Statistical Analysis Plan

A Phase 1b/2, Open-label, Multicenter Study of Novel Oncology Therapies in Combination with Chemotherapy and Bevacizumab as First-Line Therapy in Metastatic Microsatellite-Stable Colorectal Cancer (COLUMBIA-1)

Protocol Number: D910CC00001

TABLE OF CONTENTS

LIST OF TABLES

List of Abbreviations

Abbreviation or Specialized Term	Definition
AE	Adverse Event
BMI	Body Mass Index
CCI	CCI
CR	Complete Response
CRC	Colorectal Cancer
DoR	Duration Of Response
DA	Disease Assessment
DC	Disease Control
DCO	Data Cutoff
DCR	Disease Control Rate
DMPK	Clinical Pharmacology & Drug Metabolism And Pharmacokinetics
eCRF	Electronic Case Report Form
ECOG	Eastern Cooperative Oncology Group
GFR	Glomerular Filtration Rate
HR	Hazard Ratio
CCI	CCI
CCI	CCI
CCI	CCI
CCI	CCI
CCI	CCI
CCI	CCI
Ig	Immunoglobulin
IP	Investigational Product
ITT	Intent-To-Treat
IV	Intravenous
IxRS	Interactive Voice/Web Response System
CCI	CCI
MedDRA	Medical Dictionary For Regulatory Activities
MSS-CRC	Microsatellite Stable Colorectal Cancer
LRV	Lower Reference Value
OR	Objective Response
ORR	Objective Response Rate
OS	Overall Survival
OS-12	Overall Survival At 12 Months
PD	Progressive Disease
PFS	Progression-Free Survival
PFS-12	Progression-Free Survival At 12 Months
PK	Pharmacokinetics

Abbreviation or Specialized Term	Definition
PR	Partial Response
RECIST	Response Evaluation Criteria In Solid Tumors
SAE	Serious Adverse Event
SD	Stable Disease
SPP	Statistical Programming Plan
CCI	CCI

1 INTRODUCTION

This document describes the statistical analysis for D910CC00001 Clinical Study Protocol Amendment 3 dated 11 May 2022, a multidrug platform study to evaluate the safety and efficacy of standard of care in combination with novel oncology therapies (durvalumab and oleclumab) in subjects with metastatic microsatellite stable (MSS) colorectal cancer (CRC) who have received no prior therapy in the recurrent/metastatic setting.

2 STUDY OVERVIEW

2.1 Study Objectives

2.1.1 Primary Study Objectives

Safety:

- **Part 1:** To evaluate the safety and tolerability profile of FOLFOX + bevacizumab + novel oncology therapy combinations

Efficacy:

- **Part 2:** To compare the efficacy of FOLFOX + bevacizumab + novel oncology therapy combinations versus FOLFOX + bevacizumab

2.1.2 Secondary Study Objectives

Safety:

- **Part 2:** To evaluate the safety and tolerability profile of FOLFOX + bevacizumab + novel oncology therapy combinations

Efficacy:

- **Part 1:** To investigate the preliminary antitumor activity of FOLFOX + bevacizumab + novel oncology therapy combinations

Clinical activity:

- **Part 2:** To further compare the efficacy of FOLFOX + bevacizumab + novel oncology therapy combinations versus FOLFOX + bevacizumab

Pharmacokinetics (PK) & Immunogenicity:

- **Part 1 & 2:** To describe the PK of novel agents
- **Part 1:** To describe the PK of bevacizumab
- **Part 1 & 2:** To describe the immunogenicity of applicable novel agents

- **Part 1:** To describe the immunogenicity of bevacizumab

CCI [redacted]

2.2 Study Design

Study D910CC00001 is a Phase 1b/2, open-label, multicenter, randomized, multidrug platform study to evaluate the safety and efficacy of standard of care (FOLFOX plus bevacizumab) in combination with novel oncology therapies (durvalumab and oleclumab) in subjects with 1L metastatic MSS-CRC.

The study is designed to concurrently evaluate potential novel combinations with clinical promise using a 2-part approach. Part 1 is a Phase 1b study of safety, and Part 2 is a Phase 2 study of efficacy and safety. The treatment regimens evaluated in Part 2 will depend on the evaluation of safety outcomes in Part 1. Up to approximately 30-40 sites globally will participate in this study.

Following a screening period of up to 28 days, subjects will be centrally assigned (Part 1) or randomized (Part 2) to one of the open study arms. In both study parts, study treatment may be administered until disease progression or any discontinuation criteria are met. CCI [redacted]

CCI [redacted]

2.2.1 Safety Run-in (Part 1)

2.2.1.1 CCI [redacted]

CCI [redacted]


CCI

Table 1 CCI

CCI

2.2.2 Randomized Phase

Once Part 1 has been completed for a given arm, Part 2 will open for randomization to evaluate the efficacy and safety of an arm with the same treatment regimen. Randomization will be evenly distributed across all open arms (eg, 1:1, 1:1:1) initially and will be stratified based on location of the primary tumor (right sided vs left sided). After 50 subjects are randomized to the control arm, the control arm will continue to enroll subjects, but the allocation ratio to the different arms may be adjusted at the discretion of the sponsor. Enrollment to the control arm will pause if there is no active experimental arm. The exact subsequent allocation ratios across arms will be determined by the sponsor during the conduct of the study (Section 2.3).

2.3 Treatment Assignment

After eligibility is confirmed, subjects will be centrally assigned (Part 1) or randomized (Part 2) to study treatment using an interactive voice/web response system (IXRS). Randomization will be stratified by location of the primary tumor (right sided vs left sided). In Part 2, a randomization method with dynamically changing randomization ratios will be employed to account for fluctuation in the number of treatment arms open for randomization over the course of the study. At the onset, the randomization scheme will use an equal ratio to all study treatment arms open for randomization.

2.4 Sample Size

Part 1:

A minimum of 6 subjects (up to 12 subjects per dose level of novel agent) will be enrolled to complete the safety run-in. More subjects may be enrolled if dose de-escalation is needed.

CCI

Part 2:

CCI

3 STATISTICAL METHODS

3.1 General Considerations

Tabular summaries will be presented by treatment arm. Two sets of summary tables will be produced for each Part (Part 1, Safety run-in and Part 2, Randomized Phase). 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). Data listings will be sorted by treatment arm, 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.

Data analyses will be conducted using the SAS® System (SAS Institute, Inc., Cary, NC, USA) Version 9.3 or above, unless otherwise specified.

3.1.1 Evaluating the Impact of the COVID-19 Pandemic

Subjects with confirmed or suspected COVID-19 infections will be identified using MedDRA coded adverse events. Subjects with confirmed or suspected COVID-19 deaths will be identified using the MedDRA coded primary/secondary cause. A listing of visits impacted by the Pandemic will be produced.

3.2 Analysis Populations

The analysis populations are defined in Table 2.

Table 2 Analysis Populations

Population	Description
Intent-to-treat (ITT) Population	The ITT Population includes subjects who receive any study IP. 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 actually 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 8 weeks at the time of the DCO (ie, dosed at least 8 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.
DLT evaluable Population	The DLT evaluable population includes subjects enrolled in the dose-escalation phase who receive the full prescribed dose of durvalumab and ≥ 75% of the prescribed number of doses of FOLFOX plus bevacizumab and the other novel oncology therapy and complete the safety follow-up through the DLT evaluation period (defined as 28 days after the initiation of study investigational product) or experience any DLT.
PK evaluable Population	The PK evaluable population include subjects who receive at least 1 dose of IP with at least 1 reportable PK concentration.
ADA evaluable Population	The ADA evaluable population includes subjects who receive at least 1 dose of IP who have a non-missing baseline ADA result and at least 1 non-missing post-baseline ADA result.

Table 3 Summary of Outcome Variables and Populations

Outcome Variable	Population
Demographic and baseline characteristics	As-treated, Part 1 ITT, Part 2
Safety Data	As-treated
Adverse events	As-treated
Laboratory measurements	As-treated
Vital Signs/ECG/Physical examination/Other safety assessments	As-treated
DLTs	DLT evaluable
Efficacy Data	
Best Objective Response (BOR)	As-treated, Part 1 ITT and Response Evaluable, Part 2
Objective Response Rate (ORR)	As-treated, Part 1 ITT and Evaluable for Response, Part 2
Duration of Response (DoR)	As-treated, Part 1 ITT and Evaluable for Response, Part 2

Table 3 Summary of Outcome Variables and Populations

Outcome Variable	Population
Disease Control (DC)	As-treated, Part 1 ITT and Evaluable for Response, Part 2
Change in tumor size	As-treated, Part 1 ITT, Part 2
Progression-free Survival (PFS)	As-treated, Part 1 ITT, Part 2
Overall Survival (OS)	As-treated, Part 1 ITT, Part 2
Pharmacokinetics	
Serum PK concentration data	PK evaluable
Immunogenicity	
Serum ADA data	ADA evaluable

3.3 Study Subjects

3.3.1 Subject Disposition and Completion Status

Subjects screened/randomized is defined as all subjects that completed the informed consent. The number and percentage of subjects screened, (randomized for Part 2) and treated will be provided along with the reasons for screen failure (did not meet inclusion/exclusion criteria, lost to follow-up, withdrawal of consent, other), discontinuation of each study treatment (adverse event, progressive disease, lack of efficacy, lost to follow-up, initiation of subsequent anticancer treatment, protocol violation, pregnancy, subject decision, investigator decision, other) and end of study status (death, lost to follow-up, withdrawal by subject, other) for each Part and treatment arm. An overall column will also be included for Part 2. Individual end of study status will be presented for all treated subjects for Part 1 and for all randomized subjects for Part 2. Individual reasons for screen failure will be presented in a listing.

In addition, a summary of all subjects with confirmed or suspected COVID-19 infections and confirmed or suspected COVID-19 deaths will be provided.

3.3.2 Demographics and Baseline Characteristics

Demographic information related to age, age group (<65, ≥65), sex, race, ethnicity, weight, height, and body mass index (BMI) at screening will be summarized by Part and treatment arm. Descriptive statistics will be calculated for the continuous variables age, weight, height and body mass and for categorical variables age group, sex, race, ethnicity.

Disease stage diagnosis (I, II, III, IV, unknown) and at study entry (III, IIIA, IIIB, IIIC, IV, IVA, IVB, IVC, unknown), whether recurrence of local/locally advanced cancer or disease had metastasized (yes/no), histology (adenocarcinoma, mucinouse adenocarcinoma, signet-ring cell carcinomia, high-grade carcinoma, large cell neuroendocrine carcinoma, small cell neuroendocrine carcinoma, squamous cell carcinoma, adenosquamous carcinoma, medullary carcinoma) histology grade (G1, G2, G3, G4, GX, unknown), site of involvement at study entry (bone, brain/CNS, distant lymph nodes, local lymph nodes, liver, lung, spleen, other visceral), laterality of primary cancer (left, right, transverse, unknown), dihydropyrimidine dehydrogenase deficiency test result (positive, negative) will be summarized by Part and treatment arm.

Summaries of demographics and baseline disease characteristics for all subjects with confirmed or suspected COVID-19 infections will also be provided.

The incidence of brain metastasis will be summarized and listed.

CCI

[redacted]

3.3.3 Medical History

Medical history will be coded using the MedDRA Dictionary, Version 24.0. The number and percentage of subjects will be presented by system organ class and preferred term for each Part and for treatment arm. A similar summary will be produced for all subjects with confirmed or suspected COVID-19 infections. All events will be listed.

3.3.4 Concomitant Medications

Concomitant medications will include all medications taken on or after the date of first dose of investigational product or any medication started prior to first dose of study treatment that continued beyond the date of first dose of investigational product. The number and percentage of subjects who took concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) classification and generic name coded according to the WHODrugGlobalB3: 202109.

3.3.5 Prior and Subsequent Anticancer Treatment

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 previous adjuvant regimens and the best response in regimen (complete response, partial

response, stable disease, progressive disease, not evaluable, and not done). A similar summary for subsequent anticancer treatment will be provided. All prior and subsequent systemic therapies, radiation, cancer related and other therapies will be listed.

3.3.6 Study Drug Exposure

Duration of exposure is defined by last date of actual dosing (ie, a dose was actually given) in the last cycle plus 1 cycle length minus the date of first treatment with IP. For subjects who die or if a DCO occurs prior to 1 cycle length of the last dose in the last cycle, duration of exposure is defined as date of death/DCO (whichever occurs first) minus the date of first treatment plus 1 day.

Duration of exposure (cycles) is defined as the number of cycles in which at least one portion of IP was administered.

The actual duration of exposure is calculated as:

Actual duration of exposure(weeks) = duration of exposure(weeks) - total duration of dose interruptions. Duration of exposure is calculated as above and dose interruptions are defined as any length of time where the subject has not taken any of the planned dose in accordance with the protocol.

Dose intensity of IP(s) is addressed by considering relative dose intensity (RDI), where RDI is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. More specifically, RDI is defined as follows:

- RDI = 100% × d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule.

Duration of exposure in weeks and cycles, and relative dose intensity will using descriptive statistics for each IP by Part and treatment arm. The number and percentage of subjects with relative dose intensity < 90%, 90 – 110% and > 110% will also be presented.

Dosing modifications for IP(s) will be summarized with reasons for deviations for the following categories: delays, omissions, reductions and interruptions for each study drug separately.

3.3.7 Protocol Deviations

Incidence of important protocol deviations will be summarized by deviation type (concomitant medications, informed consent form, inclusion/exclusion criteria, laboratory, other, procedures/tests, study drug, visit schedule, site). Listing will be provided with protocol deviation details.

A listing will be provided for subjects affected by the COVID-19 pandemic.

3.4 Efficacy Analyses

This section describes the analysis of the primary and secondary endpoints.

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, Eisenhauer et al, 2009). Objective response rate (ORR) is defined as the proportion of subjects with a programmatically derived best overall response of confirmed CR or confirmed PR based on the site investigators RECIST data that occur prior to the initiation of subsequent anticancer treatment. Confirmation of response requirements are defined in Section 3.4.2.1for best overall response.

The final analysis of confirmed ORR for Part 1 will be based on the As-treated Population. The primary analysis of confirmed ORR for Part 2 will be based on the ITT Population with the Response-evaluable Population as supportive.

The number and proportion subjects with confirmed ORR with two-sided exact binomial 95% CI using the Clopper-Pearson method will be presented. For Part 2, ORR an estimate of the odds ratio between the experimental arm and the control arm will be reported. The experimental arm will be compared with the control arm using the Cochran-Mantel-Haenszel (CMH) test stratified by the location of the primary tumor. **Secondary Efficacy Endpoint(s) and Analyses**

The secondary efficacy endpoints include best overall response (BOR), disease control (DC), duration of response (DoR), time to response, change in tumour size, progression-free survival (PFS), progression-free survival at 12 months (PFS-12) and overall survival (OS). The analyses for each outcome will be based on the populations as indicated in Table 3.

3.4.2.1 Best Overall Response

From the investigator's review of the imaging scans, the RECIST tumor response data will be used to determine each subject's overall visit response according to RECIST 1.1 using the

information from target lesions (TLs), non-target lesions (NTLs) and new lesion and depending on the status of their disease compared with baseline and previous assessments as described in Section 7.1 Best overall response (BOR) is the best response a subject has had following first dose but prior to starting any subsequent cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorization of BOR will be based on RECIST using the following response categories: complete response (CR); partial response (PR); stable disease (SD); progressive disease (PD); and non-evaluable (NE).

Confirmation of CR and PR is required and must occur no fewer than 4 weeks after initial documentation of CR or PR with both visits occurring prior to any subsequent cancer therapy.

For subjects whose progression event is death, BOR will be calculated based upon all evaluable RECIST assessments prior to death.

At least 8 weeks from first dose date of IP must elapse without a subsequently confirmed radiological disease progression in order to assign a best overall response of SD. This time frame will include 8 weeks plus the protocol-defined disease assessment window of 3 days. Subjects must have an overall visit assessment of SD on or after study day 54 (56 days – 3), with no prior evidence of progression.

3.4.2.2 Duration of Response

Duration of response (DoR) is defined as the time from the first documentation of a subsequently confirmed response until the first documentation of a disease progression or death due to any cause, whichever occurs first. Only subjects who have achieved confirmed response (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 response + 1) / (365.25/12)

The date of PD/death or censoring is the same as defined for PFS in Section 3.4.2.5

The median DoR and its 95% CI will be estimated using the Kaplan-Meier method if data will allow.

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. DCR between each experimental arm and control arm will be compared using the Cochran–Mantel–Haenszel test stratified by the location of the primary

tumor. Median DCR and its 95% CI will be calculated using the Kaplan-Meier method if data will allow.

3.4.2.4 Change in Tumor Size

The best change and the best percentage change in tumor size from baseline will be reported, i.e. the maximum reduction from baseline or the minimum increase from baseline in the absence of a reduction from baseline based on all post baseline assessments.

Tumor size is the sum of the longest diameters (or short axis measurements for lymph nodes) of the target lesions. Target lesions are measurable tumor lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to starting treatment. The percentage change in target lesion tumor size at each week x for which data are available will be obtained for each subject taking the difference between the sum of the target lesions at each week x and the sum of the target lesions at baseline divided by the sum of the target lesions at baseline multiplied by 100 (i.e. (week x - baseline)/baseline * 100).

For best percentage change, if it cannot be calculated due to missing data, a value of +20% is imputed as the best percentage change from baseline in the following situations (otherwise best percentage change is left as missing):

- If a subject has no post-baseline assessment and has died;
- If a subject has new lesions or progression of non-target lesions (NTLs) or TLs;
- If a subject has withdrawn due to PD and has no evaluable TL data before or at PD.

Waterfall plots indicating the best percentage change from baseline in sum of the diameters of target lesions will be produced. The plot will present each subject's best percentage change from baseline in TL tumor size as a separate bar, with the bars ordered from the largest increase to the largest decrease. A reference line at the –30% change in TL tumor size level will be added to the plots, which corresponds with the definition of 'partial' response. Additionally, spider plots of the percentage change from baseline over time for each subject will be produced.

3.4.2.5 Progression-free Survival

Progression-free survival (PFS) is defined as the time from randomization (first dose date for Part 1) 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. The censoring guidance and the date of PD/death or censoring are given Table 4. PFS is defined in months as follows:

PFS (months) = (Date of PD/death or censoring – Date of randomization [first dose date for Part 1] + 1) / (365.25/12).

However, if the subject progresses or dies immediately after two or more consecutive missed visits, the subject is censored at the time of the latest evaluable disease assessment prior to the two missed visits. Note: a NE visit is not considered as a missed visit.

PFS will be summarized using the Kaplan-Meier method and Kaplan-Meier curves plotted. The number and percentage of subjects experiencing a PFS event with the median PFS and its 95% CI will be presented . The Kaplan-Meir estimate of the proportion of subjects progression free and alive at 12 months (PFS-12) and associated 95% CI will also be presented.

The difference in PFS between each experimental arm and control arm in Part 2 will be tested for significance using a stratified log-rank test. 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 by location of primary tumor location (right sided vs left sided) Cox regression model with ties handled by the Efron method (Efron et al, 1977).

The comparison of PFS-12 between each experimental arm and control arm in Part 2 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.

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

where $\hat{S}_i(t), i = 1, 2$ is the Kaplan-Meier estimate of the survival function, $\hat{\sigma}_i(t)^2 = \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}, i = 1, 2$ is the variance for $\ln\hat{S}_i(t)$ derived from the Greenwood's formula, d_i and n_i refer to the number of deaths and number of subjects at risk at time t_i, respectively.

Table 4 Summary of Censoring Guidelines for PFS

Situation	Date of PD/Death or Censoring	PFS Outcome
Documented Progressive Disease (PD) or death in the absence of progression	Date of earliest sign of PD or date of death in the absence of progression	Event
Death prior to second scheduled post-baseline disease assessment and no tumor assessment at baseline/no evaluable assessments post-baseline	Date of death	Event
No death prior to second scheduled post-baseline disease assessment and either no tumor assessment at baseline or no evaluable assessments post-baseline	Date of randomization/first dose (Day 1)	Censored
Death or PD immediately after ≥ 2 consecutive missed assessments as per the protocol specified assessment schedule	Date of randomization/first dose or last evaluable progression-free disease assessment prior to missed assessments, whichever occurred last	Censored
On-going with neither PD nor death at time of analysis or lost to follow-up/ or withdrawn consent	Date of last evaluable disease assessment	Censored
Initiation of subsequent anticancer treatment	Date of last evaluable disease assessment prior to initiation of subsequent anticancer treatment	Censored for Sensitivity Analyses only

PD = progressive disease; PFS = progression-free survival

The PFS time will always be derived based on scan/assessment dates not visit dates.

RECIST assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- The date of progression will be determined based on the earliest of the dates of the component that triggered the progression.
- When censoring a subject for PFS the subject will be censored at the latest of the dates contributing to a particular overall visit assessment.
- Note: for TLs only the latest scan date is recorded out of all scans performed at that assessment for the TLs and similarly for NTLs only the latest scan date is recorded out of all scans performed at that assessment for the NTLs.
- It will be determined that two visits were missed, if no RECIST assessment was done for more than 2*(the protocolled time between scans + the protocol allowed visit window). The length, in days, that determines 2 consecutive missed visits is dependent on when the last evaluable assessment occurred. The length of time that occurred since the last evaluable disease assessment and the event (death or disease progression)

should be calculated by taking the difference between the subject's last evaluable assessment and date of death/progressive disease. Then using the timing of the last evaluable assessment before the event, if this difference it is greater than the number of days given below, then the subject is regarded as having two consecutive missed visits.

Table 5 Determination of Two Missed Visits

RECIST assessment schedule: Every 8 weeks up to Week 48, then 12 weeks thereafter			
	Last Evaluable RECIST assessment before event	Length of 2 missed visits	Length (days)
	Baseline	(2x 8 weeks) + 3 day window	115
W8	RECIST Visit 1	2x (8 weeks + 5 day window)	122
W16	RECIST Visit 2	2x (8 weeks + 5 day window)	122
W24	RECIST Visit 3	2x (8 weeks + 5 day window)	122
W32	RECIST Visit 4	2x (8 weeks + 5 day window)	122
W40	RECIST Visit 5	(8 weeks + 5 day window) + (12 weeks + 5 day window)	150
W48	RECIST Visit 6	2 x (12 weeks + 5 day window)	178
W60	RECIST Visit 7, 8 … etc	2 x (12 weeks + 5 day window)	178

3.4.2.6 Overall Survival

Overall survival (OS) is defined as the time from randomization (first dose date for Part 1) until death due to any cause regardless of whether the subject withdraws from study therapy or receives another anticancer therapy. 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 in severity dates
- Admission and discharge dates of hospitalization
- Study treatment date
- End of 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, CCI [redacted].
- Disease assessment dates on RECIST/CCI [redacted] CRF
- Date of visit, vital signs, ECOG, electrocardiogram, and physical examination
- Start and stop dates of subsequent anticancer treatment, subsequent systemic therapy, and subsequent other therapy
- Date of procedure / surgery of subsequent anticancer surgery
- Start and end date of concomitant medication and surgical/medical procedure
- Date last known alive on the survival status CRF
- End of study date

OS is defined in months as follows:

OS (months) = (Date of death or censoring – Date of randomization [first dose date for Part 1] + 1) / (365.25/12).

If a subject is known to have died where only a partial death date is available then the date of death will be imputed as the latest of the last date known to be alive +1 from the database and the death date using the available information provided:

- For Missing day only – using the 1st of the month.
- For Missing day and Month – using the 1st of January.

If there is evidence of death but the date is entirely missing, it will be treated as missing, i.e. censored at the last known alive date.

OS will be summarized using the Kaplan-Meier method and Kaplan-Meier curves plotted. The number and percentage of subjects experiencing an OS event with the median OS and its 95% CI will be presented.

The difference in OS between groups will be tested for significance using a stratified log-rank test stratified by location of primary tumor location (right sided vs left sided). 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 with ties handled by the Efron method (Efron et al, 1977).

3.4.3 Handling of Dropouts and Missing Data

In general, missing data will not be imputed for statistical analysis other than the imputation of best percentage change in tumor size as described in Section 3.4.2. Guidance regarding the handling of dropouts and 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 reported by the investigator without consideration of missing data or censoring rules.

3.5 Immunogenicity

Anti-Drug Antibody (ADA) results will be listed for each subject and summarized for the ADA evaluable Population. The number and percentage of subjects who have ADA positive results at baseline, post-baseline, both baseline and post-baseline, post baseline and not detected at baseline, persistent positive, transient positive, not detected post-baseline and positive at baseline, treatment boosted ADA and treatment-emergent ADA will be summarized by cohort. For subjects with positive ADA, the median, 25th percentile, 75th percentile, min and max of ADA titers will be summarized for each visit.

3.6 Pharmacokinetics

PK concentrations of bevacizumab, durvalumab and oleclumab will be summarized by visit and listed for each subject for the PK evaluable Population.

3.7 Safety Analyses

All safety analyses will be performed based on the As-treated Population.

3.7.1 Adverse Events and Serious Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as events present at baseline that worsen in intensity after administration of IP or events absent at baseline that emerge after administration of IP. TEAEs will be coded by MedDRA version 22.0 or more recent version and the type incidence, severity and relationship to IP will be summarized. Specific TEAEs will be counted once for each subject for calculating percentages. In addition, if the same TEAE occurs multiple times within a particular subject, the highest severity and level of

relationship observed will be reported. In addition, all TEAEs with an onset date no more than 90 days after the last dose of IP will be listed.

The number and percentage of subjects with TEAEs, as classified by system organ class (SOC) and preferred term (PT,) will be summarized:

- Subjects with TEAEs and the number of TEAEs
- TEAEs occurring in greater than 5% of subjects in any treatment arm
- TEAEs causally related to any study treatment
- TEAEs of CTCAE grade ≥ 3
- TEAEs by maximum CTCAE grade
- Serious TEAEs
- TEAEs resulting in death
- Serious TEAEs by maximum CTCAE grade
- TEAEs leading to discontinuation of study treatment

An overall summary table of the number of subjects experiencing each category of adverse event will be produced.

The adverse events (AEs) and serious adverse events (SAEs) considered to be non-treatment emergent meaning they occurred after the signing of the informed consent and prior to the initiation of IP will be listed.

Details of infusion related reactions will be listed.

Narratives will be produced for all subjects with confirmed or suspected COVID-19 SAEs identified using MedDRA coded adverse events.

3.7.2 Adverse Events of Special Interest

Adverse events of special interest (AESI) and Adverse Events of Possible Interest (AEPI) are defined for durvalumab and oleclumab. Detailed terms of AESIs and AEPIs for each IP can be found in protocol Section 5.3. Other categories may be added or existing terms may be modified as necessary based on ongoing safety surveillance. AESIs/AEPIs and immune-

mediated AESIs/AEPIs will be summarized in a similar manner to those TEAEs as described in Section 3.7.1

3.7.3 Clinical Laboratory Evaluation

Laboratory tests will be grouped according to chemistry, hematology, coagulation and urinalysis. Listings will be provided for all laboratory results. Box plots by visit will be presented for absolute values and change from baseline of hematology, chemistry and coagulation parameters by Part and treatment arm.

Laboratory abnormalities with toxicity grades according to the NCI CTCAE version 5.0 will be derived. Laboratory abnormalities occurring from the start of IP administration to the last assessment on study will be presented. Worst toxicity grade, rates of grade 3-4 toxicity, and grade shifts of 2 or more from baseline to the maximum grade will be presented. Summaries indicating hyper- and hypo- directionality of change will be produced, where appropriate. Laboratory parameters that cannot be graded via NCI CTCAE will be summarized with frequencies of post-baseline laboratory values categorized as low (L), normal (N), or high (H) using laboratory normal ranges compared to baseline.

3.7.3.1 Thyroid Function Parameters

The number and percentage of subjects with baseline TSH within the normal range at least one post-baseline TSH below the lower limit of normal (LLN) with concurrent free thyroxine above the upper limit of normal (ULN) and the number and percentage of subjects with baseline TSH within the normal range at least one post-baseline TSH above ULN with concurrent free thyroxine below LLN will be presented.

3.7.3.2 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.

Table 6 Liver Function Parameters

Liver Function Parameters	Category
ALT	≥ 3× – ≤ 5× ULN > 5× – ≤ 8× ULN > 8× – ≤ 10× ULN > 10× – ≤ 20× ULN > 20× ULN

Liver Function Parameters	Category
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]
ULN: upper limit of normal range. [a] Total Bilirubin ≥ 2 × ULN is defined as at least one case of post-dose TBL ≥ 2 × ULN occurred at the same day or after the first incidence date of ALT or AST ≥ 3 × ULN post treatment.	

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

3.7.4 Other Safety Evaluations

3.7.4.1 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.

3.7.4.2 Electrocardiogram

Electrocardiogram (ECG) parameters will be assessed at baseline as well as throughout the study. ECG parameters (PR, QRS, QT and QTc Fridericia) will be summarized using descriptive statistic, the value of the parameters and the changes from baseline. The corrected QT interval will derived using Friderica formula:

$$QTcF = \frac{\text{QT interval (msec)}}{(60/\text{Ventricular Rate(beats/min)})^{(1/3)}}$$

The number and percentage of subjects having the following notable QTcF interval values while on treatment will be summarized:

- New QTc >450 milliseconds, > 480 milliseconds, and > 500 milliseconds.
- Change from baseline in QTc > 30, >60, and > 90 milliseconds.

For the outlier analysis on the ECG intervals, only the subject with "new" cases (as compared to baseline) will be included for summary. "New" means the category of the QTcF abnormality was not present at baseline and became present for at least one post-baseline ECG assessment. Percentages will be calculated based on the number of subjects who had a baseline and at least one post-baseline assessment.

3.7.4.3 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 visit using descriptive statistics. A shift table of baseline to the "worst" performance post-baseline on treatment period will be presented.

3.7.4.4 Pregnancy and Overdose Reports

Details of current and previous pregnancies will be listed. Any overdoses reported will be listed.

3.7.5 Subgroup Analyses

No subgroup analyses are planned for safety analysis.

4 INTERIM ANALYSIS

CCI

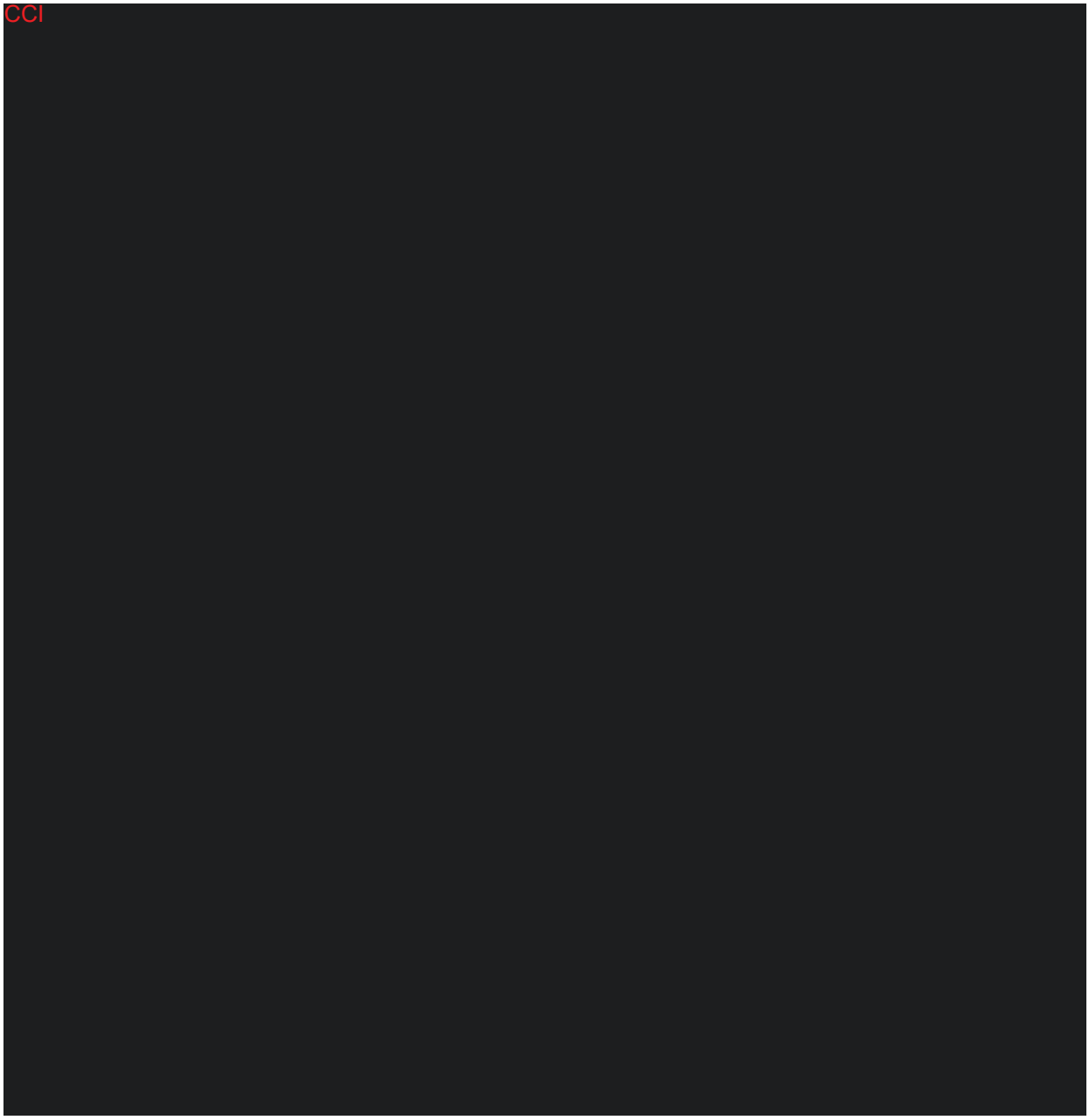
CCI

Figure 1 CCI

5 REFERENCES

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

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.

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

6 VERSION HISTORY

Version	Date	Summary of Changes	Reason for Change
1.0	05Aug2019	Initial document	Initial document
2.0	12Nov2020	General updates and corrections in preparation for futility analysis.	General updates and corrections in preparation for futility analysis.
3.0		Added information about evaluating the impact of the COVID-19 pandemic. • Updated definitions of actual and potential duration of follow-up. • Removed unnecessary guidance text about treatment duration. • Added clarification for Best Overall Response SD>8 weeks definition.	General updates

		• Updates to censoring rules for PFS to align with latest AZ guidance. • Updates to text showing number of days to determine two missed visits. • Updates to definitions for Overall Survival.	
4.0		• Updated version history table with more detail. • Further guidance added for the determination of two missed visits. CCI [redacted] • Requirement added to summaries Part 1 and Part 2 separately	• General updates to provide additional clarity. CCI [redacted] • Part 2 is randomized by laterality of primary tumor but Part 1 is not randomized.

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 (ie, 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 (ie, 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 (ie, 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 (ie, 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 (ie, 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 (ie, 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 (ie, 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 (ie, 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

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.