy arm for treatment comparisons if deemed appropriate, regardless of whether a monotherapy arm subject was enrolled concurrently or not with the respective experimental arms.

3.4.3.3 Other Endpoint

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 that occurs prior to surgery. The percentage change from baseline in target lesion sum of diameters is defined as follows:

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

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 that occur prior to surgery. 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.

3.4.4 Handling of Dropouts and Missing Data

In general, missing data are not imputed for statistical analysis. Guidance regarding the handling of dropouts and missing data and censoring will apply uniformly to all efficacy analyses resulting from an application of RECIST 1.1 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.

3.4.5 Subgroup Analyses

No subgroup analyses are planned.

3.4.6 Exploratory Analyses

[redacted]

3.4.7 Sensitivity Analyses

Sensitivity analyses may be performed for selected efficacy analyses where subjects who are randomized but violated inclusion criteria #4 or exclusion criteria #2 will be excluded. These are the subjects with incorrect disease, incorrect staging, prohibited prior therapy, etc. The same analyses will be conducted except on a smaller patient population due to exclusion of aforementioned subjects.

3.5 Pharmacodynamic Endpoint(s) and Analyses

Descriptive statistics will be the primary method for the biomarker analyses. Depending on the nature of the data, geometric mean and other appropriate statistical summaries might be used as well.

[redacted]

3.6 Safety Analyses

All safety analyses will be performed based on the As-treated Population, unless otherwise specified. AEs before and after surgery will be summarized separately where it is of scientific interest. In addition, safety data collected after initiation of subsequent anticancer therapy will not be summarized in the tables and figures, unless specified otherwise, but will be listed.

3.6.1 Adverse Events and Serious Adverse Events

Adverse events will be coded by Medical Dictionary for Regulatory Activities (MedDRA) and be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events version 5.0 (NCI CTCAE v5.0), and the type incidence, severity and relationship to study investigational product 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 highest severity 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 event (TEAE) is defined as events present at baseline that worsen in intensity after administration of study investigational product or events absent at baseline that emerge after administration of study investigational product. All TEAEs that occurred on and after first dose to the end of follow-up period 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 the end of follow-up period will be listed.

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

3.6.2 Adverse Events of Special Interest and Adverse Event of Possible Interest

An adverse event of special interest (AESI) and adverse event of possible interest (AEPI) are of scientific and medical interest specific to understanding of the investigational product and may require close monitoring and rapid communication by the investigator to the sponsor. Detailed descriptions/terms of AESIs for each IP can be found in protocol Section 5.3. Other

categories may be added or existing terms may be modified as necessary. Preferred terms used to identify each AESI group will be clarified/confirmed before database lock (DBL).

3.6.2.1 Adverse Events of Special Interest Associated with Oleclumab

The AESIs associated with Oleclumab will be summarized similarly as other AEs described in Section 3.6.1.

3.6.2.2 Adverse Events of Special Interest Associated with Durvalumab

AESI, AEPI, immune-mediated AESI, and immune-mediated AEPI will be derived and summarized following the latest version of Durvalumab Global imAE Characterization Charter.

3.6.3 Deaths and Treatment Discontinuations due to Adverse Events

For those subjects who died due to an AE, the AE contributing to death and the AE's relationship to IP will be listed.

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, urinalysis, and thyroid function analysis, where applicable. Listings will be provided for all laboratory results, including urinalysis. For chemistry and hematology tests, the change from baseline to each post-baseline visit will be summarized graphically for selected parameters of scientific interest. Descriptive statistics may be provided for the clinical laboratory results and changes from baseline by scheduled time of evaluation including maximum and minimum 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 in which CTCAE grade can be numerically derived. The analysis will present worst grade observed and the rates of subjects with Grade 3-4 toxicity. A shift table, presenting the 2-way frequency tabulation for baseline and post-baseline grade at scheduled time of evaluation as well as the worst post-baseline grade, will be provided from clinical laboratory tests. Also, laboratory parameters will be assessed by presenting tables containing information related to at least 1-grade shift and at least 2-grade

shift from baseline. Separate summaries indicating hyper- and hypo- directionality of change will be produced, where applicable. For those laboratory parameters that cannot be graded by the NCI CTCAE, the frequencies of the post-baseline laboratory value categorized as low (L), normal (N), or high (H) using laboratory reference range from baseline will be summarized.

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. Number and percentage of these subjects will be tabulated.

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.

Individual subject data where elevated ALT or AST plus total bilirubin fall into the "Potential Hy's law" will be listed.

Assessment of Nephrotoxicity

Creatinine clearance rate (C_{Cr}) will be calculated using serum creatinine and the Cockcroft-Gault formula to estimate glomerular filtration rate (GFR). Shift tables from baseline to "worst-case" among post-baseline C_{Cr} values will be provided. Baseline and "worst-case" post-baseline C_{Cr} value will be categorized for the following categories:

- 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 may be summarized by study visit which may include descriptive statistics for the value of the parameters and the changes from baseline by scheduled time of evaluation and by treatment arm including end of treatment visit as well as for the maximum and minimum post-baseline values.

3.6.5.2 Electrocardiogram

Electrocardiogram analyses may be performed.

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 by study visit which may include descriptive statistics for the value of the parameters and changes from baseline by scheduled time of evaluation and by treatment arm.

3.6.5.4 Feasibility to Surgery

Feasibility to surgery is defined as the proportion of subjects in each arm with no delay in the planned surgery (no later than 42 days after Week 1, Day 1). The feasibility to surgery will be evaluated/tabulated.

3.6.6 Subgroup Analyses

No subgroup analyses for safety are planned.

3.7 Immunogenicity

Only subjects who receive at least 1 dose of durvalumab and/or other combination study drug, and provide at least 1 post-treatment sample, will be evaluated. For each treatment arm, the immunogenic potential of durvalumab will be assessed by summarizing the number and percentage of subjects who develop detectable ADAs to durvalumab. For treatment arms that include a novel biologic agent, the immunogenic potential of the novel biologic agent will be assessed by summarizing the number and percentage of subjects who develop detectable ADAs to the novel biologic agent.

All valid assay results from subjects who receive any study drug will be included in immunogenicity summaries. All subjects with titer information will be shown in the data listing.

3.8 Pharmacokinetics

Pharmacokinetic data analyses will be performed by Clinical Pharmacology & Drug Metabolism and Pharmacokinetics (DMPK) group or its designee.

4 INTERIM ANALYSIS

Interim analyses will be conducted during the course of the study to evaluate the clinical activity and safety of any treatment arm. The first planned interim analysis will be initiated once at least 1 treatment arm has enrolled 20 efficacy evaluable subjects. Across all treatment arms, subject accrual beyond 20 subjects will be determined based on interim data. Bayesian predictive probabilities will be used to evaluate clinical activity (Lee and Liu, 2008). In this study, a target value (TV) of ΔMPR as [redacted] is set, where ΔMPR is the difference in MPR rate between a combination therapy treatment arm and the durvalumab monotherapy arm.

- A treatment arm will meet the No-Go criteria if the probability that the true ΔMPR being larger than TV is less than [redacted], i.e., Prob.(True ΔMPR > [redacted] < [redacted]. A No-Go decision will be made at the interim analysis if the predictive probability of meeting the No-Go criteria (after 40 subjects are all enrolled) given observed interim data are greater than [redacted], i.e., Predictive Probability(Prob.[ΔMPR > [redacted]] < [redacted]) > [redacted]

Table 4-1 illustrates the sample algorithm to make an early No-Go decision based on MPR at the interim analysis when combination arm and durvalumab monotherapy each have 20 subjects.

Table 4-1 Criteria of Early No-Go Decisions Based on MPR

Number of Subjects with MPR in Durvalumab Monotherapy Arm	Number of Subjects with MPR in Durvalumab Combination Therapy Arm
2	≤ 2
3	≤ 2
4	≤ 3
5	≤ 4
6	≤ 5
7	≤ 5
8	≤ 6
9	≤ 7
10	≤ 8
11	≤ 9

MPR = major pathological response.

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 MPR = 30%)	All No-Go (Interim & final)	Early No-Go (Confirmed at final)
0%	78%	36% (35%)
10%	40%	15% (12%)
15%	23%	8% (6%)
20%	12%	4% (2%)
25%	5%	2% (1%)
30%	2%	1% (0)

Criteria: Target Value (TV) δ = [redacted]; No-Go: Pr(δ > TV) < [redacted]; Futility: PP(No-Go) > [redacted]);
BM MPR: Benchmark Major Pathological Response Rate.

A treatment arm will be considered futile if it meets the futility bar for MPR according to Figure 7 in the protocol. If at least 1 treatment arm has more than 20 efficacy evaluable subjects' data at the interim analysis, slightly different futility bars compared to Figure 7 in the protocol will be provided on a case-by-case basis. In addition, for any treatment arm, enrollment will be stopped if more than 2 out of the first 20 enrolled subjects failed the feasibility to surgery.

5 DEVIATIONS FROM PLANNED ANALYSES

The definition of ITT population is slightly revised for aligning with ICH E9. The definition of efficacy evaluable population is updated in order to reflect the practical clinical considerations for the disease setting in this study.

6 REFERENCES

Ibrahim JG, Chen MH, Gwon Y, Chen F. The power prior: theory and applications. Statistics in medicine. 2015 Dec 10;34(28):3724-49.

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. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1) European Journal of Cancer. 2009;45:228-247.

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

7 VERSION HISTORY

Version	Date	Summary of Changes	Reason for Change
1.0	15Feb2019	Initial document	Initial document
2.0	15May2020	• Revised the definition of ITT population to align with the sponsor's current standard definition • Revised the definition of Efficacy Evaluable Population to reflect the practice of neoadjuvant therapy, advised by the Clinical Lead • Clarifications added about only including concurrently enrolled subjects from monotherapy arm in primary efficacy analyses • Added additional efficacy analyses utilizing non-concurrently enrolled subjects from monotherapy arm • Additionally conduct interim analysis based on ITT population as sensitivity analysis, if deemed appropriate. • Updated signature page to reflect personnel changes and new SOP • Added languages about coagulation analyses • Clarified the endpoints related to each study objectives • Corrected the unit of study drug exposure from cycles to days • Updated/added the analyses of AESI, AEPI, or imAE • Updated Table 4.1 to reflect the targeted MPR rate of 30% at monotherapy • Removed the analyses of stratified and unstratified chi-square test as it is equivalent to some other analyses we already planned to perform	Address comments raised by study team

		• Corrected the typos identified, minor issues, and/or administrative items	
3.0	24Nov2020	• [redacted] • Added sensitivity analyses of efficacy excluding subjects who had incorrect disease or setting or prohibited prior therapy at study entry, suggested by clinical team. • For the purpose of data interpretation, excluded safety data collected after initiation of subsequent anticancer therapy from safety analyses as suggested by clinical team. • Added descriptive reports of disruptions due to pandemic per study team suggestions. • Indicated the analyses deviated from what is planned in the protocol, and provided the rationale of deviations. • Administrative edits such as correcting typos, spacing, updating list of signatory due to personnel changes, etc.	Address issues pertaining to upcoming interim analysis
4.0	10Aug2021	• Editorial updates such as correcting typos, table numbering, etc. • Remove some analyses where data is not collected in this study. • Remove some analyses not indicated in the protocol and deemed unnecessary after team discussions. • Clarify or correct some descriptions of the analyses done by standard templates. • Add some languages for conducting additional analyses deemed scientifically appropriate per study team.	Address outstanding comments and issues identified during the CSR dry-run

8 APPENDIX

8.1 Derivation of RECIST 1.1 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 1.1 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.

8.1.1 Target Lesion Response

Target lesion response will be programmatically derived on the data collection instrument once RECIST 1.1 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).

8.1.2 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).

8.1.3 Disease Assessment Overall Response per RECIST1.1

Investigator visit disease response will be programmatically derived on the data collection instrument using RECIST 1.1 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 1.1, "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).

8.1.4 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.

8.1.5 Assignment of Dates of Disease Progression or Disease Response

For all analyses of endpoints resulting from an application of RECIST 1.1 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

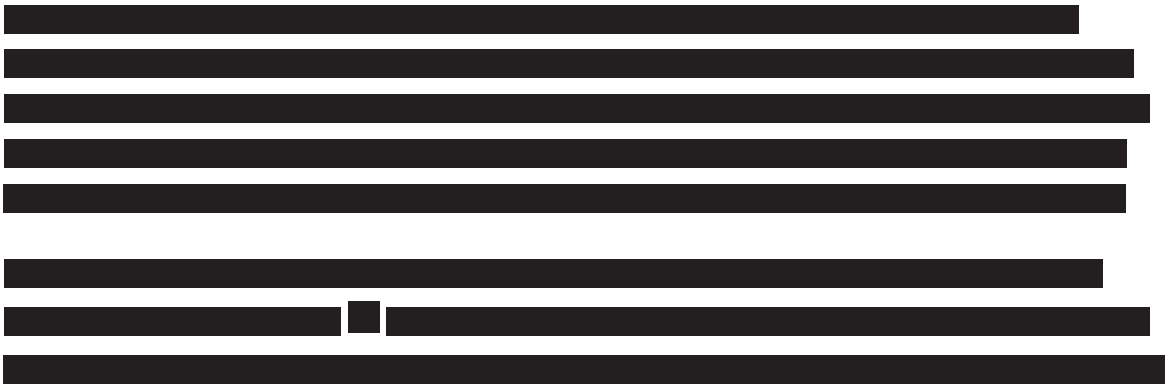
be assigned for a mid-study assessment showing SD as the date assigned for the purposes of censoring duration of response, TTP and PFS.

The date of PD will be the first date at which any objective diagnostic test provides data indicating PD. Specifically, for RECIST 1.1 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 1.1 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.

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



[redacted]

[redacted]

[redacted]

Table 8.2-1 Criteria of Early Go Decisions Based on MPR

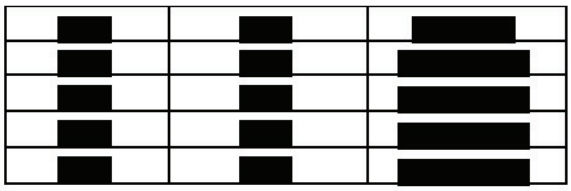
Number of Subjects with MPR in Durvalumab Monotherapy	Number of Subjects with MPR in Durvalumab Combination Therapy Arm
3	[redacted]
4	[redacted]
5	[redacted]
6	[redacted]
7	[redacted]
8	[redacted]
9	[redacted]
10	[redacted]
11	[redacted]
12	[redacted]
13	[redacted]

MPR = major pathological response.

Operating characteristics of this continuous monitoring method are presented in Table 8.2-2 based on 10000 simulations using the prespecified LRV.

Table 8.2-2 Operating Characteristics of Continuous Monitoring for Go

True δ (BM MPR = 30%)	All Go (Interim & final)	Early Go (Confirmed at final)
[redacted]	[redacted]	[redacted]



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: d9108c00002-sap-ed4		
Document Title:	Statistical Analysis Plan Edition 4	
Document ID:	Doc ID-004645333	
Version Label:	4.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
16-Aug-2021 15:45 UTC	[redacted]	Content Approval
16-Aug-2021 15:29 UTC	[redacted]	Content Approval

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