STATISTICAL ANALYSIS PLAN

Final Version 3.0 dated 10-Aug-2020

An Open-Label, Randomized Phase 3 Study to Evaluate Enfortumab Vedotin vs Chemotherapy in Subjects with Previously Treated Locally Advanced or Metastatic Urothelial Cancer (EV-301)

NCT03474107

ISN: 7465-CL-0301
IND number: 116360

Astellas Pharma Global Development, Inc. (APGD)

Table of Contents

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
ADC	Antibody drug conjugate
ADaM	Analysis data module
AE	adverse event
AESI	adverse event of special interest
Alb	albumin
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANC	absolute neutrophil count
APGD	Astellas Pharma Global Development
AST	aspartate aminotransferase
AT	aminotransferases
ATA	antitherapeutic antibodies
BOR	best overall response
BUN	blood urea nitrogen
C_{max}	maximum concentration
CMQ	customized medical query
CNS	central nervous system
COVID-19	Coronavirus Disease 2019
CPI	checkpoint inhibitor
CR	complete response
CRF	case report form
CSR	Clinical study report
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
CYP3A4	cytochrome P450 3A4
DCR	disease control rate
DILI	drug-induced liver injury
DLT	dose-limiting toxicity
DOR	duration of response
ECG	electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	electronic case report form
EDC	electronic data capture
EORTC	European Organisation for Research and Treatment of Cancer
EOT	end of treatment
ePRO	electronic patient reported outcome
EQ-5D-5L	EuroQOL 5-dimensions
EV	Enfortumab Vedotin
FAS	full analysis set

Abbreviations	Description of abbreviations
FDA	Food and Drug Administration
GCP	good clinical practice
Hct	hematocrit
Hgb	hemoglobin
HRU	healthcare resource utilization
IAP	interim analysis plan
ICF	informed consent form
ICH	International Council for Harmonisation
IDMC	Independent data monitoring committee
IND	investigational new drug
IRR	infusion-related reaction
IRT	interactive response technology
ISN	international study number
MMAE	monomethyl auristatin E
mUC	metastatic urothelial cancer
NCI-CTCAE	National Cancer Institute's Common Terminology Criteria for Adverse Events
NCI-ODWG	National Cancer Institute's Organ Dysfunction Working Group
ORR	overall response rate
OS	overall survival
PD	progressive disease
PDAS	pharmacodynamics analysis set
PD-1	programmed cell death protein 1
PD-L1	programmed death-ligand 1
PFS	progression free survival
PFS1	progression free survival on study therapy
PFS2	progression free survival on subsequent therapy
PKAS	pharmacokinetic analysis set
PPS	per protocol set
PR	partial response
PRO	patient reported outcome
QLQ-C30	EORTC Quality of Life Questionnaire
QOL	quality of life
RBC	red blood cell
RECIST	Response Evaluation Criteria in Solid Tumors
RES	response evaluable set
ROW	rest of the world
SAE	serious adverse event
SAF	safety analysis set
SAP	statistical analysis plan
SMQ	standard MedDRA query
SOC	system organ class
SOD	sum of diameter

Abbreviations	Description of abbreviations
SOP	standard operating Procedure
SSQ	sponsor specific query
SDTM	Study Data Tabulation Model
TBL	total bilirubin
TEAE	treatment emergent adverse events
TMF	trial master file
TP	total protein
ULN	upper limit of normal
US	United States
WBC	white blood cell

List of Key Terms

Terms	Definition of terms
Baseline	Assessments of subjects as they enter a trial before they receive any treatment or before randomization if they didn't receive any study drug.
Endpoint	Variable that pertains to the efficacy or safety evaluations of a trial.
Enroll	To register or enter a subject into a clinical trial. NOTE: Once a subject has received the study drug or placebo, the clinical trial protocol applies to the subject.
Investigational period	Period of time where major interests of protocol objectives are observed, and where the test drug or comparative drug (sometimes without randomization) is usually given to a subject, and continues until the last assessment after completing administration of the test drug or comparative drug.
Randomization	The process of assigning trial subjects to treatment or control groups using an element of chance to determine assignments in order to reduce bias.
Screening	A process of active consideration of potential subjects for enrollment in a trial.
Screen failure	Potential subject who did not meet 1 or more criteria required for participation in a trial.
Screening period	Period of time before entering the investigational period, usually from the time when a subject signs the consent until just before the test drug or comparative drug (sometimes without randomization) is given to a subject.
Study period	Period of time from the first site initiation date to the last site completing the study.
Variable	Any entity that varies; any attribute, phenomenon or event that can have different qualitative or quantitative values.

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes procedures for executing the statistical analysis to fulfil the objectives of the study.

The final SAP will be approved prior to any of the following: study unblinding, database hard lock, interim analysis, or accumulation of substantial amount of data in an open-label study to ensure lack of bias. Critical changes that affected the statistical analyses from the analyses planned in the SAP will be documented in the Clinical Study Report (CSR).

2 STUDY OBJECTIVE(S) AND DESIGN

2.1 Study Objective(s)

Primary:

To compare the overall survival (OS) of subjects with locally advanced or metastatic urothelial cancer treated with EV to the OS of patients treated with chemotherapy.

Secondary:

- To compare progression free survival on study therapy (PFS1) per Response Evaluation Criteria in Solid Tumors (RECIST) V1.1 of subjects treated with EV to patients treated with chemotherapy
- To compare the overall response rate (ORR) per RECIST V1.1 of EV to chemotherapy
- To evaluate the duration of response (DOR) per RECIST V1.1 of EV and chemotherapy
- To compare the disease control rate (DCR) per RECIST V1.1 of EV to chemotherapy
- To assess the safety and tolerability of EV
- To assess quality of life (QOL) and Patient Reported Outcomes (PRO) parameters

2.2 Study Design

This is a global, open-label, randomized Phase 3 study in adult subjects with locally advanced or metastatic urothelial cancer who have received a platinum-containing chemotherapy and have experienced disease progression or relapse during or following treatment with an immune checkpoint inhibitor. Subjects who discontinued CPI treatment due to toxicity are eligible provided that they have evidence of disease progression following discontinuation. Approximately 600 subjects will be randomized to EV (Arm A) or chemotherapy (Arm B) in a 1:1 ratio. Subjects will be stratified according to the following: Eastern Cooperative Oncology Group Performance Status (ECOG PS), regions of the world, and liver metastasis.

Subjects in Arm A will receive EV on Days 1, 8 and 15 of each 28-day cycle. Subjects in arm B will receive either docetaxel, paclitaxel or vinflunine (vinflunine is a choice of comparator only in countries where it is approved for urothelial cancer) as decided by the investigator prior to randomization on Day 1 of every 21-day cycle. Within the control arm, the overall proportion of subjects receiving vinflunine will be capped at approximately 35%. Subjects will continue to receive study treatment until radiological disease progression as determined

per investigator assessment or other discontinuation criteria are met or upon study termination, or study completion, whichever occurs first. No on-study crossover will be allowed. Subjects assigned to the chemotherapy arm will not be allowed to switch to a different chemotherapy treatment during study treatment period. This study will consist of three phases: screening, treatment and follow-up. Details included in protocol flow chart and schedule of assessments.

Subjects will start with cycle 1 and continue on to subsequent 28-day or 21-day cycles until one of the discontinuation criteria are met or upon study termination, or study completion, whichever occurs first. A treatment cycle is defined as 28 days for Arm A and 21 days for Arm B.

Subjects will be evaluated for response according to RECIST V1.1. Imaging for both arms will be performed at baseline and every 56 days (± 7 days) from the first dose of study treatment throughout the study until PFS1 is documented by radiological disease progression or the subject is lost to follow-up, withdraws study consent or starts a subsequent anti-cancer therapy. Baseline imaging performed prior to informed consent as standard of care may be used so long as it is performed within 28 days prior to randomization.

Following discontinuation from study drug, subjects will have a follow-up visit 30 days (+ 7 days) after their last dose of drug for safety assessments. If a subject discontinues study drug prior to radiographic disease progression (i.e., PFS1), the subject should enter the post-treatment follow-up period and continue to undergo imaging assessments every 56 days (± 7 days) until PFS1 is documented.

Following PFS1, subjects will enter the long-term follow-up period and be followed per institutional guidelines but not less than every 3 months from the date of the follow-up visit for survival status and progression status on subsequent therapy (i.e., PFS2).

Subjects will be followed until PFS2 is documented or the subject starts another anticancer treatment, whichever occurs earlier. All subsequent anticancer therapy including date and site of progression for PFS2 will be recorded on the case report form.

Following PFS2, subjects will enter the survival follow-up period and be followed every 3 months for survival status until death, lost to follow-up, withdrawal of study consent, or study termination by sponsor. This study is expected to end once final survival analysis is complete. Subjects will be eligible to continue receiving treatment in this study until they meet a discontinuation criterion as outlined in Section 6 of the study protocol or upon study termination, or study completion, whichever occurs first. QOL assessments and PRO will be collected at protocol-specified time points from all randomized subjects. The following validated tools will be used: European Organisation for Research and Treatment of Cancer (EORTC), Quality of Life Questionnaire (QLQ-C30) and EuroQOL 5-dimension 5-Level Questionnaire (EQ-5D-5L). Healthcare resources utilization (HRU) information will be collected at protocol-specified time points with particular focus on the number of subjects who have an unplanned use of healthcare resources related to clinical or AEs from subjects assigned to treatment arms A and B.

Blood samples for pharmacokinetics and ATA will be collected throughout the study for subjects randomized into Arm A. Validated assays will be used to measure the concentrations of EV ADC and monomethyl auristatin E (MMAE) in serum or plasma and to assess ATA. Pharmacokinetic samples will not be collected from subjects randomized into Arm B. Samples for exploratory biomarkers will be collected at protocol-specified timepoints. Biomarker assessments will not be used for subject selection.

An Independent Data Monitoring Center (IDMC) will be chartered to oversee safety and the planned interim efficacy analysis, which will occur after at least 285 OS events (about 65% of the total planned events) are observed. The IDMC may recommend to the sponsor whether the trial should be terminated, modified or continue unchanged based on ongoing reviews of safety data and interim efficacy analysis. Further details will be outlined in the IDMC charter and interim analysis plan (IAP).

2.3 Randomization

Subjects randomization will occur centrally through Interactive Response Technology (IRT). There are two treatment arms. Subjects will be assigned randomly in a 1:1 ratio to Arm A (EV) OR Arm B (investigator's choice of paclitaxel, docetaxel or vinflunine).

Investigators must select one treatment among the Arm B options before randomization occurs, to use in the event that the subject is randomized to the Arm B. Within the control arm (Arm B), the overall proportion of subjects receiving vinflunine will be capped at approximately 35%.

Randomization will be stratified according to the following factors:

(1) ECOG PS: 0 vs. 1

(2) Region of the world: Western EU vs. US vs. Rest of World

(3) Liver metastasis status: Yes vs. No

3 SAMPLE SIZE

The study design is a group sequential design with two planned analyses: one interim analysis and one final (primary) analysis. The interim and final analyses will be performed after the pre-specified number of death events. The family-wise type I error rate for this study is strongly controlled at 0.025 (one-sided) per multiplicity adjustment rule specified in Section 6.4. OS will be tested at 1-sided 0.00541 significance level for efficacy according to the O'Brien-Fleming boundary as implemented by Lan-DeMets alpha spending function [Lan-DeMets, 1983]. Secondary endpoints PFS1, ORR and DCR will be tested sequentially when primary OS is rejected.

Approximately 600 subjects will be randomized in a 1:1 ratio to 2 treatment arms: EV (Arm A) and chemotherapy (Arm B). Randomization will be stratified by ECOG PS (0 vs 1), regions of the world (Western EU, US, or the Rest of World), and liver metastasis (Yes or No). Assuming HR = 0.75 (median OS in Arm A and Arm B are 10.7 months and 8 months, respectively), drop-out rate of 10%, the final analysis at the planned 439 death events and

1 interim analysis at 65% of the total planned events (285 death events), this sample size will provide 85% power to detect a statistically significant difference at overall type I error rate of 1-sided 0.025.

Sample size is determined by primary endpoint OS. The planned sample size will provide more than 90% power to detect statistically significant differences on selected secondary endpoints: PFS1(assuming median PFS1 in Arm A and Arm B are 6 months and 4 months, respectively), ORR and DCR (assuming 15% treatment difference between Arm A and Arm B for both ORR and DCR).

4 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the following analysis sets will be used for the analyses.

The determination of whether subjects are included or excluded from the safety and efficacy analysis sets will be made prior to database hardlock.

4.1 Full Analysis Set

The full analysis set (FAS) consists of all subjects who are randomized. This analysis set is in compliance with the ITT principle that includes all randomized subjects, the FAS is equivalent to the ITT population. The FAS will be the primary analysis set for efficacy analyses except for response related efficacy endpoints. Demographic and baseline characteristics will be summarized for the FAS.

4.2 Safety Analysis Set

The SAF consists of all subjects who received any amount of study drug, and will be used for safety analyses.

4.3 Response Evaluable Set

The response evaluable set (RES) is defined as all subjects in FAS who have measurable disease (per RECIST V1.1) per investigator at baseline. RES will be used for primary efficacy analysis of response related endpoints, e.g., ORR and DCR.

4.4 Pharmacokinetics Analysis Set

Pharmacokinetics Analysis Set (PKAS) includes subjects who received active drug for whom at least one blood sample was collected and assayed for measurement of the enfortumab vedotin (ASG-22CE) serum/plasma concentrations and for whom the time of sampling and the time of dosing on the day of sampling is known. The PKAS is used for all tables and graphical summaries of the PK data.

5 SUMMARY OF ENDPOINTS

5.1 Primary Efficacy Endpoint

The primary endpoint for this study is the overall survival.

5.1.1 Overall Survival (OS)

OS is defined as the time from the date of randomization until the documented date of death from any cause. All events of death on or prior to analysis cutoff date will be included, regardless of whether the event occurred while the subject is still taking study drug or after the subject discontinue study drug. Subjects who are still alive at the time of data cutoff date will be censored at the last known alive date or at the analysis cutoff date, whichever is earlier. All dates on or prior to cutoff date (e.g., lab testing date, drug administration date), which can support subject's survival status will be used to derive the last known alive date. Subjects with death or last known alive date after the analysis cutoff date will be censored at the analysis cutoff date.

OS (in days) is calculated as: (Date of death or censored) – (Date of randomization) +1.

The primary analysis population for OS is FAS.

5.2 Secondary Endpoints

The secondary endpoints include the PFS1, ORR, DOR, DCR, EORTC-QLQ-C30 and EQ-5D-5L.

5.2.1 Progression Free Survival on Study Therapy (PFS1)

For each subject, PFS1 is defined as the time from the date of randomization until the date of documented radiological disease progression per investigator based on RECIST V1.1, or until death due to any cause, whichever occurs first. If a subject has neither progressed nor died, the subject will be censored at the date of last radiological assessment or at the date of randomization if no post-baseline radiological assessment is available. Subjects who receive any further anti-cancer therapy for the disease before radiological progression will be censored at the date of the last radiological assessment before the anticancer therapy started. In addition, subjects who have PD or death after 2 or more missed disease assessments will be censored at the last disease assessment prior to the 2 or more missed disease assessments.

PFS1 (in days) is calculated as:

(Date of death or documented disease progression or censoring) – (Date of randomization) +1.

To apply the cut-off date to PFS1 is to exclude tumor assessments, death and anti-cancer therapy date after cutoff date in the analysis.

Note: Patient cannot be censored at "Not Evaluable (NE)". If NE is the only previous assessment, then PFS will be censored at randomization date.

In order to evaluate the robustness of the PFS1 endpoint, a sensitivity analysis will be performed. The sensitivity analysis is the same as the primary analysis except that when PD or death is documented after more than one missed disease assessment it will not be censored. See below table for PFS1 censoring rules for both primary analysis and sensitivity analysis:

Censoring Rules for Primary and Sensitivity Analysis of PFS1

Situation	Primary Analysis	Sensitivity Analysis
No PD and no death; new anticancer treatment is not initiated	Censored at last disease assessment or randomization date if no post-baseline tumor assessment	Censored at last disease assessment or randomization date if no post-baseline tumor assessment
No PD and no death; new anticancer treatment is initiated	Censored at last disease assessment before new anticancer treatment or randomization date if no post-baseline tumor assessment before new anticancer treatment	Censored at last disease assessment before new anticancer treatment or randomization date if no post-baseline tumor assessment before new anticancer treatment
PD or death documented after ≤ 1 missed disease assessment	Progressed at date of documented PD or death	Progressed at date of documented PD or death
PD or death documented after ≥ 2 missed disease assessments	Censored at last disease assessment prior to the ≥ 2 missed disease assessment	Progressed at date of documented PD or death

The primary analysis population for PFS1 is FAS.

5.2.2 Overall Response Rate (ORR)

Overall response rate (ORR) is defined as the proportion of subjects with best overall response (BOR) as confirmed complete response (CR) or partial response (PR), per RECIST v1.1 as assessed by investigator. Derivation of BOR is specified in Section 5.2.5.1. The primary analysis population for ORR is RES. ORR based on BOR regardless of confirmation will also be calculated.

5.2.3 Duration of Response (DOR)

DOR is defined as the time from the date of the first CR/PR (whichever is first recorded) that is subsequently confirmed as assessed by investigator to the date of documented disease progression or death due to any cause whichever occurs first. If a subject has neither progressed nor died, the subject will be censored at the date of last radiological assessment or at the date of first CR/PR if no subsequent post-baseline radiological assessment is available. Subjects who receive any further anti-cancer therapy for the disease before radiological progression will be censored at the date of the last radiological assessment before the anticancer therapy started. In addition, subjects who have PD or death after 2 or more missed disease assessments will be censored at the last disease assessment prior to the 2 or more missed disease assessments.

DOR (in days) is calculated as:

(Date of documented progression or death or censoring) – (Date of the first response CR/PR) +1.

To apply the cutoff date to DOR is to exclude tumor assessments, death and anti-cancer therapy date after cutoff date in the analysis.

The primary analysis population for DOR is all subjects in RES who achieved BOR of either confirmed CR or confirmed PR.

5.2.4 Disease Control Rate (DCR)

Disease control rate (DCR) is defined as the proportion of subjects with BOR of confirmed CR or confirmed PR or SD, per RECIST v1.1 as assessed by investigator. Derivation of BOR with confirmation is specified in Section 5.2.5.1. The primary analysis population for DCR is RES.

5.2.5 Derivation of BOR and SOD

Best overall response (BOR) and sum of diameters (SOD) will be derived for the purpose of analyzing response related endpoints, e.g., ORR, DOR and DCR.

5.2.5.1 Best Overall Response (BOR)

Best overall response (BOR) is determined once all tumor timepoint response data for the subject is available. Responses recorded after new anticancer therapy or progressive disease (PD), will be excluded from BOR derivation.

BOR regardless of Confirmation

The BOR regardless of confirmation for a subject is defined as the best timepoint response in the order of CR, PR, SD, PD and NE. SD must be documented as present at least once at or after 7 weeks (49 days) post first dose or randomization (for subjects who didn't receive the study drug) for BOR to be SD.

BOR with Confirmation

Confirmation of CR or PR should occur at the next scheduled assessment (not less than 4 weeks following the initial assessment at which CR or PR is observed).

The BOR with confirmation will be derived according to below criteria per RECIST V1.1:

- If a patient has at least two CR and the first and the last CR dates are at least 28 days apart, then the best overall response is defined as confirmed CR
- If a patient has PR and another CR/PR with at least 28 days apart, then the best overall response for this patient is confirmed PR
- For those patients who do not have confirmed CR or PR, if the patient had at least one tumor assessment record of CR/PR/SD which is at least 49 days after date of first dose (date of randomization will be used for subjects who randomized but received no study drug), then best overall response is defined as SD
- For those patients who do not have confirmed CR, confirmed PR or SD defined as above, but they have a tumor assessment as PD, their best overall response is PD
- Otherwise, best overall response is defined as Not Evaluable (NE) or No Data (ND) for subjects without any post-baseline tumor assessment data

5.2.5.2 Sum of Diameters (SOD)

Per RECIST, tumor size is measured by SOD, which is defined as the sum of the diameter of all target lesions at a tumor assessment. Percentage change of SOD at post-baseline assessment visit from SOD at baseline and from the smallest SOD prior to the visit (Nadir) will be used to determine CR, PR, SD and PD in timepoint response assessment.

5.2.6 EORTC QLQ-C30

EORTC QLQ-C30 was developed to assess the quality of life of cancer subjects. It has been translated and validated into 81 languages and used in more than 3,000 studies worldwide. It contains 5 functioning scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, nausea, pain) and additional single symptom items. It is scored on a 4 point scale (1=not at all, 2=a little, 3= quite a bit, 4=very much). The EORTC QLQ-C30 instrument also contains 2 global scales that use 7 point scale scoring with anchors (1=very poor and 7=excellent).

The EORTC QLQ-C30 scale scores will be calculated using the EORTC QLQ-C30 Scoring Manual (Fayers et al, 2001). The instrument yields the following scales:

- Global health status / Quality of Life (QL2) (2 items)
- Functional scales
    - Physical functioning (PF2) with 5 items
    - Role functioning (RF2) with 2 items
    - Emotional functioning (EF) with 4 items
    - Cognitive functioning (CF) with 2 items
    - Social functioning (SF) with 2 items
- Symptom scales / items
    - Fatigue (FA) with 3 items
    - Nausea and vomiting (NV) with 2 items
    - Pain (PA) with 2 items
- Symptom items
    - Dyspnea (DY), insomnia (SL), appetite loss (AP), constipation (CO), diarrhea (DI) with 1 item each
- Financial impact (FI) with 1 item

Physical and role functioning Revised scales, as well as the global QoL, have revised scale names which are those that have been changed since version 1.0, and their short names are indicated a suffix “2” – for example, PF2, according to the instrument’s manual (Fayers et al, 2001).

The principle for scoring is the same for all scales. Briefly, outcome scores are computed by standardizing the average of the items (i.e., a raw score) making up the scale. Outcome scores are computed using a linear transformation of the raw score such that scores range from 0 to 100. A higher score represents a higher (“better”) level of functioning, or a higher (“worse”) level of symptoms, i.e. a high score for a functional scale represents a high/healthy level of functioning, a high score for the GHS represents a high QoL, but a high score for a symptom

scale/item represents a high level of symptomatology/problems. Note that the global health status scale is based on only the 2 specific HRQoL items and not the entire questionnaire.

If at least half the items of a scale are present for a timepoint then the score will be calculated using the average of all items answered; otherwise the score will be set to missing. For single measures, if the item is missing the scale score is set to missing (Fayers et al, 2001).

Change from baseline defined as post-baseline value minus baseline value will be calculated for each assessment. At each post-baseline assessment, the change from baseline will be calculated for each scale/item.

This assessment will be completed at various time points as specified in the study Flow Chart in study protocol, beginning with Baseline (Day -7 to -1) and on Day 1 of each week for the first 12 weeks and then every 12 weeks afterward, EOT visit and at the follow-up visit. Please refer to Protocol Appendix Section 12.5 for a sample questionnaire. The primary analysis population is FAS.

5.2.7 EQ-5D-5L

The EQ-5D-5L is a generic preference-based measure that indirectly measures the utility for health that generates an index-based summary score based upon societal preference weights (Pickard et al 2007). The EQ-5D-5L consists of 6 items that cover 5 main domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and a general visual analog scale for health status. For the 5 main domains, each domain has one question with 5 answer choices indicating the 5 severity levels. This assessment will be completed at various time points as specified in the study Flow Chart in study protocol, beginning with Baseline (Day -7 to -1) and on Day 1 of each week for the first 12 weeks and then every 12 weeks afterward, EOT visit and at the follow-up visit. Please refer to Protocol Appendix Section 12.6 for a sample questionnaire. The primary analysis population is FAS.

5.3 Exploratory Efficacy Endpoints

5.3.1 Biomarkers

Nectin-4 and PD-L1 expression in tissue will be tested using screening/baseline tumor tissue sample. Nectin-4 expression in tissue will be assessed by IHC H-score. PD-L1 expression in tissue will be assessed by IHC combined positive score (CPS).

Subjects may have also been tested for PD-L1 expression prior to enrollment and these PD-L1 test results (i.e., prior PD-L1 testing) were collected when available.

Additional biomarkers (e.g., TCGA subtyping) may be evaluated to explore effects on antitumor activity and safety of the study drug.

5.3.2 Incidence of antitherapeutic antibodies (ATA)

The incidence of ATA formation to the antibody drug conjugate (ASG-22CE).

Subjects with positive ATA post-baseline are defined in two ways:

- Transiently positive: if subjects had at least one post-baseline positive result;

- Persistently positive: if subjects had two or more consecutive samples were confirmed positive.

5.3.3 PFS in the next line of therapy (PFS2)

PFS2 is defined as the time from randomization to subsequent disease progression after initiation of new anti-cancer therapy, or death from any cause, whichever occurs first. If progression after next-line therapy cannot be reliably determined, a PFS2 event is defined as starting of a different treatment or death from any cause, whichever occurs first. Patients alive and for whom a subsequent disease progression has not been observed will be censored at the last time known to be alive and without second disease progression. Among these censoring patients, if there are subjects who didn't receive any new anti-cancer therapy but received the study drug, the subjects will be censored at the last dosing date of the study drug or cutoff date (for subjects who are on study treatment); and if subjects who received neither study drug nor any other anti-cancer therapy, subjects will be censored at the randomization date.

PFS2 (in days) is calculated as:

(Date of death or disease progression on next-line therapy or starting date of another anti-cancer therapy after next-line therapy or censoring) – (Date of randomization) +1.

The analysis population for PFS2 is FAS.

5.3.4 Time to Response

Time to response (TTR) is defined as the time from the date of randomization until the date of the first CR/PR (whichever is first recorded) that is subsequently confirmed as assessed by investigator. TTR is only derived for subjects who achieve confirmed CR or PR.

5.3.5 Healthcare resources utilization (HRU)

The Health Resource Utilization (HRU) questionnaire focuses on unplanned use of healthcare resources related to clinical or adverse events from subjects assigned to treatment arms. Healthcare resources utilization such as incidence of emergency room visits, hospitalizations, and outpatient (either primary care or specialist) office visits are collected via the HRU tool. Date and length of each hospitalization will also be collected.

5.4 Safety Endpoints

Safety endpoints are AEs, laboratory tests, vital signs, ECGs and ECOG performance status. Safety will be assessed by evaluation of the following variables:

- Treatment-emergent adverse events (TEAEs; frequency, severity, seriousness, and relationship to study drug).
  - TEAE is defined as an adverse event observed after starting administration of the study drug and within 30 days after taking the last dose of study drug. If the adverse event occurs on the first dosing date and the onset check box is marked "Onset after first dose of study drug" or the onset check box is left blank, then the adverse event will be considered treatment emergent. If the adverse event occurs on the first dosing

date and the onset check box is marked "Onset before first dose of study drug", then the adverse event will not be considered treatment emergent. If a subject experiences an event both during the pre investigational period and during the investigational period, the event will be considered as TEAE only if it has worsened in severity (i.e., it is reported with a new start date). If a complete onset date is unknown, and the onset check box is marked "Onset after first study drug taken", then the adverse events will be considered treatment emergent. If the onset check box is marked "Onset before study drug taken", then the adverse event will not be considered treatment emergent. If onset check box is left blank, imputed onset date as specified in Section 6.9.1 will be used to determine whether an adverse event is treatment emergent.

- A drug-related TEAE is defined as any TEAE with relationship "yes" to study treatment as assessed by the investigator or with missing assessment of the causal relationship.

- Serious adverse events (SAEs) include adverse events that are flagged as serious by the investigator on eCRF or the SAE flag by the investigator on CRF is missing, or upgraded by the Sponsor based on review of the Sponsor's list of Always Serious term.
- Clinical laboratory variables
  Below is a table of the laboratory tests that will be performed during the conduct of the study. Refer to the Schedule of Assessments in the protocol for study visit collection dates.

Laboratory Assessments

Panel/Assessment	Parameters to be analyzed	NCI-CTCAE Grading (Hyper/Hypo)
Biochemistry	Sodium (Na)	Both
	Magnesium(Mg)	Both
	Potassium (K)	Both
	Calcium (Ca)	Both
	Chloride (Cl)	N/A
	Phosphate (P)	Hypo
	Serum Creatinine	Hyper
	Glucose (Gl)	Both
	Blood Urea Nitrogen (BUN)	N/A
	Alkaline Phosphatase (ALP)	Hyper
	Aspartate Aminotransferase (AST)	Hyper
	Alanine Aminotransferase (ALT)	Hyper
	Lactate Dehydrogenase (LDH)	N/A
	Bilirubin Total (TBL) (total and direct)	Hyper
	Total Protein (TP)	N/A
	Albumin (Alb)	Hypo
	Bicarbonate (HCO3)	Hypo
	Amylase	Hyper
	Lipase	Hyper
	Uric Acid	Hyper
	Hemoglobin A1c (screening only)	N/A
	Serum HCG for female subjects of childbearing potential	N/A

Table continued on next page

Panel/Assessment	Parameters to be analyzed	NCI-CTCAE Grading (Hyper/Hypo)
Hematology	Red Blood Cell Count (RBC)	N/A
	Hematocrit (Hct)	N/A
	Hemoglobin (Hgb)	Both
	Platelets	Hypo
	White blood cell count (WBC)/differential	Both
	• Absolute Neutrophils	Hypo
	• Eosinophils	N/A
	• Basophils	N/A
	• Lymphocytes	Both
	• Monocytes	N/A
Urinalysis	Pregnancy	N/A

- Vital signs (systolic and diastolic blood pressures (mmHg), radial pulse rate (beats/minute) and body temperature (C)) and weight
- 12-lead electrocardiogram (ECG)
- ECOG performance scores

5.5 Other Endpoints

5.5.1 Drug exposure

- Duration of exposure (day)

  Duration of exposure = min (Last date of exposure, death, cutoff date) – First dose date + 1

  For EV subject, last date of exposure = (initial dose date of the last cycle + 28– 1)

  For Chemo subject, last date of exposure = (initial dose date of the last cycle + 21– 1)

- Number of cycles

  Total number of cycles with non-zero dosing.

- Cumulative dose

  Sum of (Total Dose Administrated) across all days

- Planned dose intensity

  Initial dose of the drug multiplied by planned number of dosing days per cycle.

  For example, for EV, the planned number of dosing days per cycle is 3 and the planned dose intensity is $1.25 \times 3 = 3.75$ mg/kg per cycle. For vinflunine, the planned number of dosing days per cycle is 1, if the subject is given initial dose as 320 mg/m^2, the planned dose intensity is 320 mg/m^2 per cycle.

- Dose intensity

  For EV:

$$Dose\ intensity = \frac{\sum_{i=1}^{NC}\sum_{j=1}^{3}(TD_{i,j}/W_{i,j})}{(last\ date\ of\ exposure - first\ dose\ date + 1)/28}$$

  Where $TD_{i,j}$ is the actual total drug administrated at cycle i day j, $W_{i,j}$ is the body weight of the subject at cycle i day j, and NC is the total number of cycles with non-zero dosing. If the body weight is greater than 100kg, then $W_{i,j} = 100kg$. If there is no dose administrated in a planned dosing day in a cycle, $TD_{i,j} = 0mg$.

For Chemotherapy (docetaxel, vinflunine and paclitaxel):

$$Dose\ intensity = \frac{\sum_{i=1}^{NC}(TD_i/BSA_i)}{(last\ date\ of\ exposure - first\ dose\ date + 1)/21}$$

Where TD_i is the actual total drug administrated at cycle i, BSA_i is the body surface area for the subject at cycle i, and NC is the total number of cycles with non-zero dosing.

- Relative dose intensity (%)

$$\frac{dose\ intensity}{Planned\ dose\ intensity} \times 100$$

5.5.2 Previous and concomitant medication

Previous medication is defined as medication with at least one dose taken before the date of the first dose of study drug.

Concomitant medication is defined as medication with at least one dose taken between the date of first dose (inclusive) and the date of last dose plus 30 days (inclusive) of study drug.

5.5.3 Previous and concomitant non-medication therapy

Previous non-medication therapy is defined as non-medication therapy administered at least once before the date of the first dose of study drug.

Concomitant non-medication is defined as non-medication therapy administered at least once between the date of first dose (inclusive) and the date of last dose (inclusive) of study drug.

5.5.4 Subsequent Anti-Cancer Therapy

Subsequent Anti-Cancer therapies include all systemic therapies and radiation, palliative radiation and other therapies which irradiated or affected either target lesion or non-target lesions.

6 STATISTICAL METHODOLOGY

6.1 General Considerations

For continuous variables, descriptive statistics will include the number of subjects (n), mean, standard deviation, median, minimum and maximum. When needed, the use of other percentiles (e.g., 10%, 25%, 75% and 90%) will be mentioned in the relevant section. Frequencies and percentages will be displayed for categorical data. Percentages by categories will be based on the number of subjects with no missing data, i.e., will add up to 100%. Kaplan-Meier survival curves will be displayed for time-to-event variables and median survival time will be estimated with 2-sided 95% confidence interval (CI).

Summaries based on FAS and RES (e.g., disposition, baseline and efficacy data) will be presented by randomized treatment per IRT, unless specifically stated otherwise. Safety analysis and other summaries based on SAF or PKAS will be presented by actual treatment received.

All nominal comparisons will be made using two-sided test at the α=0.05 significance level. Multiplicity adjustment for conducting formal statistical testing on efficacy endpoints is specified in Section 6.4.

All data processing, summarization, and analyses will be performed using SAS® Version 9.2 or higher on Unix. Specifications for table, figure, and data listing formats can be found in the TLF specifications for this study.

Study day will be calculated in reference to the first dose date. For assessments conducted before the first dose, study day will be calculated as (assessment date – first dose date). For assessments conducted on or after the first dose, study day will be calculated as (assessment date – first dose date + 1).

Baseline value is defined as the last non-missing measurement taken prior to the first dose of the study drug unless specified otherwise. For subjects who were randomized but received no dose, baseline value is defined as the last non-missing measurement taken prior to the randomization date.

For the definition of subgroups of interest please refer to Section 6.7.

6.2 Study Population

All summary statistics specified in this section will be presented by treatment groups and overall on FAS and other analysis populations if necessary, unless specifically stated otherwise.

6.2.1 Disposition of Subjects

The following subject data will be presented:

- Number and percentage of subjects with informed consent, discontinued before randomization, randomized (overall only);
- Number and percentage of subjects included in each analysis set, by treatment group and overall;
- Number and percentage of subjects discontinued treatment, by primary reason for treatment discontinuation, by treatment group for all randomized subjects;
- Number and percentage of subjects completed or discontinued the 30-day follow-up visit, by primary reason for 30-day follow-up discontinuation, by treatment group for all randomized subjects;
- Number and percentage of subjects discontinued post-treatment period, by primary reason for post-treatment discontinuation, by treatment group for all randomized subjects;
- Number and percentage of subjects discontinued long term follow-up period, by primary reason for long term follow-up discontinuation, by treatment group for all randomized subjects;
- Number and percentage of subjects discontinued survival follow-up period, by primary reason for survival follow-up discontinuation, by treatment group for all randomized subjects;

- Number and percentage of screen failure subjects, by primary reason for screen failure for screen failure subjects only;

6.2.2 Major Protocol Deviations

Major protocol deviations as defined in the study protocol (Section 8.1.6 Major Protocol Deviations) will be assessed for all subjects randomized. The number and percentage of subjects meeting any criteria will be summarized for each criterion and overall, by treatment group and overall as well as by study site. Subjects deviating from a criterion more than once will be counted once for the corresponding criterion. Any subjects who have more than one major protocol deviation will be counted once in the overall summary. A data listing will be provided by site and subject. The major protocol deviation criteria will be uniquely identified in the summary table and listing.

The unique identifiers will be as follows:

PD1 - Entered into the study even though they did not satisfy entry criteria,

PD2 - Developed withdrawal criteria during the study and was not withdrawn,

PD3 - Received wrong treatment or incorrect dose,

PD4 - Received excluded concomitant treatment.

6.2.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized by descriptive statistics. Number and percentage of subjects enrolled in each country and site will also be presented.

Demographic variables:

- Age, Weight, Height, BMI and BSA
- Sex, Ethnicity, and Race
- Age categories (<65, ≥ 65 to <75, ≥ 75 years)
- EudraCT age categories (≥ 18 to ≤ 64, ≥ 65 to ≤ 84, ≥ 85 years)
- Weight categories (<=100, >100 kg)
- BMI categories (<18.5, ≥ 18.5 to <25, ≥ 25 to <30, ≥ 30 kg/m^2)

Baseline variables:

- Stratification factors used at randomization
  - ECOG PS
  - region
  - liver metastasis
- Nectin-4 IHC H-score (including 25th percentile and 75th percentile)
- PD-L1 IHC combined positive score (<10, >=10)
- Prior PD-L1 testing
- Hemoglobin
- HbA1c (<5.7%, >=5.7% to <6.5%, >=6.5%)
- Albumin
- Hemoglobin categories (<10 g/dL, ≥10 g/dL)

- Albumin categories (<LLN, ≥LLN)
- Bellmunt risk score (0-1, ≥2) (Bellmunt 2010)
- Renal function group based on estimated creatinine clearance (by Cockcroft-Gault formula)
  Normal: ≥90 mL/min
  Mild: ≥60 and <90 mL/min
  Moderate: ≥30 and <60 mL/min
  Severe: ≥15 and < 30 mL/min
- History of diabetes/hyperglycemia defined as any hyperglycemia SSQ/CMQ
- Hepatic dysfunction group defined per NCI-ODWG criteria below
  Normal: Total bilirubin ≤ULN and AST ≤ULN
  Mild: (Total bilirubin >1 - 1.5 x ULN) or (total bilirubin <= ULN and AST >ULN)
  Moderate: Total bilirubin >1.5 - 3 x ULN
  Severe: Total bilirubin >3 x ULN
- Prior tobacco history (never used, former user, current user, unknown)
- Urothelial cancer disease history
  - Primary disease site of origin
  - Current extent of disease
  - Histology type at initial diagnosis
  - Visceral metastasis
  - Lymph node only metastasis
  - CNS metastases
  - Time since metastatic disease or time since locally advanced disease: defined as time from metastatic disease or locally advance disease to the date of randomization

Medical history other than mUC and conditions existing at baseline will be coded in MedDRA and summarized by System Organ Class (SOC) and Preferred Term (PT) as well as by PT alone, by treatment group and overall for the FAS and other analysis populations if necessary. Baseline conditions are defined as those ongoing at the time of informed consent and before the first dose of study drug. For ongoing medical conditions, Common Terminology Criteria for Adverse Events (CTCAE) grade will be provided in listing.

6.2.4 Previous and Concomitant Medications

Previous medications are coded with WHO-DD, and will be summarized by therapeutic subgroup (ATC 2nd level), chemical subgroup (ATC 4th level), preferred WHO name for the SAF.

As with previous medication, concomitant medication will be summarized for each treatment group by therapeutic subgroup (ATC 2nd level), chemical subgroup (ATC 4th level), preferred WHO name for the SAF. Subjects taking the same medication multiple times will be counted once per medication and investigational period. All previous and concomitant medications will be coded by indication-specific ATC.

6.2.5 Prior Systemic Anti-Cancer Therapy

Prior systemic anti-cancer therapy will be summarized by descriptive statistics. Number and percentage of subjects by below categories will be presented (continuous variables below will also be summarized by mean, standard deviation, median, min and max):

- Number of prior lines of systemic therapy (1, 2 and ≥3 lines)
- Types of prior CPI received (e.g., Pembrolizumab, Atezolizumab): In addition, number of subjects received prior PD-L1 inhibitor only, PD-1 inhibitor only and both will be derived and displayed
- Types of prior platinum-based treatment received (cisplatin, carbo-platin and both)
- CPI as the most recent therapy (yes vs. no)
- Best response to prior CPI

6.2.6 Prior Procedures for Primary Cancer

Number and percentage of subjects who took any prior surgery or procedure for the treatment of the primary cancer will be presented. Number and percentages of subjects under different types of the procedures will be tabulated.

6.2.7 Prior Radiation Therapy

Frequency tabulations of subjects with prior radiation therapy will be presented.

6.2.8 Previous and Concomitant Non-Medication Therapy

Subjects with previous and concomitant non-medication therapy and its reason for use will be presented in the listing.

6.2.9 Cytology Procedures

Subjects who had fluid tap will be presented. The location of the cavity, lateral location, name of the procedure, and cytology results will also be presented in the listing.

6.2.10 Subsequent Anti-Cancer Therapy

Number and percentage of subjects who took at least one subsequent anti-cancer therapy will be presented. The best overall response to the first subsequent anti-cancer therapy will be also tabulated.

6.3 Study Drug Exposure

The following information on drug exposure will be presented by treatment group for the SAF:

- Descriptive statistics for duration of exposure, number of cycles, cumulative dose, planned dose intensity, dose intensity and relative dose intensity (RDI)
- Number and percent of subjects with dose adjustment and their reasons
- Number of cycles will be categorized per below categories. Counts and percentages of subjects in each of these categories will be summarized.
    - less than or equal to 1 cycle
    - at least 2 cycle, less than 4 cycles

  - at least 4 cycles, less than 6 cycles
  - 6 cycles or more
  - Unknown
- RDI will be categorized according to the following categories. Counts and percentages of subjects in each of these categories will also be summarized.
  - less than 50%
  - at least 50%, less or equal to 80%
  - greater than 80%
  - Unknown

6.4 Analysis of Efficacy

Efficacy analyses will be conducted on the FAS and RES. The interpretation of results from statistical tests will be based on the FAS or RES. All randomized subjects will be analyzed according to the treatment to which they are randomized.

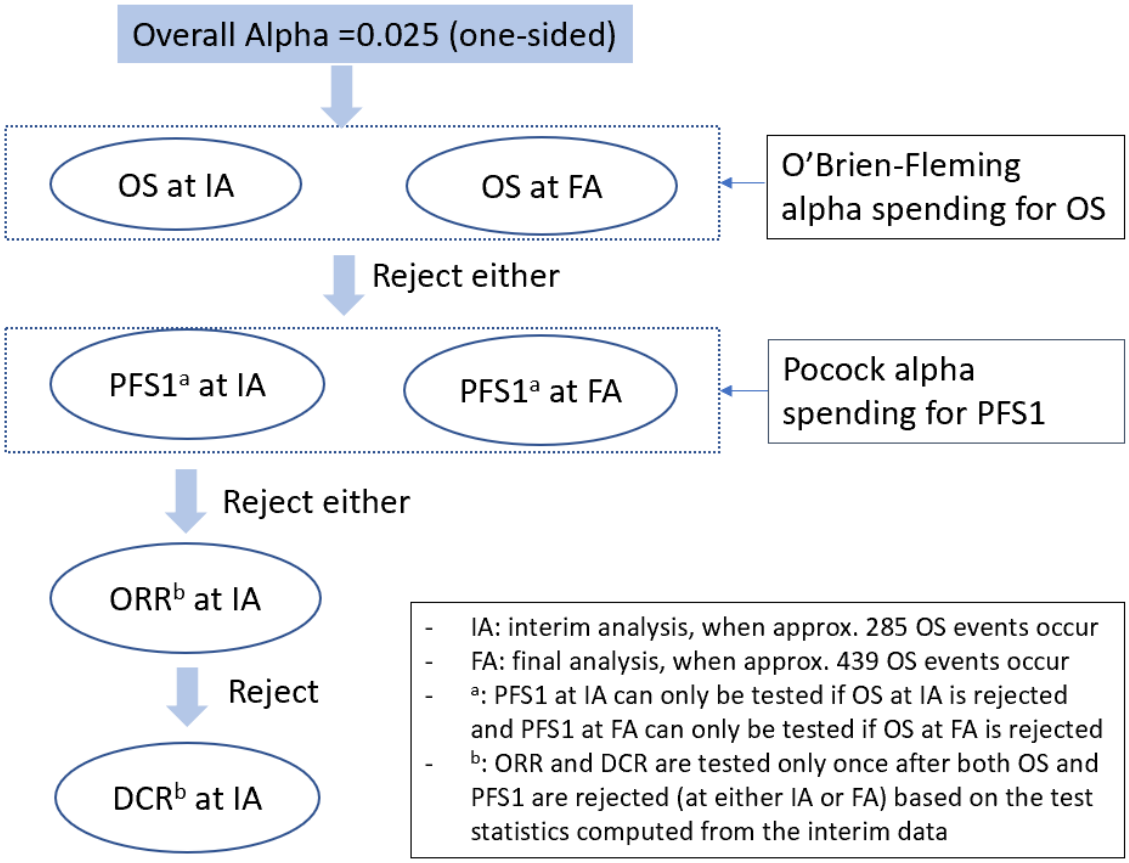
Multiplicity Adjustment

Multiplicity adjustment was incorporated within the interim and final (primary) analysis of the primary endpoint, and between the primary and the selected secondary endpoints.

The family-wise type I error rate for this study is strongly controlled at 2.5% (one -sided) that allows the study to declare positive on primary endpoint OS on the FAS population. OS will be formally tested at both the interim analysis and final analysis according to the O'Brien-Fleming boundary per Lan and DeMets method [Lan-DeMets 1983] via EAST 6.5.

Formal hypothesis testings on the selected secondary endpoints including PFS1, ORR and DCR, will be performed hierarchically (per the order of PFS1-> ORR->DCR) only when the OS testing result is rejected. PFS1 is planned to test at either the interim analysis or final analysis when OS is rejected. The significance level of PFS1 at the interim and final analysis will be based on Pocock boundary per Lan and DeMets method [Lan-DeMets 1983] via EAST 6.5. ORR and DCR will be tested only once after both OS and PFS1 are rejected (at either IA or FA) and the test statistics will be computed from the interim data. The significance level of both ORR and DCR are 0.025. DCR will be tested after ORR is rejected. Below diagram summarized the multiplicity procedure for all formal hypothesis testings.

Multiple Testing Procedure for All Formal Hypothesis Testings



The details about the significance levels at interim analysis and final analysis for each efficacy endpoints (OS, PFS1, ORR and DCR) is specified in below table.

Summary of Timing, Sample Size and Decision Guidance at the Planned Analyses

Analysis	Criteria for conducting analysis (projected timing)	Endpoint/Analysis Set	Efficacy Boundary	
			P-value(1-sided) at the boundary	Approx. observed HR/Z value at the boundary
Interim Analysis	Approx. 285 OS events are observed	OS/FAS	0.00541	HR=0.74
		PFS1/FAS	0.02072	HR=0.82
		ORR* /RES	0.025	Z=1.96
		DCR*/RES	0.025	Z=1.96
Final (Primary) Analysis	Approx. 439 OS events are observed	OS/FAS	0.02332	HR=0.83
		PFS1/FAS	0.01196	HR=0.82

**: ORR and DCR are tested only once after both OS and PFS1 are rejected (at either IA or FA) and the test statistics will be computed from the interim data.*

In above table, P-value for efficacy boundaries are based on 65% and 75% information fraction for OS and PFS1, respectively. The total PFS1 event at the final analysis was estimated approximately 530 based on the assumption specified in Section 3. At the timing of IA (i.e. when 285 OS events are observed), the total PFS1 event is estimated as about

398 (i.e., 75% of the number of total PFS1 events). Boundaries for OS and PFS1 will be updated based on actual observed information fractions at the interim. Since the study enrollment was completed in Jan 2020, based on the definition of RES, ORR and DCR analysis are ready after enrollment completion. Per pre-specified multiplicity procedure, ORR will be tested once on the interim data (i.e. data per interim analysis cutoff) and it will be only tested when both OS and PFS1 are rejected either at interim or final analysis (note: ORR analysis will apply the interim data cutoff date when it is performed at the final). ORR test p-value will be compared to the significance level 0.025. Testing on DCR is similar to the testing on ORR, but DCR will be tested only when ORR is rejected.

6.4.1 Analysis of Primary Endpoint

Overall Survival (OS) is the primary efficacy endpoint. The distribution of OS will be estimated for each treatment arm using Kaplan-Meier methodology and the primary analysis on comparing Arm A and Arm B will be conducted using the log-rank test stratified by ECOG PS (0 vs. 1), region (US, EU and the Rest of world) and liver metastasis status (Yes vs. No) per IRT. In addition, stratified Cox proportional hazard model (same stratification factors as used for stratified log-rank test) will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

Comparison of Arm A (EV) and Arm B (Chemotherapies) will be tested at a planned interim analysis and the final (primary) analysis.

The null hypothesis is:

- H_{01}: OS of Arm A is not better than the comparator Arm B

The accompanying alternative hypothesis is:

- H_{11}: OS of Arm A is better than the comparator Arm B

OS primary analysis will be evaluated on the FAS. The final (primary) analysis will be performed when approximately 439 OS events have been observed. One planned interim analysis will be performed when approximately 285 OS (about 65% of the total OS events) events have been observed. The efficacy boundary for interim analysis and final analysis will be determined according to the O'Brien-Fleming boundary as implemented by Lan-DeMets alpha spending function [Lan-DeMets 1983]. Details about the interim analysis will be included in interim analysis plan (IAP).

Additional sensitivity analysis for OS will include:

- OS analysis based on unstratified log-rank test
- Imbalance in the use of subsequent anti-cancer therapy (including subsequent use of EV) may potential bias the inference in OS. The following methods may be used to assess the impact of the use of subsequent anti-cancer therapy on OS:
    - Rank-preserving structural failure time (RPSFT) method [Robins and Tsiatis, 1991]. This method is to assess the impact of Arm B patients who take EV as a subsequent therapy by reconstructing the survival duration of subjects, as if they had never received EV.

  - Inverse probability of censoring weights (IPCW) method [Robins and Finkelstein, 2000]. Subjects who took subsequent therapy will be censored at the time of sequent anti-cancer therapy, but subjects are weighted according to their probability to take subsequent therapy.
- To assess potential COVID-19 impact, below sensitivity analyses maybe conducted:
  - Same as OS primary analysis, except that subjects died due to COVID-19 infection will be censored at the death date.
  - Same as OS primary analysis, but exclude the subjects died due to COVID-19 infection from the analysis.

6.4.2 Analysis of Secondary Efficacy Endpoints

Primary analysis of PFS1, ORR, DOR, and DCR will be based on RECIST v1.1 assessed by investigator.

6.4.2.1 PFS1

Progression-free survival as assessed by the investigator (PFS1) on the FAS is the secondary endpoint. In order to compare the PFS1 between Arm A (EV) and Arm B (the comparator Chemo), the null hypothesis is constructed:

H_{02}: PFS1 of Arm A is not better than the Arm B

The accompanying alternative hypothesis is:

H_{12}: PFS1 of Arm A is better than the Arm B

Formal statistical comparison of Arm A and Arm B will be performed per planned multiplicity procedure, see multiplicity adjustment section.

The distribution of PFS1 will be estimated for each treatment arm using Kaplan-Meier methodology and compared between Arm A and Arm B using log-rank test stratified by ECOG PS (0 and 1), region (US, EU and the Rest of world) and liver metastasis status (Yes vs. No) per IRT. In addition, stratified Cox proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

Additional sensitivity analysis for PFS1 will include:

- PFS1 analysis based on unstratified log-rank test
- Sensitivity analysis on FAS when removing the censoring of PD or death events occurred after missing 2 consecutive tumor assessments as specified in Section 5.2.1
- To access potential COVID-19 impact, below sensitivity analyses maybe conducted:
  - Same as PFS1 primary analysis, except that subjects died due to COVID-19 infection without PD will be censored at the death date.
  - Same as PFS1 primary analysis, but exclude the subjects died due to COVID-19 infection from the analysis.

6.4.2.2 ORR

The comparison of ORR between Arm A and Arm B will be performed using stratified CMH test. The primary analysis will be performed on RES. In addition, ORR for each arm will be

estimated and corresponding 95% confidence interval will be constructed. The formal statistical comparison of ORR between Arm A and Arm B will be conducted only per planned multiplicity procedure, see multiplicity adjustment section. The null hypothesis of ORR is:

H_{03}: ORR of Arm A is not better than the Arm B

The accompanying alternative hypothesis is:

H_{13}: ORR of Arm A is better than the Arm B

Additional sensitivity analysis for ORR will include the comparison of ORR regardless of confirmation.

6.4.2.3 DOR

The distribution of DOR will be estimated for subjects in RES and achieved confirmed CR or PR by each treatment arm using Kaplan Meier method.

6.4.2.4 DCR

The comparison of DCR between Arm A and Arm B will be performed using stratified CMH test. In addition, DCR for each arm will be estimated and corresponding 95% confidence interval will be constructed. The formal statistical comparison of Arm A and Arm B will be conducted only per planned multiplicity procedure, see multiplicity adjustment section. The null hypothesis of DCR is:

H_{04}: DCR of Arm A is not better than the Arm B

The accompanying alternative hypothesis is:

H_{14}: DCR of Arm A is better than the Arm B

Additional sensitivity analysis for DCR will include the comparison of DCR regardless of confirmation.

6.4.2.5 EORTC QLQ-C30

Instrument completion rate at each analysis visit will be reported for EORTC QLQ-C30:

- Completion rate (i.e., unadjusted) at each analysis visit will be calculated as the number of subjects meeting the minimum requirements for scoring at least one domain of the instrument divided by the number of subjects in the FAS population.
- Compliance rate (i.e., adjusted) at each analysis visit will be calculated among subjects who are expected to have PRO assessments. The following will be provided:
    - The number and % of subjects with all questions of EORTC QLQ-C30 completed
    - The number and % of subjects with at least one subscale of EORTC QLQ-C30 can be calculated (minimum requirements for scoring of the instrument)
    - The number and % of subjects with at least one question of EORTC QLQ-C30 completed

The completion (unadjusted) and compliance (adjusted) rates by treatment group at each analysis visit will also be provided graphically by means of a line graph.

Means and standard deviations and change from baseline at each scheduled assessment will be reported for each of the QLQ-C30 subscales. The analyses will include data from the baseline assessment through the last available data for all subjects in FAS.

Additionally, the change from baseline in the QL2 score will be analyzed using a restricted maximum likelihood (REML) based repeated measures approach (MMRM – Mixed Model Repeated Measures) (Brown et al 2006). The primary objective of this analysis is to compare EV versus chemotherapy at Week 12. The Week 12 timepoint was selected to minimize the impact of missing data given that median of PFS1 for the chemotherapy arm is 4 months (as indicated in the sample size calculation section).

The analysis will be based on observed data, i.e., data collected at each time point without carrying forward previous values. Only subjects with a baseline and at least one post-baseline score will be included in the analysis. Data from a limited number of PRO assessments may be used in case of substantial dropout (i.e., analysis will be limited to time points at which at least 10% of subjects have non-missing data in both treatment groups) by week 12.

The response variable will be the change from baseline to each PRO assessment. The model will include the following covariates:

- Fixed effects
    - Treatment arm (EV vs Chemotherapy)
    - Timepoint (categorical: visit)
    - Baseline PRO score (continuous)
    - ECOG PS (0 versus 1)
    - Liver Metastasis (yes vs no)
    - Regions of the world (US, Western Europe and Rest of World)
- Interactions
    - Baseline PRO score x time
    - Treatment arm x time

Both fixed effects and the interaction terms will remain in the model, regardless of significance. The model will present least squares (LS) mean estimates, standard errors, 95% CIs and p-values (where applicable) for mean changes from baseline to each visit. A plot of the LS means accompanied by the 95% CI will be produced.

The analysis will be conducted using PROC MIXED in SAS. The model will assume unstructured covariance among the within-subject repeated measurements. If the algorithm does not converge, a heterogeneous Toeplitz (the TOEPH option in SAS PROC MIXED) will be tried first and then AR(1) as a covariance structure to achieve convergence. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom.

Variables listed as categorical in the list above will be included in the CLASS statement of the procedure. The unique subject identifier will also be included as a class variable. A

REPEATED statement over the visits will be included with the unique subject identifier as the SUBJECT variable in the REPEATED statement.

Due to potential ePRO data integrity issue for site 39004, sensitivity analysis for the above MMRM analysis will be conducted for EORTC QLQ-C30 and for the mean change from baseline for each of the domains for the EORTC QLQ C30 tables by removing the data from site 39004.

To further understand patient experience and outcomes, additional analyses for EORTC QLQ-C30 will be specified in a separate PRO SAP.

Detailed scoring instructions for these scales are provided in Appendix 9.3. All time point data will be included and summarized.

6.4.2.6 EQ-5D-5L

Instrument completion rate at each analysis visit will be reported for EQ-5D-5L:

- Completion rate (i.e., unadjusted) at each analysis visit will be calculated as the number of subjects with either utility index or the VAS can be calculated (minimum requirements for scoring of the instrument) divided by the number of subjects in the FAS population.
- Compliance rate (i.e., adjusted) at each analysis visit will be calculated among subjects who are expected to have PRO assessments. The following will be provided:
    - The number and % of subjects with all questions completed (i.e., all 5 items and VAS)
    - The number and % of subjects with either utility index or the VAS can be calculated (minimum requirements for scoring of the instrument)
    - The number and % of subjects with at least one question of EQ-5D or VAS completed

The completion (unadjusted) and compliance (adjusted) rates by treatment group at each analysis visit will also be provided graphically by means of a line graph.

Descriptive characteristics of EQ-5D-5L will be summarized from baseline assessment through the last available data. Frequency and the percentage of reported problems for each level for each dimension will be provided. Detailed scoring instructions for these dimensions are provided in Appendix 9.4. All time point data will be included and summarized. EQ-5D-5L VAS will be summarized by treatment group at each visit using descriptive statistics (mean, standard deviation, minimum, maximum and median). 13 subjects from 4 sites were removed from the VAS analysis due to device display error on VAS scale and their data will not be included in the VAS analysis. Details specified in Appendix 9.5

Due to potential ePRO data integrity issue for site 39004, sensitivity analysis for the above EQ-5D-5L analysis will be conducted by removing the data from site 39004.

Additional analyses for EQ-5D-5L will be specified in a separate PRO SAP.

6.4.3 Analysis of Exploratory Endpoints

6.4.3.1 Nectin-4 and PD-L1 expression

Baseline Nectin-4 IHC H-score will be summarized using mean, standard deviation, minimum, maximum and median. An exploratory assessment of Nectin-4 IHC H-score with clinical outcomes will be conducted. If appropriate, efficacy data will be summarized using different H-score thresholds based on its distribution and historical data.

PD-L1 IHC combined positive score (CPS) will be summarized by two categories: <10 and >=10. An exploratory assessment of PD-L1 with clinical outcomes may be conducted.

If additional biomarkers are evaluated, the analysis will be described in a separate biomarker statistical analysis plan.

6.4.3.2 Incidence of antitherapeutic antibodies (ATA)

In the case that there are subjects with positive ATA post-baseline, the number and percentage of subjects for their post-baseline ATA status will be presented by the baseline ATA status. In addition, for subjects with positive ATA, individual subject titer levels will be displayed at each visit in a listing.

6.4.3.3 PFS in the next line of therapy (PFS2)

The distribution of PFS2 will be estimated on the FAS for each treatment arm using Kaplan-Meier methodology and compared between Arm A and Arm B using stratified log-rank test stratified by ECOG PS (0 vs. 1), region (US, EU and the Rest of world) and liver metastasis status (Yes versus No) per IRT. In addition, stratified Cox proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

6.4.3.4 Healthcare resources utilizations

Healthcare resources utilization such as incidence of emergency room visits, hospitalizations, and outpatient (either primary care or specialist) office visits will be summarized by treatment group by each visit. Duration of hospital stays will also be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median).

Due to potential ePRO data integrity issue for site 39004, sensitivity analysis for the above healthcare resources utilization analysis will be conducted by removing the data from site 39004.

6.5 Analysis of Safety

All analysis of safety will be presented by actual treatment group and overall for SAF, unless specified otherwise.

6.5.1 Adverse Events

All adverse event (AE) recorded on treatment including within 30 days from the last study treatment will be summarized.

The coding dictionary for this study will be MedDRA. It will be used to summarize AEs by SOC and PT. AEs will be graded using National Cancer Institute's Common Terminology Criteria for AEs (NCI-CTCAE).

An overview table will include the following details by treatment group and overall:

- Number and percentage of subjects with TEAEs,
- Number and percentage of subjects with causally drug-related TEAEs,
- Number and percentage of subjects with serious TEAEs and Astellas upgraded serious TEAE,
- Number and percentage of subjects with serious drug-related TEAEs and Astellas upgraded serious drug-related TEAE,
- Number and percentage of subjects with TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with drug-related TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with grade 3 or higher TEAE
- Number and percentage of subjects with grade 3 or higher drug-related TEAE
- Number and percentage of subjects with TEAEs leading to dose reduction,
- Number and percentage of subjects with drug-related TEAEs leading to dose reduction,
- Number and percentage of subjects with TEAEs leading to drug interruption,
- Number and percentage of subjects with drug-related TEAEs leading to drug interruption,
- Number and percentage of subjects with TEAEs leading to death,
- Number and percentage of subjects with drug-related TEAEs leading to death
- Number and percentage of subjects with TEAEs leading to death excluding disease progression,
- Number and percentage of subjects with drug-related TEAEs leading to death excluding disease progression, and
- Number of deaths

The above overview table will be repeated to report the number of events (all TEAEs which may include multiple events of the same preferred term with the same or different CTCAE grades) and the number of events adjusted by patient year (defined as the total duration of exposure in years).

The number and percentage of subjects with TEAEs, as classified by SOC and PT will be summarized by treatment group and overall. Summaries will be provided for:

- TEAEs
- Drug-related TEAEs,
- serious TEAEs and Astellas upgraded serious TEAE,
- drug-related serious TEAEs and drug-related Astellas upgraded serious TEAE,
- TEAEs leading to permanent discontinuation of study drug,
- Drug-related TEAEs leading to permanent discontinuation of study drug,

- TEAEs leading to dose reduction,
- Drug-related TEAEs leading to dose reduction,
- TEAEs leading to drug interruption,
- Drug-related TEAEs leading to drug interruption,
- TEAEs leading to death
- Drug-related TEAEs leading to death
- TEAEs leading to death excluding disease progression
- Drug-related TEAEs leading to death excluding disease progression
- grade 3 or higher TEAEs
- grade 3 or higher drug-related TEAEs
- TEAEs excluding serious adverse events that equal to or exceed a threshold of 5% in any dose level

AE summary tables will include subject counts as opposed to AE counts. If a subject experiences more than one episode of a particular AE, that subject will be counted only once for that event. If a subject has more than one AE that code to the same preferred term, the subject will be counted only once for that preferred term. Similarly, if a subject has more than one AE within a body system, the subject will be counted only once in that body system.

TEAEs and the number and percentage of subjects with TEAEs, as classified by SOC and PT will also be summarized by NCI-CTCAE severity grade and by relationship to study drug. If an adverse event changes in severity grade or relationship, then the subject will be counted only once with the worst severity grade and highest degree of relationship. The adverse event however will be presented in each category they were classified to. If a subject has an event more than once with missing severity grade and with non-missing severity grade, then the subject will be counted as the highest non-missing grade. If a subject has an event more than once with missing relationship and with non-missing relationship, then the subject will be counted as relationship=Yes. Drug-related TEAEs will be presented in a similar way by severity only.

The number and percentage of subjects with treatment-emergent adverse events of interest (AESI) such as IRR (infusion related reaction), ocular disorders, skin reactions hyperglycemia and neuropathy, as classified by SSQ/CMQ or SMQ and PT will also be summarized by treatment group and overall. The list of adverse events of interest to be summarized may change during the course of the study due to ongoing pharmacovigilance. It will be finalized before the database hard lock. All AEs, deaths, SAEs and withdrawals due to adverse events will be displayed in listings.

For selected AESI, time to onset, will be analyzed as appropriate. Time to onset of a specific AESI will be calculated as time from the first dose of study drug to the start of first treatment-emergent event that meets the respective search criteria. In the analysis of time to onset of AE of a specific grade (e.g., grade 3 or higher), episode of events that are improved from a previous higher grade will not be included.

Time to onset will be summarized at the subject level.

6.5.2 Clinical Laboratory Evaluation

Laboratory assessments are from central lab. The baseline visit is the last measurement taken prior to initial study drug administration. Plots of median plus 25th and 75th percentile lab values at each scheduled assessment time by treatment group will be provided for each laboratory parameter.

Laboratory results will be graded using NCI-CTCAE, where possible. Parameters that have criteria available for both low and high values, i.e., hypo- and hyper-, will be summarized for both criteria. The same subject can be counted for both values if the subject has different laboratory values meeting each criterion. NCI-CTCAE grade of laboratory evaluations will be summarized by number and percentage of subjects for each visit. Shift tables of NCI-CTCAE grade change from baseline to worst post-baseline grade for selected laboratory parameters will also be presented. Laboratory abnormalities reported in ≥ 10% (All Grade) or ≥ 5% (Grade 3-4) of all subjects in SAF will be presented by treatment groups.

6.5.2.1 Liver Safety Assessment

The following potentially clinically significant criteria in liver function tests for Alkaline Phosphatase (ALP), Alanine Transaminase (ALT), total bilirubin, Aspartate Transaminase (AST) and their combination are defined. The subject's highest value during the investigational period will be used.

Parameter	Criteria
ALT	> 3xULN > 5xULN > 8xULN > 10xULN > 20xULN
AST	> 3xULN > 5xULN > 8xULN > 10xULN > 20xULN
ALT or AST	> 3xULN > 5xULN > 10xULN > 20xULN
Total Bilirubin (TB)	> 2xULN
Alkaline Phosphatase (ALP)	> 1.5xULN

Parameter	Criteria
International Normalized Ratio (INR)	>1.5
ALT and/or AST AND TB (*)	(ALT and/or AST > 3xULN) and (TB > 2xULN)
ALT and/or AST AND INR (*)	(ALT and/or AST > 3xULN) and (INR> 1.5)
ALT and/or AST AND ALP AND TB (*)	(ALT and/or AST > 3xULN) and (ALP < 2xULN) and (TB > 2xULN)

(*) Combination of values measured within same day or within 1 day apart

The number and percentage of subjects with potentially clinically significant values in liver enzymes and total bilirubin during the investigational period will be presented by treatment group.

6.5.3 Vital Signs and Weight

The baseline visit is the last measurement taken prior to initial study drug administration. Vital signs (systolic blood pressure, diastolic blood pressure, pulse rate and body temperature) and weight will be summarized using mean, standard deviation, minimum, maximum and median by treatment group and visit. Additionally, a within-subject change will be calculated per visit as the post-baseline measurement minus the baseline measurement and summarized by treatment group and visit.

6.5.4 Electrocardiograms

Number and percentage of subjects with normal and abnormal results as assessed locally for the overall interpretation will be tabulated at each treatment visit.

6.5.5 Pregnancies

A detailed listing of all pregnancies will be provided.

6.5.6 Eastern Cooperative Oncology Group (ECOG) Performance Scores

Summary statistics (number and percent of subjects) for each category of the ECOG performance status at each assessment will be provided. Negative change scores indicate an improvement. Positive scores indicate a decline in performance. ECOG will also be summarized using shift table from baseline to worst post-baseline score by visit.

6.5.7 Ophthalmologic Assessment

Ophthalmologic assessment for subjects with recent ocular complaints (within 3 months of screening) are required. Assessments will include the following: visual acuity, slit lamp, tonometry examination and dilated fundus examination.

Below variables will be summarized by descriptive statistics and will be presented at each visit (whenever data are available) by treatment groups and by eyes (left vs. right).

- Visual acuity: method of BCVA Assessment, BCVA score and result
- Slit lamp: biomicroscopy interpretation
- Tonometry examination: testing result
- Dilated fundus examination: testing result

6.6 Analysis of Pharmacokinetics

Descriptive statistics will include n, mean, SD, geometric mean, minimum, median, maximum, %CV, and geometric %CV. PK analysis will be conducted on the PKAS.

Descriptive statistics will be presented for serum concentrations of Tab and ADC and for plasma concentrations of MMAE by scheduled time. If appropriate, plots of the mean concentrations over time will be produced for each analyte.

A separate PK Analysis Plan may be produced to describe potential model-based analyses assessing the relationship between PK concentrations and select safety and/or efficacy endpoints. The results and the model development will be described in detail in a separate population PK report.

6.7 Subgroup of Interest

To determine whether the treatment effect is consistent across various subgroups, the estimate of the between-group treatment effect (with a nominal 95% CI) for OS, PFS1 and ORR, will be estimated and plotted within each category of the following classification variables:

A subgroup analysis may not be performed if the number of subjects in the subgroup in each treatment group is not sufficiently large (e.g., <5%). In the case of the subgroup variable with more than two levels, pooling may be considered when there is no sufficient sample size within the level.

Grouping variable	Subgroups
Age	<65
	≥65
	<75
	≥75
Sex	female
	male

Grouping variable	Subgroups
Region	Western EU
	US
	Rest of the world
ECOG PS	0
	1
Liver metastasis	yes
	no
Pre-selected control therapy by investigator	Paclitaxel
	Docetaxel
	Vinflunine
Primary site of tumor	upper tract (renal pelvis or ureter)
	bladder/other (including urethra, bladder and other)
Prior lines of systemic therapy	1-2 lines
	≥ 3 lines
Best response to prior CPI	Responder
	Non-Responder

6.8 Interim Analysis (and Early Discontinuation of the Clinical Study)

An interim efficacy analysis is planned to occur after approximately 285 OS events (about 65% of the total planned events) are observed. OS will be tested at 1-sided 0.00541 significance level for efficacy according to the O'Brien-Fleming boundary as implemented by Lan-DeMets alpha spending function [Lan-DeMets, 1983]. The IDMC may recommend terminating the trial at the interim analysis based on statistically significant OS results favoring EV. When total deaths reach 439, the final OS analysis will be conducted at the 1-sided 0.02332 significance level. If the exact number of events at interim and final analyses are different than planned, the significance level will be adjusted accordingly, based on the O'Brien-Fleming method with a Lan-DeMets alpha spending function.

The interim analysis will be conducted by the IDAC and the results will be reviewed by IDMC. In addition, safety data reviews during the trial will be conducted by the IDMC on a periodic basis. For example, the IDMC will review safety data after the first 50 subjects have

been randomized and on study drug for approximately 3 months. The full procedures for IDMC safety review and interim analysis will be described in IDMC charter and IAP.

6.9 Additional Conventions

6.9.1 Missing Data Imputation

In the imputation of missing or partial dates, if the imputed date is after min (death date, cutoff date), min (death date, cutoff date) will be used as imputed date.

Missing or partial start and stop dates of adverse events and concomitant medication will be imputed using the following algorithm:

- Imputation rules for partial or missing stop dates:
  - If the month and year are present, then impute as the last day of that month.
  - If only the year is present, impute as December 31 of that year.
  - If the stop date is entirely missing, assume the event or medication is ongoing.
- Imputation rules for partial or missing start dates:

Start Date		Stop Date						
		Complete: *yyyymmdd*		Partial: *yyyymm*		Partial: *yyyy*		missing
		< 1st dose	≥ 1st dose	< 1st dose *yyyymm*	≥ 1st dose *yyyymm*	< 1st dose *yyyy*	≥ 1st dose *yyyy*	
Partial: *yyyymm*	= 1st dose *yyyymm*	2	1	n/a	1	n/a	1	1
	≠ 1st dose *yyyymm*		2	2	2	2	2	2
Partial: *yyyy*	= 1st dose *yyyy*	3	1	3	1	n/a	1	1
	≠ 1st dose *yyyy*		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute as the date of first dose; 2 = Impute as the first of the month; 3 = Impute as January 1 of the year; 4 = Impute as January 1 of the stop year

The imputed dates will be used to determine whether an AE is/is not treatment emergent. Listings of AEs and concomitant medications will present the actual partial dates; imputed dates will not be shown.

In the case of partial starting date of subsequent anti-cancer therapy, the date will be imputed to the first day of the month but not earlier than the last dosing date of the study drug. A month and year must be present or the date will remain missing.

Partial missing date of prior therapy: for start date, the date will be imputed to the first day of the month; for the end date, the date will be imputed to the last day of the month or 14 days before the first dose of study drug, whichever is earlier. A month and year must be present or the date will remain missing.

For continuous variables (e.g., clinical laboratory measurement, vital signs), subjects with missing baseline variable will be excluded from the analysis of change from baseline.

Subjects who do not satisfy the criteria to be counted as responders or have insufficient data to determine or confirm a response per the RECIST guidelines (Version 1.1) will be considered as non-responders in the final analysis of response rates. No imputation of data will be done to determine individual subject response.

For all analyses other than PK analysis, all values will be included in the analyses. For analysis of PK data, only samples for which the time of sampling relative to the dose administration and the exact dose is known will be included.

6.9.2 Outliers

All values will be included in the analyses.

6.9.3 Analysis Windows

CRF visit will be used for analysis. For safety analysis, in the case of multiple observations at a specific visit, the observation which is the latest will be used. If more than one observation is made on the same day, an average value if continuous or the worst value if categorical will be included in the analysis.

For PRO endpoints, visit windows will be defined for each time point when PROs were collected prior to the study drug (arm A) or chemotherapy (arm B) administration as presented in below Table. According to the protocol, QOL questionnaires should be completed by the subject at home on hand-held devices prior to coming to the clinic visit with the exception of Baseline Day 1 of the first week, the EOT and the follow-up visits, at which the QOL questionnaires will be completed by the subject at the clinic. In the case of multiple observations in the same analysis visit window, the observation which is closest to the target date will be used. Additional derivation details will be provided in the ADaM specifications.

Analysis visit windows for PRO

Analysis Visit	Target day	Analysis window
Baseline (randomized but not treated)	1 (randomization date)	[Day -7, Day 7]
Baseline (treated)	1(first dose date)	[Day -7, Day 1]
Week 1	8	[Day 2, Day 11]
Week 2	15	[Day 12, Day 18]
Week 3	22	[Day 19, Day 25]
Week 4	29	[Day 26, Day 32]
…	…	…
Week 12	85	[Day 82, Day 127]
Week 24	169	[Day 128, Day 211]
Week 36	253	[Day 212, Day 295]
…	…	…
EOT	last dose date	[last dose date, last dose date+29]
30-day follow-up	last dose date+30	[last dose date+30, last dose date+60]

HRU questionnaires will be completed monthly (Day 1 of every 4 weeks), starting on Week 5 Day 1, and at the EOT and follow-up visit. HRU questionnaires will be completed by the subject at home on hand-held devices prior to coming to the clinic visit with the exception of the EOT and the follow-up visits at which the HRU questionnaires will be completed by the subject at the clinic. For HRU, visit windows will be defined for each time point when data were collected as presented in below Table.

Analysis visit windows for HRU

Analysis Visit	Target day	Analysis window
Week 5	36	[Day 22, Day 49]
Week 9	64	[Day 50, Day 77]
Week 13	92	[Day 78, Day 105]
Week 17	120	[Day 106, Day 133]
…	…	…
EOT	last dose date	[last dose date, last dose date+29]
30-day follow-up	last dose date+30	[last dose date+30, last dose date+60]

6.9.4 Blinding process

Although the study is an open label study, to maintain trial integrity and increase the credibility of study results, aggregate analyses or summaries by randomized treatment assignment or actual treatment assignment will be limited and documented before the primary database lock. The study team consists of medical, clinical operations, statistics, statistical programming and data management personnel. For the purpose of preventing unintentional efficacy and safety summary generated by treatment assignment, the below process will be implemented for this study (more details can be found in Appendix 9.2):

- The study statistician, support statisticians, study lead programmer and support programmers will form a study data analysis team (*sdat*). The *sdat* will have no access to the randomized/actual treatment assignment before the primary database lock. Dummy treatment codes will be used to prepare analysis programs.
- The datasets which may reveal subject's treatment information, e.g., dose administration, PK concentration data are considered as restricted data and will be stored in a separate study folder and can only be accessed and handled by a separate team call the restricted data analysis team (*rdat*).
- Communication between the *sdat* and the *rdat* will be limited to un-restricted data.
- Other study team members (including the study manager, medical monitor and data manager) may have the access to the treatment information at the individual subject level for data review and cleaning purpose.
- There will be no summaries by treatment assignment generated during the study until primary database lock. Treatment level aggregated results at interim analysis will be provided by IDAC to IDMC.
- Limited additional sponsor personnel may be unblinded to the treatment level aggregated results of the interim analysis, if required in order to act on the recommendations of the IDMC or facilitate regulatory filing after interim analysis.

7 REVISION AND RATIONALE

Version	Date	Changes	Comment/rationale for change
1.0	19-JUN-2018	NA	Document finalized
2.0	11-Feb-2020	Section 2.2 Sample size changes from 550 to 600	Per protocol amendment 2, the planned enrollment is being increased from 550 patients to 600
		Section 3 ○ Total OS event at final analysis changes from 384 to 439 ○ Add secondary endpoint PFS1 total event size	Per protocol amendment 2, the targeted number of death events increased from 384 to 439 to increase the power of the study from 80% to 85% Add PFS1 total event to make it clear when calculating the information fraction of PFS1 at the interim
		Section 4 ○ Removed PPS ○ Clarified SAF for subjects received any amount of EV ○ Clarified RES to add 6 months follow up time	To be consistent with protocol amendment 2
		Section 5 ○ Fixed some minor typos ○ Clarified primary analysis set for some secondary endpoints and PRO endpoints ○ Clarified the BOR definition for subjects who randomized but not dosed ○ Clarified the definition of PFS2 ○ Clarified the definition of time to response ○ Corrected the formula to calculate the duration of exposure and dose intensity ○ Clarified the definition of concomitant medication	To make the SAP more clear and precise
		Section 6 ○ Removed PPS related analyses ○ Clarified the baseline definition for subjects randomized but not dosed ○ Refined analysis variables in demographic and baseline characteristics ○ Refined analysis variables in prior systemic treatment ○ Removed infusion related analysis in drug exposure ○ Clarified the multiplicity adjustment rule	To be consistent with protocol amendment 2 To refine analysis per TLFs dry run review comments To make the SAP easy to read and more precise

Version	Date	Changes	Comment/rationale for change
		○ Added AESI analyses including time to onset and time to resolution ○ Clarified the ECOG endpoint analysis ○ Modified the subgroup analyses and removed some of the subgroup analysis variables which are either with small sample size or considered not useful per team's discussion ○ Removed the imputation rules for death and PD dates because no longer allow partial date for both dates ○ Added imputation algorithm for starting and end date of prior therapy ○ Added the analysis window definition for PRO endpoints ○ Modified the interim analysis planned event size	
3.0	10-AUG-2020	Section 4.3 RES definition change: Removed the criterion of 'have the opportunity to be followed at least 6 months since randomization'.	Due to FDA comment received on 5/7/2020
		Section 5.1.1 and 6.4.1: Remove the sensitivity analysis censoring OS at the time of new anticancer therapy for patients who started new therapy	We already included the sensitivity analyses based on RPSFT and IPCW methods to evaluate the impact of the use of subsequent anti-cancer therapy.
		Section 5.2.1 and 5.2.3: Update the censor rules for PFS1 and DOR to also censor the PD or death after 2 or more missed disease assessments	Due to FDA comment received on 5/7/2020
		Section 5.2.6 and 6.4.2.5: Clarify the definition for EORTC QLQ-C30 scale scores and add MMRM analysis on QL2	MMRM analysis on QL2 is added in this SAP
		Section 5.3.2 and 6.4.3.2 refine the language of ATA positive definition and analysis.	Clarify the definition and analysis of ATA.
		Section 5.5.4 Subsequent Anti-cancer Therapy: Add this section.	Clarify the definition of subsequent anti-cancer therapy
		Section 6.2.5: Add additional variables to summarize the prior CPI	Provide the variables to summarize prior CPI as PD-L1 inhibitor only, PD-1 inhibitor only and both
		Section 6.2.6 and 6.2.7: Delete the unnecessary summaries.	These data will be presented in listings.

Version	Date	Changes	Comment/rationale for change
		Section 6.4: Update multiplicity adjustment	Due to enrollment completion earlier, the multiplicity strategy was updated accordingly.
		Section 6.4.1 and 6.4.2.1: Add sensitivity analyses of OS and PFS1 due to COVID-19	Add these sensitivity analyses to assess potential COVID-19 impact
		Section 6.4.2.5, 6.4.2.6 and 6.4.3.4: Add sensitivity analysis to remove site 39004	Due to potential ePRO data integrity issue for site 39004, sensitive analyses were conducted for EORTC-QLQ-C30, EQ-5D-5L and HRU when removing the data from site 39004.
		Section 6.4.2.6 and Appendix 9.5, Explain 13 subjects will be removed from VAS analysis	13 subjects from 4 sites were removed from the VAS analysis due to device display error on VAS scale
		Section 6.5.1: Add the summaries for overview and by SOC and PT for TEAEs and drug-related TEAEs leading to death excluding disease progression	Add these analysis to exclude the death due to disease progression.
		Section 6.5.1: Remove time to resolution/improvement analysis	Due to the limitation of eCRF data collection method, it is difficult to identify records from the same event to allow performing this type of analysis.
		Section 6.9.1: Add the general rule of imputation of missing or partial dates to ensure imputed date won't be after death date or cutoff date.	Ensure imputed date won't be after death date or cutoff date
		Section 6.10.3: Update the analysis window for PRO and add the analysis window for HRU	Align with the analysis windows used in PRO SAP.
		Appendix 9.6: Add this Appendix to provide descriptions of analysis to be performed to support PMDA submission work.	Describe the additional analysis required by PMDA.

8 REFERENCES

Bellmunt J, Choueiri TK, Fougeray R, Schutz FA, Salhi Y, Winquist E, Culine S, von der Maase H, Vaughn DJ and Rosenberg JE (2010). Prognostic factors in patients with advanced transitional cell carcinoma of the urothelial tract experiencing treatment failure with platinum-containing regimens. J Clin Oncol 28: 1850-5.

Brown H and Prescott R (2006). Applied Mixed Models in Medicine. John Wiley and Sons Ltd, 2nd Edition.

Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1) EU J of Cancer 2009:45, 228-247

Fayers P et al (2001). EORTC QLQ-C30 Scoring Manual (3rd edition). Brussels: EORTC Quality of Life Group.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika.1983;70:659-63.

Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun Stat Theory Methods. 1991;20(8):2609–31.

Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000;56(3):779–88.

9 APPENDICES

9.1 Appendix 1: Key Contributors and Approvers

List of Key Contributors and Approvers

Key Contributors

The following contributed to or reviewed this Statistical Analysis Plan as relevant to their indicated discipline or role.

Primary author (s)

PPD	Global Data Science
	Astellas Pharma Global Development, Inc.

Contributors and Reviewers

PPD	Global Data Science
	Astellas Pharma Global Development, Inc.
PPD	Global Data Science
	Astellas Pharma Global Development, Inc.
PPD	Global Medical Science
	Astellas Pharma Global Development, Inc.
PPD	Global Medical Science
	Astellas Pharma Global Development, Inc.
PPD	Global Medical Science
	Astellas Pharma Global Development, Inc.
PPD	Global Oncology TA
	Astellas Pharma Global Development, Inc.

PPD	Pharmacovigilance Medical Safety Astellas Pharma Global Development, Inc.
PPD	Global Clinical Science Astellas Pharma Global Development, Inc.
PPD	PK & Exploratory Statistics, Global Data Science, Astellas Pharma Global Development, Inc.
PPD	Global Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.
PPD	Global Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.
PPD	Global Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.
PPD	Global Data Science Astellas Pharma Global Development, Inc.
PPD	GDO Management Global Data Science Astellas Pharma Global Development, Inc.

Author and Approver Signatories

(E-signatures are attached at end of document)

PPD, APGD was the study statistician for this study.

PPD, APGD was the biostatistics peer reviewer of this Statistical Analysis Plan

This Statistical Analysis Plan was approved by:

PPD

Global Medical Science

9.2 Appendix 2: Open-Label Blinding Process

As specified in SAP Section 6.9.4, although the study is an open label study, to maintain trial integrity and increase the credibility of study results, aggregate analyses or summaries by randomized treatment assignment or actual treatment assignment will be limited and documented before the primary database lock (DBL).

Data Analysis Organization

To fulfill the open-label blinding process for this study, there will be two separate analysis teams to support internal data review and preparation of analysis for CSR before primary database lock:

Study data analysis team (sdat): consists of the study statistician, support statisticians, study lead statistical programmer and support programmers. The major responsibility for *sdat* is to develop STDM/ADaM datasets and TLFs programs in preparation of CSR.

Resticted data analysis team(rdat): consists of an independent statistician and support statistical programmers. The major responsibility for *rdat* is to develop datasets and TLFs programs for the restricted data. Restriced data are the data which may reveal subject's treatment information, e.g., dose administration, PK concentration data and lab data for this study.

The *sdat* will have no access to the randomized treatment assignment as well as the restricted data before the primary database lock. The *sdat* is not allowed to review any listings on restricted data.

The *rdat* will only be involved in developing datasets and TLFs programs on the restricted data.

Communication between the *sdat* and the *rdat* will only be limited to un-restricted data.

No summaries by treatment assignment can be generated by either *sdat* or *rdat*. Listings without treatment assignment or with dummy treatment codes may be generated for data review purpose, but should be limited.

Handling of Treatment Codes

For this study, treatment codes can be obtained from two sources: IRT vendor and EDC system. The following procedures should be applied when handling data extraction, data transferring and TLFs preparation to maintain the blinding:

- The treatment codes collected in EDC system will not be extracted until the primary DBL. The treatment variable will be blocked for any data extraction during the study conducting period. Program used to block treatment variable will be set up before the first data extraction.
- Dummy treatment codes will be created by *sdat* for preparing the SDTM/ADaM datasets and TLFs for primary CSR.

- Raw data (SDTM/ADaM if needed) without treatment codes (or with dummy codes) will be provided by *rdat* to IDAC to conduct interim analysis or other IDMC requested analysis.
- Separate IRT transfer request will be sent directly by Astellas Independent Statistician to IRT vendor for transferring treatment codes. The Astellas Independent Statistician will transfer the treatment codes to IDAC for the purpose of performing formal interim analysis or other IDMC requested analysis (e.g., safety summaries to support routine safety monitoring meetings).

Handling of Restricted Data

For this study, there are three major types of restricted data:

- Study Drug Administration
- PK data (Only Arm A (EV) has PK)
- Safety data by visit (cycle length different, EV=28 days, chemo=21 days)

The following procedures should be applied when handling the restricted data:

- The restricted data will be extracted/transferred to a seperate folder which is only accessible to the *rdat*.
- For the IDMC analysis, the restricted data will be included in a password protected folder during the data transfer. The password will be sent to the IDAC statistican by the *rdat* directly per IDAC statistician request. No *sdat* are allowed to receive the password.
- Certain variables may be necessary for the purpose of developing programs by the *sdat*, e.g., last dosing date. These variables are considered as un-restricted data and allowed to share with the *sdat*. The variables should be reviewed by study statistican in determining whether the sharing is necessary, before they are created and shared.
- Listings without treatment codes or with dummy treatement codes may be generated on restricted data for data review purpose. These listings should be uploaded to a secure sharepoint site which can’t be assessed by the *sdat*.

9.3 Appendix 3: EORTC QLQ-30 Scoring Algorithm

Table 1: Scoring the QLQ-C30 version 3.0

	Scale	Number of items	Item range*	**Version 3.0** Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

Then for **Functional scales**:

$$Score = \left\{1 - \frac{(RS - 1)}{range}\right\} \times 100$$

and for **Symptom scales / items** and **Global health status / QoL**:

$$Score = \{(RS - 1)/range\} \times 100$$

Examples:

Emotional functioning

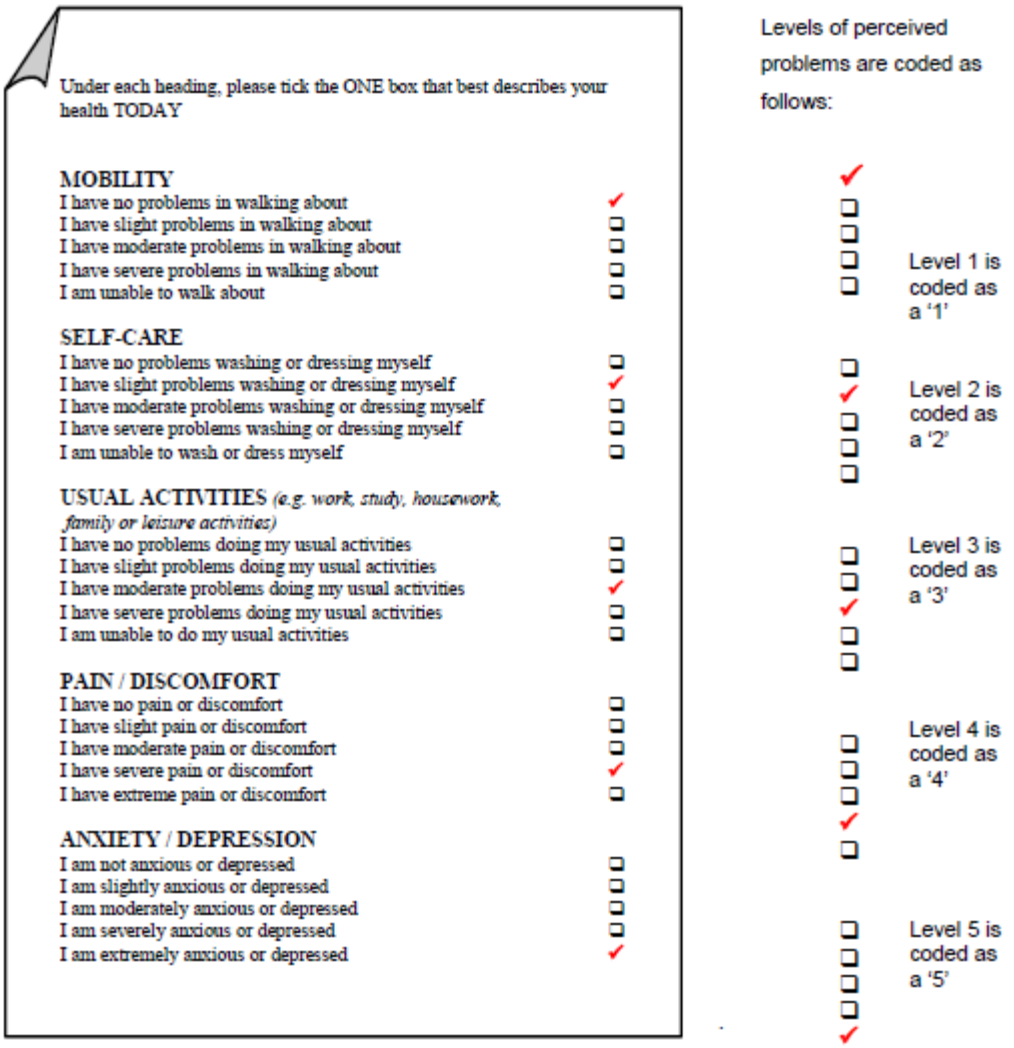
$$RawScore = (Q_{21} + Q_{22} + Q_{23} + Q_{24})/4$$
$$\text{EF } Score = \{1 - (RawScore - 1)/3\} \times 100$$

Fatigue

$$RawScore = (Q_{10} + Q_{12} + Q_{18})/3$$
$$\text{FA } Score = \{(RawScore - 1)/3\} \times 100$$

9.4 Appendix 4: EQ-5D-5L Scoring Algorithm

The EQ-5D-5L descriptive system should be scored, for example, as follows:



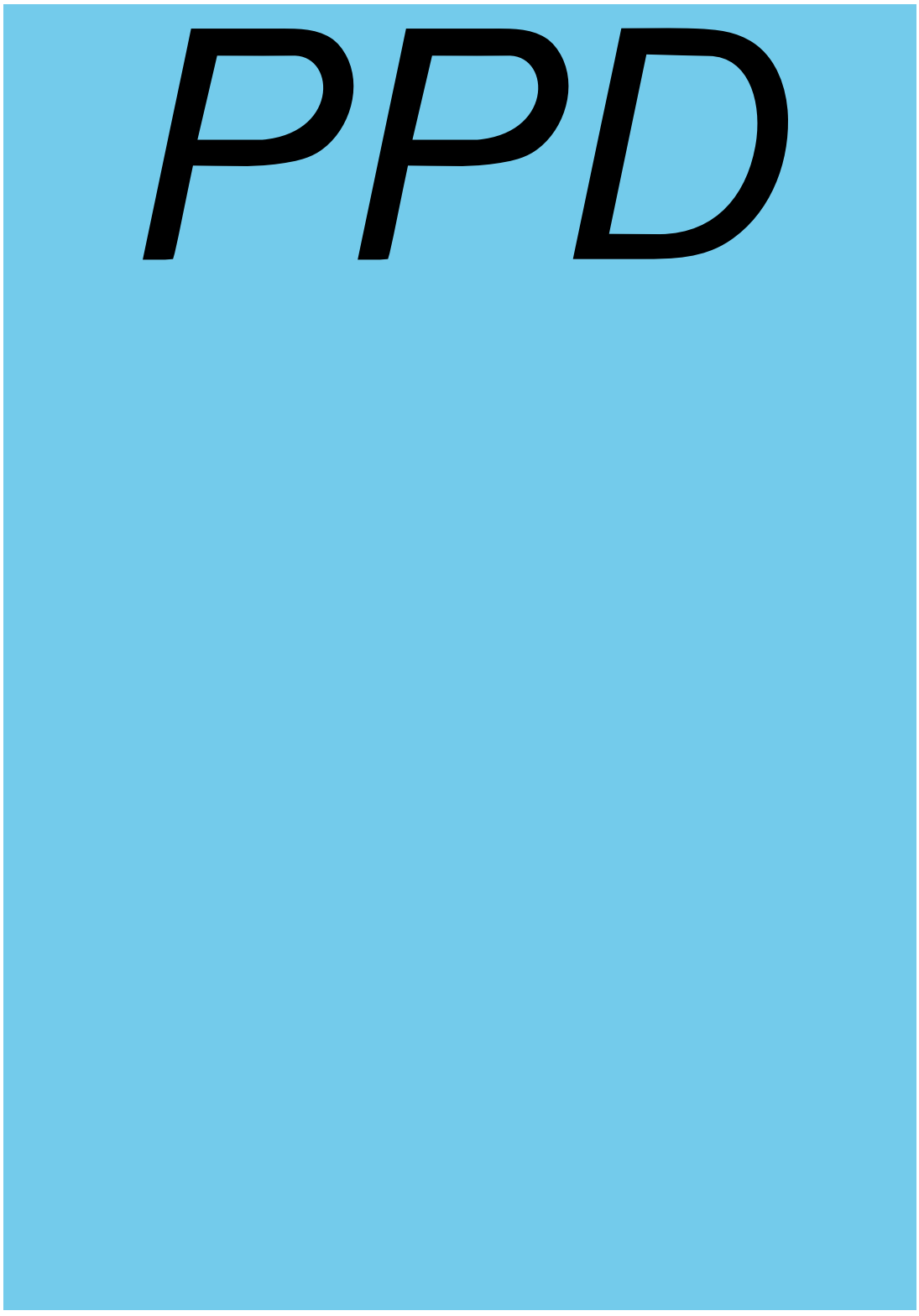
This example identifies the health state '12345'.

***NB:** There should be only <u>ONE</u> response for each dimension*

***NB:** Missing values can be coded as '9'.*

***NB:** Ambiguous values (e.g. 2 boxes are ticked for a single dimension) should be treated as missing values.*

9.5 Appendix 5: Memo of Device Display Error on VAS in EQ-5D-5L




PPD

9.6 Appendix 6: Japan Specific Analysis

9.6.1 Introduction

This appendix describes the planned regional specific analyses in addition to the analyses defined by the main body of this SAP. The regional analyses specified in this appendix are for the PMDA submission only.

9.6.2 Region Specific Analyses of Japan

The following regional specific analyses of Japan will be performed as appropriate:

- All analyses as described in the main body of this SAP will be repeated for the subgroups of Japan vs Non-Japan
- All listings will be presented by sites ordered as Japan followed by Non-Japan sites

Below shows a couple of examples where the regional specific analyses of Japan will be handled differently from what are described in the main body of this SAP. The exceptions will be reflected in the corresponding TLF specifications, where the detailed Specifications for table, figure, and data listing formats can be found.

Example(s) of Region Specific Analyses of Japan with Difference

Section(s)	Change(s)	Comment/rationale for change
6.4.1 and 6.4.2	The endpoints of OS, PFS1, ORR and DCR will be evaluated using unstratified analysis only.	Due to the small number of Japan patients, these efficacy analysis will be performed using un-stratified analysis instead of stratified analysis.