Statistical Analysis Plan

Study Code	D419BC00001
Edition Number	Version 5.0
Date	Dec 4, 2019

A Phase III, Randomized, Open-label, Controlled, Multi-Center, Global Study of First-Line MEDI4736 Monotherapy and MEDI4736 in Combination with Tremelimumab Versus Standard of Care Chemotherapy in Patients with Unresectable Stage IV Urothelial Cancer

A Phase III, Randomized, Open-label, Controlled, Multi-Center, Global Study of First-Line MEDI4736 Monotherapy and MEDI4736 in Combination with Tremelimumab Versus Standard of Care Chemotherapy in Patients with Unresectable Stage IV Urothelial Cancer

STUDY STATISTICIAN

PPD

12 December 2019

PPD (AstraZeneca) Date

A Phase III, Randomized, Open-label, Controlled, Multi-Center, Global Study of First-Line MEDI4736 Monotherapy and MEDI4736 in Combination with Tremelimumab Versus Standard of Care Chemotherapy in Patients with Unresectable Stage IV Urothelial Cancer

GLOBAL PRODUCT STATISTICIAN PPD 12 December 2019

PPD (AstraZeneca) Date

TABLE OF CONTENTS

PAGE

TABLE OF TABLES

TABLE OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
3i score	Immune Immediacy Index
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse events of special interest
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
APF12	Proportion of patients alive and progression free at 12 months from randomization
AST	Aspartate aminotransferase
BlCS	Bladder cancer subscale
BDRM	Blinded data review meeting
BICR	Blinded independent central review
BoR	Best objective response
CI	Confidence Interval
CR	Complete Response
CRF/eCRF	Case Report Form (electronic)
CSR	Clinical Study Report
CSP	Clinical Study Protocol
CT	Computed tomography
CTC/CTCAE	Common Terminology Criteria for Adverse Event (National Institutes of Health, National Cancer Institute)
bTMB	Tumour mutational burden in circulating tumour DNA
CV	coefficient of variation
DAE	Discontinuation of Investigational Product due to Adverse Events
DBL	Database lock
DCR	Disease control rate
DCO	Data cut off
DoR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EQ-5D-5L	EuroQol 5-dimension, 5-level health state utility index
EWB	Emotional well-being
FACT-BL	Functional Assessment of Cancer Therapy-Bladder Cancer
FAS	Full analysis set
FST-FS	First subsequent therapy free survival
FWB	Functional well-being
HR	Hazard ratio
HRQoL	Health-related Quality of Life
ICU	intensive care unit
IDMC	Independent data monitoring committee
IHC	Immunohistochemistry
IP	Investigational Product
imAE	Immune-mediated adverse event
irND	Immune-related No evidence of disease
irRECIST	Immune-related Response Evaluation Criteria in Solid Tumors

Abbreviation or special term	Explanation
IPD	Important protocol deviations
ITT	Intention to Treat
IV	Intravenous
IVRS	Interactive Voice Response System
LD	Longest diameter
LSI	Last patient in
LSMEANS	Least squared means
MEDI4736	Durvalumab
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
MMRM	Mixed model for repeated measures
MRI	Magnetic Resonance Imaging
MTP	Multiple testing procedure
NA	Not applicable
NE	Not evaluable
ND	No disease, by BICR
NFBlSI-18	National Comprehensive Cancer Network-FACT Bladder Symptoms Index-18
NTL	Non-target lesions
OAE	Other Significant Adverse Event
ORR	Objective Response Rate
OS	Overall Survival
OS24	Proportion of patients alive at 24 months from randomization
PD	Progressive disease
PD-L1	Programmed death ligand 1
PDx	Pharmacodynamic
PFS	Progression-free survival
PFS2	Time from randomization to second progression
PGIC	Patient global impression of change
PID	Percentage intended dose
PK	Pharmacokinetics
PR	Partial Response
PRO	Patient Reported Outcome
PRO-CTCAE	Patient-reported outcomes version of the Common Terminology Criteria for Adverse Events
PWB	Physical well-being
QoL	Quality of Life
QTcF	QT interval (corrected for heart rate using Fridericia's correction)
RECIST 1.1	Response Evaluation Criteria In Solid Tumours, Version 1.1
REML	Restricted maximum likelihood
RDI	Relative dose intensity
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Safety analysis set
SD	Stable disease
sf	Significant figures
SoC	Standard of Care
SST-FS	Second subsequent therapy free survival

Abbreviation or special term	Explanation
SWB	Social/family well-being
TEAE	Treatment emergent adverse event
TL	Target lesions
TTR	Time to response
UC	Urothelial bladder cancer

AMENDMENT HISTORY

Date	Brief description of change
V1.0 (02 Dec 2015)	Original version
V2.0 (02 Mar 2017)	Updated to reflect changes from version 3, 4, and 5 of the protocol
V3.0 (21 Nov 2017)	Updated to reflect changes from version 6 of the protocol
V4.0 (04 Sep 2018)	Added an appendix for detail information about the significant level to be applied to the tests in the MTP Dropped relative dose intensity calculation in SoC Added sample size estimation for assessing the efficacy of MEDI4736 + tremelimumab combination therapy versus SoC in terms of OS in patients with PD-L1 low/negative UC Added exploratory analyses of OS for MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy in PD-L1 high UC, and MEDI4736 monotherapy versus SoC in PD-L1 low/negative UC Added exploratory analyses of PFS per cisplatin eligible status at baseline per eCRF

Date	Brief description of change
V5.0 (4 Dec, 2019)	Removed IA3 according to the protocol v8.0. Significance level, critical value, etc for final analysis were updated accordingly. Updated the languages for when to perform the final analysis according to the protocol v8.0. Added stratified max-combo test as an exploratory analysis for the primary analysis. Added 3i score analysis as an exploratory analysis. Clarified PFS2 definition. Added Time to deterioration analysis for some PRO subscales and total scores. Defined bTMB analysis set. Added secondary analyses of PFS and OS in bTMB high (≥ 24 mut/Mb). Added exploratory analyses of PFS and OS based on bTMB using cuts-off of 16 and 20 mut/Mb. Removed subgroup analysis for safety data. Changed imAE analysis based on the latest imAE Charter. Replaced language of “causally related" AEs with language of "possibly related" AEs. In the ECG section, the text was updated to indicate that a listing will be done and not a summary table. Removed “tumour grade” from the baseline characteristics data. Removed text regarding a summary of the number of post-discontinuation treatment regimens.

1. STUDY DETAILS

1.1 Study objectives

All objectives will be evaluated for all randomized patients, unless otherwise indicated, and where appropriate for patients with PD-L1 (programmed death ligand 1)-high and/or PD-L1–low/negative Urothelial cancer (UC).

1.1.1 Primary objective

Primary objectives:	**Outcome measures:**
To assess the efficacy of MEDI4736 + tremelimumab combination therapy versus SoC in terms of OS in patients with unresectable Stage IV UC	OS
To assess the efficacy of MEDI4736 monotherapy versus SoC in terms of OS in patients with unresectable Stage IV PD-L1-High UC	OS

OS Overall survival; UC Urothelial cancer.

1.1.2 Secondary objectives

Secondary objectives:	**Outcome measures:**
To assess the efficacy of MEDI4736 monotherapy compared to SoC in terms of PFS in patients with PD-L1-High UC	PFS using investigator assessments according to RECIST 1.1[a]
To assess the efficacy of MEDI4736 + tremelimumab combination therapy versus SoC in terms of PFS in patients with UC	PFS using investigator assessments according to RECIST 1.1[a]
To assess the efficacy of MEDI4736 monotherapy compared to SoC in terms of PFS and OS in patients with UC	PFS using investigator assessments according to RECIST 1.1[a] OS
To assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of PFS and OS in patients with PD-L1-Low/Neg UC	PFS using investigator assessments according to RECIST 1.1[a] OS
To assess the efficacy profile of MEDI4736 monotherapy in patients who are not cisplatin-eligible	ORR, DoR, DCR, TTR and PFS, using BICR assessments according to RECIST 1.1 [b] OS

Secondary objectives:	**Outcome measures:**
To further assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to MEDI4736 monotherapy in terms of PFS, OS, OS24, APF12, ORR, DoR, DCR, and PFS2 in patients with PD-L1-Low/Neg UC and all patients with UC	OS OS24 PFS, APF12, ORR, DoR, and DCR using investigator assessments according to RECIST 1.1 [a] PFS2 as defined by local standard clinical practice
To further assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of OS24, APF12, ORR, DoR, DCR, and PFS2 in patients with UC	OS24 APF12, ORR, DoR, and DCR using investigator assessments according to RECIST 1.1 [a] PFS2 as defined by local standard clinical practice
To further assess the efficacy of MEDI4736 monotherapy compared to SoC in terms of OS24, APF12, ORR, DoR, DCR, and PFS2 in patients with PD-L1-High UC	OS24 APF12, ORR, DoR, and DCR using investigator assessments according to RECIST 1.1 [a] PFS2 as defined by local standard clinical practice
To further assess the efficacy of MEDI4736 monotherapy compared to SoC in terms of OS24, APF12, ORR, DoR, DCR, and PFS2 in patients with UC	OS24 APF12, ORR, DoR, and DCR using investigator assessments according to RECIST 1.1 [a] PFS2 as defined by local standard clinical practice
To further assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in terms of OS24, APF12, ORR, DoR, DCR, and PFS2 in patients with PD-L1-Low/Neg UC	OS24 APF12, ORR, DoR, and DCR using investigator assessments according to RECIST 1.1 [a] PFS2 as defined by local standard clinical practice
To further assess the efficacy of MEDI4736 + tremelimumab combination therapy compared to SoC in patients with PD-L1-High UC	PFS, APF12, ORR, DoR, and DCR using investigator assessments according to RECIST 1.1 [a] OS and OS24 PFS2 as defined by local standard clinical practice
To assess disease-related symptoms and HRQoL in UC patients treated with MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy compared with SoC and each other using the FACT-BL questionnaire	FACT-BL: Fatigue, Pain, Derived NFBlSI-18 score, FACT-BL TOI, and FACT-BL Total score

Secondary objectives:	Outcome measures:
To assess the PK of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy	Serum concentration of MEDI4736/tremelimumab and PK parameters (such as peak concentration and trough, as data allow; sparse sampling)
To investigate the immunogenicity of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy	Presence of ADAs for MEDI4736 and tremelimumab (confirmatory results: positive or negative)

a. The analysis will be based on programmatically derived investigator assessments according to RECIST 1.1. See Section 8 of the CSP for further details.

b. The analysis will be based on independently reviewed and adjudicated BICR assessments according to RECIST 1.1. See Section 8 of the CSP for further details.

ADA Anti-drug antibody; APF12 Proportion of patients alive and progression free at 12 months from randomization; BICR Blinded Independent Central Review; DCR Disease control rate; DoR Duration of response; FACT-BL Functional Assessment of Cancer Therapy - Bladder Cancer; FACT-BL TOI Functional Assessment of Cancer Therapy - Bladder Cancer Trial Outcome Index; HRQoL Health-related quality of life; NFBlSI-18 National Comprehensive Cancer Network - FACT Bladder Symptoms Index-18; ORR Objective response rate; OS Overall survival; OS24 Proportion of patients alive at 24 months from randomization; PD-L1 Programmed cell death ligand 1; PFS Progression-free survival; PFS2 Time from randomization to second progression; PK Pharmacokinetics; RECIST 1.1 Response Evaluation Criteria in Solid Tumors, version 1.1; SoC Standard of care; TTR time to response; UC Urothelial cancer.

1.1.3 Safety objective

Safety Objective:	Outcome Measures:
To assess the safety and tolerability profile of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy compared to SoC	Adverse events (AE), laboratory findings, vital signs and ECGs
To assess the safety and tolerability profile of MEDI4736 monotherapy and SoC in patients who are not cisplatin-eligible	AEs, laboratory findings, vital signs, and ECGs

AE Adverse event; ECG Electrocardiogram; SoC Standard of care.

1.1.4 Exploratory objectives

Exploratory objectives:	Outcome measures:
To explore irRECIST as an assessment methodology for clinical benefit of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy versus SoC with assessment by BICR	PFS, APF12, ORR, DoR, and DCR using BICR assessment according to irRECIST
To assess AEs directly by patient self-reporting of specific PRO-CTCAE symptoms	Sixteen PRO-CTCAE symptoms considered relevant to study treatments
To assess disease-related symptoms and HRQoL in UC patients treated with MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy versus SoC and each other using the FACT-BL questionnaire	FACT-BL: PWB, SWB, EWB, FWB, BlCS, and FACT-G Total score
To investigate the relationship between PK exposure and clinical outcomes, efficacy, AEs, and/or safety parameters, if deemed appropriate	A graphical and/or a data modeling approach will be used to analyze PK exposure and the relationship with clinical outcomes, efficacy, AEs, and/or safety parameters, as deemed appropriate
To describe and evaluate resource use associated with assigned treatments and underlying disease during assigned treatment	Health resource utilization measures including hospitalization, outpatient visits, or emergency department visits, measured via the HOSPAD module
To explore the impact of treatment and disease state on health state utility using the EQ-5D-5L during assigned treatment	EQ-5D-5L health state utility index will be used to derive health state utility based on patient-reported data
To collect blood and tissue samples for defining biological responses to MEDI4736 and tremelimumab and for identifying candidate markers that may correlate with likelihood of clinical benefit	Protein expression detected by IHC (eg, PD-L1) miRNA/mRNA T-cell and MDSC phenotyping SNP genotyping Urine-derived factors, where applicable
To assess patients' overall impression of the change in their health status since the start of study treatment	Single-item PGIC measure

Note: Exploratory objective analyses may be reported separately from the main clinical study report.

AE Adverse event; APF12 Proportion of patients alive and progression free at 12 months from randomization; BICR Blinded Independent Central Review; BlCS Bladder Cancer Subscale;

DCR Disease control rate; DoR Duration of response; EQ-5D-5L EuroQol 5-dimension, 5-level health state utility index; EWB Emotional well-being; FWB Functional well-being; FACT-BL Functional Assessment of Cancer Therapy - Bladder Cancer; FACT-G Functional Assessment of Cancer Therapy - General; FACT-BL TOI Functional Assessment of Cancer Therapy - Bladder Cancer Trial Outcome Index; HRQoL Health-related quality of life; IHC Immunohistochemistry irRECIST Immune-related Response Evaluation Criteria in Solid Tumors; MDSC Myeloid-derived suppressor cells; miRNA Micro-ribonucleic acid; mRNA Messenger ribonucleic acid; ORR Objective response rate; PD-L1 Programmed cell death ligand 1; PFS Progression-free survival; PGIC Patient Global Impression of Change; PK Pharmacokinetics; PRO-CTCAE Patient-reported outcomes version of the Common Terminology Criteria for Adverse Events; PWB Physical Well-Being; SNP Single nucleotide polymorphism; SoC Standard of care; SWB Social/family well-being; UC Urothelial cancer..

A further objective to meet China FDA requirement is to evaluate consistency in efficacy and safety among Chinese patients for benefit-risk assessment of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy compared to SoC. Details of China cohort and Asia population analysis will be specified in China supplementary SAP.

1.2 Study design

This is a randomized, open-label, controlled, multi-center, global Phase III study to determine the efficacy and safety of MEDI4736 monotherapy (1.5 g intravenous [IV] q4w) and MEDI4736 (1.5 g IV q4w) in combination with tremelimumab (75 mg IV q4w) for up to 4 doses/cycle each followed by MEDI4736 1.5 g IV q4w) versus SoC (cisplatin + gemcitabine or carboplatin + gemcitabine doublet) first-line chemotherapy in treatment-naïve patients with histologically or cytologically documented, unresectable, Stage IV (i.e., T4b, any N; or any T, N2-N3; or M1) transitional cell carcinoma (transitional cell and mixed transitional/non-transitional cell histologies) of the urothelium (including renal pelvis, ureters, urinary bladder, and urethra). Crossover from SoC to MEDI4736 monotherapy or MEDI4736 + tremelimumab combination therapy will not be permitted. A schematic diagram and flow chart of the overall study design is shown in Figure 1 and Figure 2, respectively.

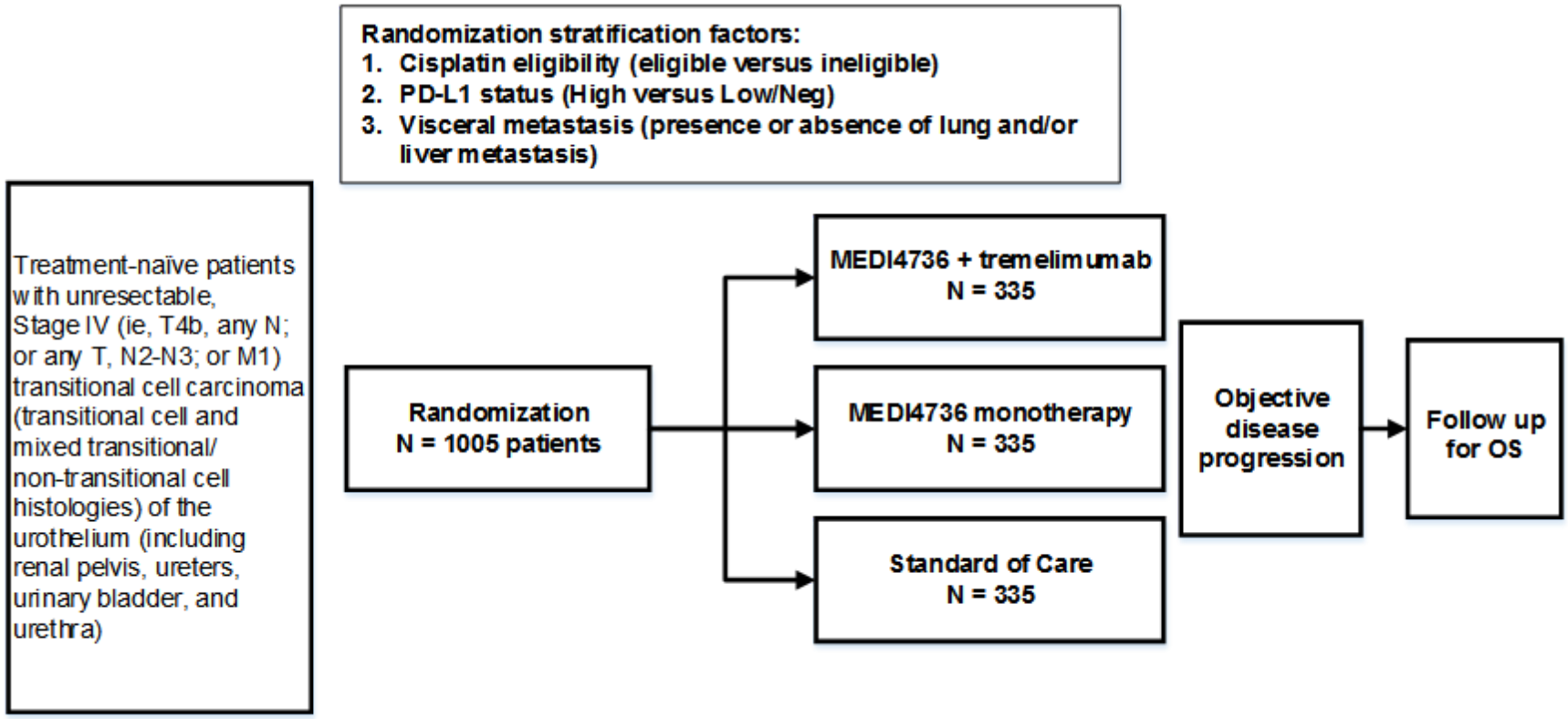
This study will randomize approximately 1005 patients globally. After the end of global enrollment (i.e. last patient in [LSI], globally), recruitment into an expansion cohort will continue in China until approximately 180 Chinese patients have been randomized. Closing of global cohort enrollment will be defined as closing recruitment across all sites except for those located in China. Identification of China cohort patients will be clearly defined in Section 2.3 and by distinct E-codes.

Patients will provide a tumor tissue sample at screening to determine PD-L1 status for stratification.

Patients enrolled in the study will be randomized (1:1:1) to treatment with MEDI4736 + tremelimumab combination therapy, MEDI4736 monotherapy, or SoC (cisplatin + gemcitabine or carboplatin + gemcitabine, based on cisplatin eligibility). Based on stratification factors from Interactive Voice Response System (IVRS), patients will be stratified according to cisplatin eligibility (eligible or ineligible; see Inclusion Criterion 7 in Section 3.1 of the clinical study protocol CSP), PD-L1 status (High or Low/Neg,using the VENTANA PD-L1(SP263) Assay), and visceral metastasis (presence or absence of lung and/or liver metastasis). Doses and treatment regimens are described in

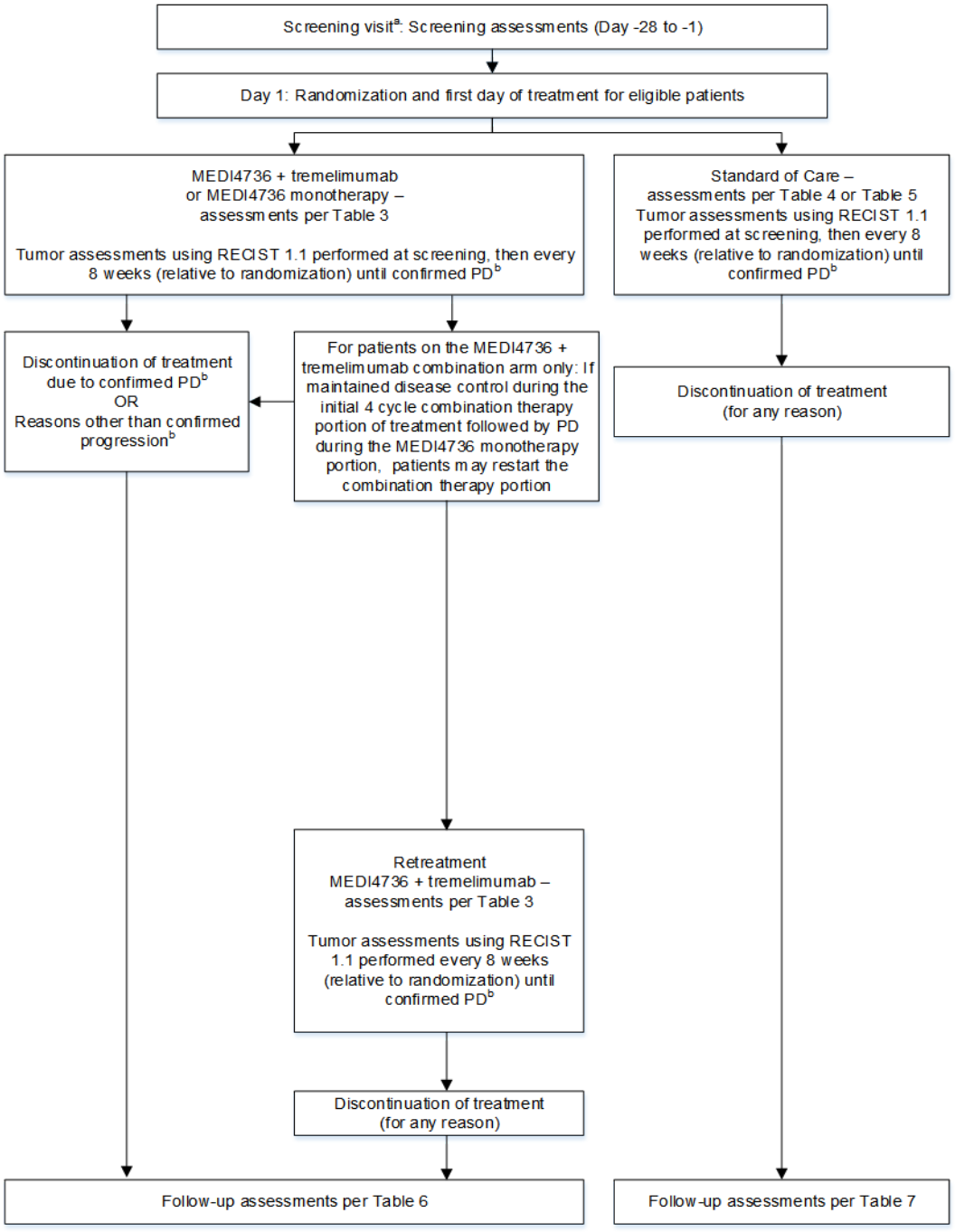
Sections 7.1 and 7.2 of the CSP. Assessments will be conducted as indicated in Table 3, Table 4, Table 5, Table 6, and Table 7 of the CSP.

Figure 1 Overall study design



OS Overall survival; PD-L1 Programmed cell death ligand 1.

Figure 2 Study flowchart



[a] Informed consent may be obtained prior to the 28-day screening window, if necessary, in order to permit tumor biopsy sample acquisition and analysis prior to randomization.

[b] In addition to PR and CR, a confirmatory scan is required following the initial demonstration of PD (patients enrolled in the SoC arm will discontinue study drug at the first assessment of disease progression). (See Section 5.1 of the CSP for more information.)

CR Complete response; PD Progressive disease; PR Partial response; RECIST 1.1 Response Evaluation Criteria in Solid Tumors, version 1.1; SoC Standard of care.

Independent Data Monitoring Committee (IDMC)

An IDMC will be established to perform an interim assessment of the safety and efficacy of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy in this population. The IDMC will be comprised of independent experts. The committee will meet approximately 6 months after the study has started or after the randomization of 30 patients, whichever happens first. The second IDMC meeting will occur approximately 3 months after the first IDMC meeting or when 90 patients are enrolled, whichever occurs first. A subsequent IDMC meeting will occur 3 months after 90 patients are enrolled. Further IDMC meetings will occur every 6 months, unless otherwise requested by the IDMC. IDMC members will be consulted to ensure appropriate frequency. Following each meeting, the IDMC will report to AstraZeneca and may recommend changes in the conduct of the study.

In addition to safety review meetings outlined above, two interim efficacy analyses of the DANUBE study will also be reviewed by the IDMC at:

- The first interim analysis (Interim 1) will focus on ORR and DoR in patients who are not cisplatin eligible, and treated with MEDI4736 monotherapy. Interim 1 will be conducted when all patients in the global cohort have at least 6 months follow-up.
- The second interim analysis (Interim 2) will focus on co-primary OS endpoints. The analysis will be conducted when approximately 80% of deaths occur in the UC patients across the MEDI4736 + tremelimumab combination therapy and SoC treatment arms AND in the PD-L1 High UC patients across the MEDI4736 monotherapy and SoC treatment arms.

Full details of the IDMC procedures, processes, and interim analyses can be found in the IDMC Charter.

1.3 Number of patients

The global study will plan to enroll approximately 1340 patients globally in order to randomize approximately 1005 patients in (1:1:1) ratio to MEDI4736 + tremelimumab combination therapy, MEDI4736 monotherapy, or SoC (cisplatin + gemcitabine or carboplatin + gemcitabine). Therefore, approximately 335 patients will be randomized to each of the treatment arms. Once global enrollment completes, recruitment to the expansion cohort will continue in China until approximately 180 Chinese patients have been randomized.

The global study is sized to characterize OS benefit of MEDI4736 in combination with tremelimumab versus SoC in patients with unresectable Stage IV UC and OS benefit of MEDI4736 versus SoC in patients with unresectable Stage IV PD-L1-High UC. The sample size calculation assumes a 6-month delay in separation of the OS curves between test arm and SoC arm, hence the use of average hazard ratios (HRs).

Non-uniform accrual of patients (with k=2) is assumed when estimating the analysis times. The total proportion of patients randomized at time t [t≤16 months] following the start of the study is assumed to be $(t/16)^2$.

The final analysis of OS will be performed when approximately 327 OS events (81%maturity) have occurred in PD-L1-High UC patients treated across the MEDI 4736 monotherapy and SoC treatment arms.

For the co-primary OS endpoint in the intent-to-treat (ITT) population (MEDI4736 + tremelimumab combination therapy versus SoC), 1 interim analysis will also be undertaken; and the interim analyses for OS endpoint in the PD-L1-High UC population (MEDI4736 monotherapy versus SoC) will be conducted at the same time.

- The interim analysis will be conducted when approximately 80% of final OS analysis events have occurred in the UC patients across the MEDI4736 + tremelimumab combination therapy and SoC treatment arms (440 events, 66% maturity) AND across the MEDI4736 monotherapy and SoC treatment arms in the PD-L1-High UC population (262 events, 65% maturity)

MEDI4736 + tremelimumab versus SoC (OS in all-comers UC)

The assumed OS treatment effect under the alternative hypothesis is an average HR of 0.73 for MEDI4736 + tremelimumab combination therapy versus SoC. The OS on the control arm is assumed to be characterized by an exponential distribution with a median OS of 11.3 months. The OS on the MEDI4736 + tremelimumab combination therapy treatment arm is therefore assumed to be characterized by a piece-wise exponential distribution with a median OS of 14.8 months. Assuming that the survival curves of the two treatment arms do not separate for 6 months then the HR after that point would need to be 0.61 to produce an average HR of 0.73 over the follow-up period.

The SoC control arm OS assumptions are based on a weighted average between median OS from gemcitabine/carboplatin (40% weighting) published in De Santis et al 2012 and gemcitabine/cisplatin published in Bellmunt et al 2012 (60% weighting), and it was calculated to be 0.4×9.3 months+0.6×12.7 months=11.3 months.

Final analysis of OS based on 550 events for the comparison of MEDI4736 + tremelimumab combination therapy versus SoC (82% maturity, 550/670), from all randomized patients, is expected to occur approximately 46 months after the first patient is randomized and will provide at least 87% power to demonstrate a statistically significant difference in OS at a 2-sided alpha level of 1.33% (with overall alpha for OS of 1.5%). With a minimum follow-up time of 30 months from the end of patient recruitment, this yields an anticipated overall average HR of 0.73, with a critical value for statistical significance of 0.81.

MEDI4736 versus SoC (OS in PD-L1-High UC)

It will be assumed that approximately 60% of patients will have PD-L1-High tumors.

The assumed OS treatment effect under the alternative hypothesis is an average HR of 0.71 for MEDI4736 monotherapy versus SoC, and the OS on the control arm is assumed to be the same regardless of PD-L1 status. The OS on the MEDI4736 monotherapy treatment arm is therefore assumed to be characterized by a piece-wise exponential distribution with a median OS of 15.3 months, and a 12-month OS rate of 56%. Assuming that the survival curves of the two treatment arms do not separate for 6 months then the HR after that point would need to be 0.57 to produce an average HR of 0.71 over the follow-up period.

By the time of the final analysis of OS of combination therapy versus OS in the ITT population, it is expected that there will be around 327 OS events in PD-L1-High UC patients across the MEDI4736 monotherapy and SoC arms, from approximately 402 PD-L1-High patients in total (81% maturity, 327/402), which will provide at least 84% power to

demonstrate a statistically significant difference in OS at a 2-sided alpha level of 3.03% (with overall alpha for OS of 3.5%). With a minimum follow-up time of 30 months from the last patient randomized, this yields an anticipated overall average HR of 0.71, with a critical value for statistical significance of 0.79.

MEDI4736 versus SoC (OS in all comers UC)

The analysis of OS for MEDI4736 monotherapy versus SoC in all UC patients is a key secondary endpoint. For illustrative purposes, it will be assumed that the significance level applied to the test will be a 2-sided 5% significance level, that is, assuming the analysis of both co-primary endpoints are significant at the 1.5% and 3.5% alpha level, respectively.

The assumed OS treatment effect under the alternative hypothesis is an average HR of 0.75 for MEDI4736 monotherapy versus SoC, and the median OS on the control arm is assumed to be 11.3 months. The OS on the MEDI4736 monotherapy treatment arm is therefore assumed to be characterized by a piece-wise exponential distribution with a median OS of 14.4 months. Assuming that the survival curves of the two treatment arms do not separate for 6 months, then the HR after that point would need to be 0.63 to produce an average HR of 0.75 over the follow-up period.

By the time of the final analysis of the co-primary endpoints of OS, it is expected that there will be around 553 OS events in the 670 UC patients (ITT population) across the MEDI4736 monotherapy and SoC arms, (83% maturity, 553/670), which will provide approximately 91% power to demonstrate a statistically significant difference in OS at a 2-sided 4.29% significant level (with overall alpha for OS of 5%). With a minimum follow-up time of 30 months from the end of patient recruitment, this yields an anticipated overall average HR of 0.75, with a critical value for statistical significance of 0.84.

MEDI4736 + tremelimumab versus SoC (OS in PD-L1-low/negative UC)

The analysis of OS for MEDI4736 + tremelimumab versus SoC in PD-L1 low/negative UC patients is a secondary endpoint in the 3rd level of MTP. For illustrative purposes, it will be assumed that the significance level applied to the test will be a 2-sided 5% significance level, that is, assuming the analysis of co-primary endpoints are significant at the 1.5% and 3.5% alpha level, respectively, and the key secondary endpoint, OS for MEDI4736 monotherapy versus SoC in all UC patients, is also significant at 5% level.

The assumed OS treatment effect under the alternative hypothesis is an average HR of 0.76 for MEDI4736 + tremelimumab versus SoC, and the median OS on the control arm is assumed to be 11.3 months. The OS on the MEDI4736 monotherapy treatment arm is therefore assumed to be characterized by a piece-wise exponential distribution with a median OS of 14.4 months. Assuming that the survival curves of the two treatment arms do not separate for 6 months, then the HR after that point would need to be 0.63 to produce an average HR of 0.75 over the follow-up period.

By the time of the final analysis of the co-primary endpoints of OS, it is expected that there will be around 222 OS events in the 268 PD-L1 low/negative UC patients across the MEDI4736 + tremelimumab and SoC arms, (83% maturity, 222/268), which will provide approximately 57% power to demonstrate a statistically significant difference in OS at a 2-sided 4.29% significant level (with overall alpha for OS of 5%). With a minimum follow-up

time of 30 months from the end of patient recruitment, this yields an anticipated overall average HR of 0.75, with a critical value for statistical significance of 0.77.

These power calculations are illustrative only and rely on the assumptions specified above. Any results presented from these tests will consider the impact of the percent prevalence of PD-L1-High tumors and the significance level applied on the power of the test when interpreting the results.

Table 1 provides a summary of the statistical assumptions for primary endpoints.

Table 1 Summary of statistical assumptions for primary endpoints

Endpoint	Analysis set	Events (number of patients) at FA	Alpha at FA (%)	Power at FA (%)	Alternative hypothesis (average HR)
OS (Final) MEDI4736 + tremelimumab versus SoC	FAS (ITT analysis set)	550 (670)	1.33	87	0.73
OS (Final) MEDI4736 mono versus SoC	PD-L1- High analysis set	327 (402)	3.03	84	0.71

OS Overall Survival; PD-L1 Programmed death ligand 1; FA: final analysis

2. ANALYSIS SETS

2.1 Definition of analysis sets

Full analysis set (Intention to treat (ITT))

The full analysis set (FAS) will include all randomized patients prior to the end of global recruitment. Any patients recruited in China, after global recruitment has ended, will not be included in the FAS (see Section 8.6 of the CSP). Unless otherwise specified, the FAS will be used for all efficacy analyses (including patient reported outcomes [PROs]). Treatment groups will be compared on the basis of randomized study treatment, regardless of the treatment actually received. Patients who were randomized but did not subsequently go on to receive study treatment are included in the analysis in the treatment group to which they were randomized.

PD-L1-High analysis set

The PD-L1-High analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1-High as defined by an Immunohistochemistry (IHC) assay developed by Ventana (see Table 2).

PD-L1-Low/Negative analysis set

The PD-L1-Low/Negative analysis set will include the subset of patients in the FAS whose PD-L1 status is PD-L1-Low/Negative as defined by an IHC assay developed by Ventana (see Table 2).

Table 2 PD-L1 status defined by scoring of an IHC assay developed by Ventana for stratification in D419BC00001*

Interpretation	Staining description
PD-L1 High	≥25% tumor cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control OR ≥25% tumor associated immune cell positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control
PD-L1 Low/Negative	<25% tumor cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control AND <25% tumor associated immune cell positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control

IHC Immunohistochemistry; PD-L1 Programmed cell death ligand 1.

*Definition of PD-L1 High versus low/negative expression will be used for stratification. Different cutoff of PDL1 expression may be utilized for analysis based on emerging data.

Note: PD-L1 High is defined as (1) If >=25% tumor cell membrane positivity for PD-L1; or (2) If IC area >1% and >=25% tumor associated immune cell positivity for PD-L1; or (3) If IC area=1% and 100% tumor associated immune cell positivity for PD-L1). PD-L1 Low/Neg if criteria not met for PD-L1 High.

Cisplatin ineligible analysis set

The cisplatin ineligible analysis set will include the subset of patients in the FAS who are not eligible for cisplatin treatment at baseline, per electronic case report form (eCRF).

MEDI4736 cisplatin ineligible population

All patients who have received MEDI4736 monotherapy and are not eligible for cisplatin treatment at baseline (per eCRF) in the Cisplatin ineligible analysis set will be included in this analysis set.

Safety analysis set

All patients recruited prior to the end of global recruitment who received at least 1 dose of investigational product (IP) will be included in the Safety analysis set. Any patients recruited in China, after global recruitment has ended, will not be included in the safety analysis set. Safety data will be summarized according to the treatment received, that is, erroneously

treated patients (e.g., those randomized to treatment A but actually given treatment B) will be summarized according to the treatment they actually received.

Cisplatin ineligible safety analysis set will include the subset of patients in the safety analysis set who are not eligible for cisplatin treatment at baseline (per eCRF) and have received either MEDI4736 monotherapy or SoC (carboplatin + gemcitabine).

bTMB analysis set

The bTMB analysis set includes patients with samples analyzed for TMB on circulating tumour DNA using the Guardant Health OMNI panel. Subgroups are defined as bTMB high ($\geqslant$ 24 mut/Mb), low (< 24 mut/Mb), bTMB-no call and missing results. Further subgroup analyses include TMB cut points at 16 and 20 mut/Mb.

Pharmacokinetic analysis set

All patients who received at least 1 dose of MEDI4736 or tremelimumab per the protocol and had at least one post-dose evaluable PK data of MEDI4736 or tremelimumab will be included in the PK analysis set. The analysis set will be defined by AstraZeneca/Medimmune, the Pharmacokineticist, and the Statistician prior to any analyses being performed.

Definitions of the analysis sets for each outcome variable are provided in Table 3

Table 3 Summary of outcome variables and analysis set

Outcome variable	Population
Efficacy data	
OS	Full analysis set (ITT population) PD-L1-High analysis set PD-L1-Low/Neg analysis set MEDI4736 cisplatin ineligible population (For cis-ineligible interim analysis [IA1] only) bTMB analysis set
PFS, OS24, APF12, ORR, DoR, DCR, PFS2, PROs, and symptom endpoints	Full analysis set (ITT population) bTMB analysis set PD-L1-High analysis set PD-L1-Low/Neg analysis set – ORR will be based on the subset of patients in each analysis set with measurable disease at baseline. – DoR will be based on the subset of patients in each analysis set which achieves objective tumor response. MEDI4736 cisplatin ineligible population (For cis-ineligible interim analysis [IA1] only) – ORR, DoR, TTR, DCR and PFS bTMB analysis set – PFS
Demography	Full analysis set (ITT population) Cisplatin ineligible safety analysis set (For cis-ineligible interim analysis [IA1] only)
PK data	PK analysis Set*
Safety Data	
Exposure	Safety analysis Set Cisplatin ineligible safety analysis set (For cis-ineligible interim analysis [IA1] only)
AEs	Safety analysis Set Cisplatin ineligible safety analysis set (For cis-ineligible interim analysis [IA1] only)
Laboratory measurements	Safety analysis Set Cisplatin ineligible safety analysis set (For cis-ineligible interim analysis [IA1] only)

Outcome variable	Population
ECOG performance status	Safety analysis Set Cisplatin ineligible safety analysis set (For cis-ineligible interim analysis [IA1] only)
Vital signs	Safety analysis Set Cisplatin ineligible safety analysis set (For cis-ineligible interim analysis [IA1] only)

* For cis-ineligible interim analysis [IA1] only, PK analysis will be restricted to a subset of patients in the PK analysis set who are not eligible for cisplatin treatment at baseline (per eCRF) and have received MEDI4736 monotherapy.

AE adverse event; APF12 proportion of patients alive and progression free at 12 months from randomization; DCR disease control rate; DoR duration of response; ECOG Eastern Cooperative Oncology Group; ITT Intent-to-treat; ORR objective response rate; OS overall survival; OS24 proportion of patients alive at 24 months from randomization; PD-L1 programmed cell death ligand 1; PFS progression-free survival; PFS2 time from randomization to second progression; PK pharmacokinetic; PRO patient-reported outcomes; TTR time to response.

2.2 Deviations

The following general categories will be considered important protocol deviations (IPD's) and be listed and discussed in the clinical study report (CSR).

- Deviation 1: Patients randomized but who did not receive study treatment.
- Deviation 2: Patients randomized who received treatment other than that to which treatment arm they were randomized to.

  Note: Patients randomized to the SoC arm will only be counted in this category if they received MEDI4736 + tremelimumab combination or MEDI4736 monotherapy treatment instead ie the type of SOC will not be considered.

- Deviation 3: Patients who deviate from key entry criteria as per the CSP. These are inclusion criteria 3, 4, 5 and exclusion criteria 3, 4, and 6.
- Deviation 4: Baseline RECIST scan > 42 days before date of randomization.
- Deviation 5: No baseline RECIST 1.1 assessment on or before date of randomization.
- Deviation 6: Received prohibited concomitant medications (including other anti-cancer agents). Please refer to the CSP section 7.7 for those medications that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to database lock to identify those likely have an impact on efficacy.

In addition to the programmatic determination of the deviations above, monitoring notes or summaries will be reviewed to determine any important post entry deviations that are not identifiable via programming, and to check that those identified via programming are correctly classified. The final classification of deviations will be made at the blinded data review

meeting (BDRM) prior to database lock or data freeze. Decisions made at the BDRM will be documented and approved by AstraZeneca prior to analysis.

None of the important deviations identified above will lead to exclusion from the FAS, PD-L1-High analysis set, PD-L1-Low/Negative analysis set or cisplatin ineligible analysis set. Deviation 1 will lead to exclusion from the Safety analysis set.

A per-protocol analysis excluding patients with important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis will be performed by excluding such patients if > 10% of patients on either treatment arm had one or more of these important deviations.

Errors in treatment dispensing, will be summarized and listed separately to the important protocol deviations. When a patient is not randomized or treated according to the randomization schedule it is envisaged that there will be 2 sub categories of this:

- Patients who receive no treatment whatsoever for a period of time due to errors in dispensing of medication. Note, this is not due to tolerability issues where patients may stop taking drug.
- The patient receives a treatment pack with a different code to their randomization code. However, the actual treatment may still match the randomized treatment. For example, a patient is given randomization code 0001, which according to the randomization schedule is MEDI4736. However, at the randomization visit they are given treatment pack 0003, but this still contains MEDI4736.

The summary will include all patients with a dispensing error but will also include information on how many of those patients received at least one dose of the wrong treatment at any time. Patients who receive the wrong treatment at any time will be included in the Safety analysis set as described in Section 2.1. During the study, decisions on how to handle mis-randomizations will be made on an individual basis with written instruction from the study team leader and/or statistician.

Errors in stratifications (based upon stratification information recorded in Interactive Voice Response System [IVRS] and eCRF), will also be summarized and/or listed separately to the important protocol deviations. The definition of cisplatin eligible/ineligible based upon eCRF information is detailed in Section 3.1 of the CSP. In addition, the location of visceral metastases (Lung, Liver, or Both) based on eCRF will be summarized and/or listed.

2.3 China cohort

China cohort consists of all patients from China sites and Taiwan sites accredited by China regulation recruited prior to the end of global recruitment and after the completion of global recruitment. The global cohort includes Chinese patients recruited prior to the end of global recruitment. Once global enrollment has completed, recruitment into an expansion cohort continues. Hence a patient randomized in China cohort prior to the end of global recruitment will be included in both the (globally recruited) FAS and the China FAS. A patient randomized in China after the end of global recruitment will be included only in the China FAS.

In addition to the evaluation of global cohort data for primary, secondary and safety objectives, evaluation of consistency in efficacy and safety in China and Asia population is

required to facilitate the benefit-risk assessment for Chinese patients. Hence, the safety and efficacy data in China cohort will be analyzed separately where the same endpoint definitions and the same analysis methods are applied.

The China full analysis set (China FAS) will include all patients randomized in China cohort and will be used for all China only efficacy analyses.

The China safety analysis set will consist of all patients recruited in China cohort who received at least 1 dose of study treatment.

Efficacy analyses for China cohort will be performed when the OS data from the China patients is of similar maturity where significant clinical efficacy is established in the global cohort, e.g. if OS efficacy is established at the second OS interim analysis, a similar maturity to this will be used for consistency evaluation.

All statistical analyses will be considered exploratory and only performed if sufficient numbers of events or patients are available (e.g. ≥20 OS events) unless specified, otherwise descriptive statistics only will be presented. No adjustment for multiplicity will be made and so the multiple testing procedure (MTP) testing detailed in Section 4.2.1 will not be followed. OS efficacy evaluation for China cohort will be performed once.

Details of China cohort and Asia population analysis, including vendor to perform the analysis, will be specified in China supplementary SAP, which is to be finalized before global cohort data lock for analysis.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST Visit Responses

RECIST 1.1 criteria will be used to assess patient response to treatment by determining PFS, proportion of patients alive and progression free at 12 months from randomization (APF12), ORR, DoR, TTR and DCR. The RECIST 1.1 guidelines for measurable, non-measurable, target, and non-target lesions and the objective tumor response criteria (complete response [CR], partial response [PR], stable disease [SD] or progressive disease [PD]) are presented in Appendix E in the CSP.

The methods of assessment of tumor burden used at baseline are computed tomography (CT) and magnetic resonance imaging (MRI) scans of the chest, abdomen, and pelvis. Any other areas of disease involvement should be additionally imaged based on the signs and symptoms of individual patients.

The baseline assessment should be performed no more than 28 days before randomization and ideally as close as possible to the start of the assigned IP. Efficacy for all patients will be assessed by objective tumor assessments every 8 weeks (relative to the date of randomization. If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits.

For patients who discontinue IP (including SoC) due to toxicity in the absence of confirmed objective progression, objective tumor assessments per the scheduled assessments should be continued every 8 weeks ± 7 days until confirmed objective disease progression.

In addition to PR and CR, a confirmatory scan is required following the initial demonstration of PD (exception is that patients enrolled in the SoC arm will discontinue study drug at the first assessment of disease progression). The confirmatory scan should occur preferably at the next scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinically significant deterioration. Treatment will continue between the initial assessment of progression and confirmation for progression for all patients randomized MEDI4736 + tremelimumab combination therapy or MEDI4736 monotherapy arms.

If a patient discontinues treatment (and/or receives a subsequent anticancer therapy) prior to progression, then the patient should still continue to be followed until confirmed objective disease progression.

Categorization of all objective tumor response assessments will be based on the RECIST 1.1 criteria of response: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Target lesion progression will be calculated in comparison to when the tumor burden was at a minimum (i.e., smallest sum of diameters previously recorded on study). In the absence of progression, tumor response (CR or PR) and SD will be calculated in comparison to the baseline tumor assessments obtained before randomization.

Objective tumor response (CR or PR) should be confirmed preferably at the next scheduled visit and preferably not less than 4 weeks after the visit when the response was first observed.

3.1.1 Investigator RECIST 1.1-based assessments: Target lesions (TLs)

At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, or PD depending on the status of their disease compared with baseline and previous assessments. Baseline should be assessed within the 28 days prior to randomization. If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of not evaluable (NE; unless there is evidence of progression in which case the response will be assigned as PD).

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (except lymph nodes which must have short axis ≥ 15 mm) with CT or MRI and which is suitable for accurate repeated measurements.

A patient can have a maximum of 5 measurable lesions recorded at baseline with a maximum of 2 lesions per organ (representative of all lesions involved suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). Lymph nodes are collectively considered as a single organ. If more than one baseline scan is recorded then measurements from the one that is closest and prior to the date of randomization will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

All other lesions (or sites of disease) not recorded as TL should be identified as non-target lesions (NTL) at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

Measurable disease (i.e., at least one TL) is one of the entry criteria for the study. However, if a patient with NTLs only is enrolled in the study, the evaluation of overall visit responses will

be based on the overall NTL assessment and the absence/presence of new lesions. If a patient does not have measurable disease at baseline then the TL visit response will be not applicable (NA).

Table 4 TL visit responses

Visit Reponses	Description
Complete Response (CR)	Disappearance of all TLs. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive disease (PD)	A $\geq$ 20% increase in the sum of diameters of TLs and an absolute increase of $\geq$ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD
Not Evaluable (NE)	Only relevant in certain situations (i.e. if any of the target lesions were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a target lesion response
Not applicable (NA)	No TLs are recorded at baseline

Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to 1 decimal place before assigning a TL response. For example 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%.

Missing TL data

For a visit to be evaluable, all TL measurements should be recorded. However, a visit response of PD should still be assigned if any of the following occurred:

- A new lesion is recorded.
- A NTL visit response of PD is recorded.

- The sum of TLs is sufficiently increased to result in a 20% increase, and an absolute increase of ≥ 5mm, from nadir even assuming the non-recorded TLs have disappeared.

Note: the nadir can only be taken from assessments where all the TLs had a lesion diameter recorded.

If there is at least one TL measurement missing and a visit response of PD cannot be assigned, the visit response is NE.

Lymph nodes

For lymph nodes, if the short axis diameter reduces to < 10mm then these are considered non-pathological. However a size will still be given and this size should still be used to determine the TL visit response as normal. In the special case where all lymph nodes are < 10mm and all other TLs are 0mm then although the sum may be >0mm the calculation of TL response should be over-written as a CR.

TL visit responses subsequent to CR

A CR can only be followed by CR, PD or NE. If a CR has occurred then the following rules at the subsequent visits must be applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node longest diameter (LD) increases by 20% but remains < 10mm.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response will be set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.
- Step 3: If not all lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD.
- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR.

TL too big to measure

If a TL becomes too big to measure this should be indicated in the database and a size ('x') above which it cannot be accurately measured should be recorded. If using a value of x in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team blinded to treatment assignment. It is expected that a visit response of PD will remain in the vast majority of cases.

TL too small to measure

If a TL becomes too small to measure a value of 5mm will be entered into the database and used in TL calculations, unless the radiologist has indicated and entered a smaller value that can be reliably measured. If a TL response of PD results then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e. lesion irradiated prior to entry into the study) should be recorded as NTLs and should not form part of the TL assessment.

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolisation), should be handled in the following way and once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD this will remain as a valid response category.
- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and if ≤ 1/3 of the TLs have missing measurements then scale up as described in the 'Scaling' section below. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.
- Step 3: If after both steps PD has not been assigned, then, if appropriate, a scaled sum of diameters will be calculated (as long as ≤ 1/3 of the TLs have missing measurements), treating the lesion with intervention as missing, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention also have a value of 0 recorded. If scaling up is not appropriate due to too few non-missing measurements then the visit response will be set as NE.

At subsequent visits the above steps will be repeated to determine the TL and overall visit response. When calculating the previous minimum, lesions with intervention should be treated as missing and scaled up where appropriate (as per step 2 above).

Scaling (applicable only for irradiated lesions/lesion intervention)

If > 1/3 of TL measurements are treated as missing (because of intervention) then TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (ie if using a value of 0 for missing lesions, the sum of diameters has still increased by 20% or more compared to nadir and the sum of TLs has increased by ≥5mm from nadir).

If ≤ 1/3 of the TL measurements are treated as missing (because of intervention) then the results will be scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters and this will be used in calculations; this is equivalent to comparing the visit sum of diameters of the non-missing lesions to the nadir sum of diameters excluding the lesions with missing measurements.

Table 5 Example of scaling

Lesion	Longest diameter at nadir visit	Longest diameter at follow-up visit
1	7.2	7.1
2	6.7	6.4
3	4.3	4.0
4	8.6	8.5
5	2.5	Intervention
Sum	29.3	26

Lesion 5 is missing at the follow-up visit.

The sum of lesions 1-4 at the follow-up is 26 cm. The sum of the corresponding lesions at nadir visit is 26.8 cm.

Scale up as follows to give an estimated TL sum of 28.4cm:

$$\frac{26}{26.8} \times 29.3 = 28.4\ cm$$

CR will not be allowed as a TL response for visits where there is missing data. Only PR, SD or PD (or NE) could be assigned as the TL visit response in these cases. However, for visits with ≤1/3 lesion assessments not recorded, the scaled up sum of TLs diameters will be included when defining the nadir value for the assessment of progression.

Lesions that split in two

If a TL splits in two, then the LDs of the split lesions should be summed and reported as the LD for the lesion that split.

Lesions that merge

If two TLs merge, then the LD of the merged lesion should be recorded for one of the TL sizes and the other TL size should be recorded as 0cm.

Change in method of assessment of TLs

CT, MRI and clinical examination are the only methods of assessment that can be used within a trial, with CT and MRI being the preferred methods and clinical examination only used in special cases. If a change in method of assessment occurs between CT and MRI this will be considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g. CT changes to clinical examination or vice versa), any affected lesions should be treated as missing.

3.1.2 Investigator RECIST 1.1-based assessments: Non-target lesions (NTLs) and new lesions

At each visit an overall assessment of the NTL response should be recorded by the investigator. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit.

NTL response will be derived based on the Investigator's overall assessment of NTLs as follows:

Table 6 NTL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
Progressive Disease (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Non-CR/Non-PD	Persistence of one or more NTLs with no evidence of progression.
Not Evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For patients without TLs at baseline, this is relevant if any of the NTLs were not assessed at this visit and the progression criteria have not been met.
Not Applicable (NA)	Only relevant if there are no NTLs at baseline

To achieve 'unequivocal progression' on the basis of NTLs, there must be an overall level of substantial worsening in non-target disease such that, even in the presence of SD or PR in TLs, the overall tumour burden has increased sufficiently to merit a determination of disease progression. A modest 'increase' in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more brand new lesions is assessed as progression.

A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

- The finding of a new lesion should be unequivocal: i.e. not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumour.
- New lesions will be identified via a Yes/No tick box. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses.
- A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.
- If the question 'Any new lesions observed at this specific visit that have not been previously recorded?' has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present, but should not overtly affect the derivation.

Symptomatic/clinical progression is not a descriptor for progression of NTLs: it is a reason for stopping study therapy and will not be included in any assessment of NTLs.

Patients with 'symptomatic/clinical progression' requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumour assessments where possible until objective disease progression is observed.

3.1.3 Investigator RECIST 1.1-based assessments: Overall visit response

Table 7 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 7 Overall visit responses

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
CR	CR or NA	No (or NE)	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	Non-PD or NE or NA	No (or NE)	PR
SD	Non-PD or NE or NA	No (or NE)	SD
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NE	Non-PD or NE or NA	No (or NE)	NE

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
NA	CR	No (or NE)	CR
NA	Non-CR/Non-PD	No (or NE)	SD
NA	NE	No (or NE)	NE

3.1.4 Blinded Independent Central Review (BICR) of RECIST 1.1-based assessments

The BICR of radiological images will be carried out using RECIST 1.1 criteria. Radiological scans (including those at unscheduled visits or outside visit windows) will be collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for quality checking (QC) and storage. For patients undergoing BICR, images will be reviewed by 2 primary independent radiologists using RECIST 1.1 and will be adjudicated, if required. The adjudicator must choose all of the assessments of one of the two primary reviewers. For each patient, the BICR will define the overall visit response data (CR, PR, SD, PD, No disease [ND] or not evaluable [NE]) and the relevant scan dates for each timepoint (i.e., for visits where response or progression is/is not identified). If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). Endpoints (of PFS, ORR, TTR, and DoR) will be derived from the scan dates contributing to the timepoint responses.

PFS by irRECIST criteria using BICR assessments may be performed for exploratory purposes. The original definitions of irCR, irPR, irSD/irNN, irPD, irNE and irND (ie responses according to immune-related Response Criteria, or irRC), were outlined by Wolchok et al 2009 using sums of cross-products from bi-dimensional (modified WHO criteria-based) TL diameters. In this project irRECIST using a RECIST base will be implemented where the target lesions will be measured unidimensionally (Nishino et al 2013).

In irRECIST the presence of new lesions will not automatically trigger a declaration of Progressive Disease, but instead the new lesions will be measured and these measurements will be added to the sum of diameters of the target lesions. Based on the sum of these measurements and % calculations thereof, the target lesion response assessment will be derived. The overall response assessment (irCR, irPR, irSD/irNN, irPD, irNE or irND) will be obtained at the BICR and confirmation of irPD is required.

A BICR of cis-ineligible patients only will be performed for the Interim 1 database lock, which will cover all of the scans up to the Data Cut Off (DCO). Further details of the BICR will be documented in the Independent Review Charter.

3.2 Outcome Variables

The analysis of the secondary endpoints of PFS, APF12, ORR, DoR and DCR on all analysis sets (excluding MEDI4736 cisplatin ineligible population) will be based on investigator

assessments according to RECIST 1.1. For analyses regarding the MEDI4736 cisplatin ineligible population, see Section 5.1.

3.2.1 Primary endpoints – overall survival

OS in patients with UC, and OS in patients with PD-L1-High UC are the co-primary endpoints.

OS is defined as the time from the date of randomization until death due to any cause (i.e., date of death or censoring – date of randomization + 1). Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Note: Survival calls will be made in the week following the date of DCO for the analysis, and if patients are confirmed to be alive or if the death date is post the DCO date these patients will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the interim and final OS analyses should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly-available death registries. In the event that the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly-available resources where it is possible to do so under applicable local laws.

3.2.2 Secondary endpoints

3.2.2.1 Progression free survival

Progression free survival (PFS) (per RECIST 1.1, as assessed by investigator or BICR) will be defined as the time from the date of randomization until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomized therapy or receives another anticancer therapy prior to progression. Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the patient progresses or dies after consecutive 2 or more missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the missed visits (Note: NE visit is not considered as missed visit). Given the scheduled visit assessment scheme (i.e. eight-weekly) the definition of 2 consecutive missed visits will equate to 18 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks).

If the patient misses all visits or does not have baseline data, the patient will be censored at Day 1 unless he/she dies within two visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

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

RECIST 1.1 assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- The date of progression will be determined on the earliest of the scan dates of the component that triggered the progression.
- For investigator assessments, the date of progression will be determined based on the earliest of the RECIST assessment/scan dates of the component that indicates progression.
- When censoring a patient for PFS, the patient will be censored at the latest of the dates contributing to a particular overall visit assessment.

Note: For target lesions, only the latest scan date is recorded out of all scans performed at that assessment for the target lesions, and similarly for non-target lesions, only the latest scan date is recorded out of all scans performed at that assessment for the non-target lesions.

3.2.2.2 Proportion of patients alive at 24 months

Alive at 24 months (OS24) will be defined as the Kaplan-Meier estimate of OS at 24 months.

3.2.2.3 Proportion of patients alive and progression-free at 12 months

Alive and progression-free at 12 months (APF12) will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed by investigator) at 12 months.

3.2.2.4 Objective response rate

Objective response rate ORR (per RECIST 1.1 as assessed by investigator) is defined as the number (%) of patients with at least 1 visit response of CR or PR. If any patients do not have measurable disease at baseline as measured by investigator then the analysis of ORR will exclude these patients, so that the denominator is a subset of the ITT analysis set who have measurable disease at baseline.

Data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue randomized treatment without progression, receive a subsequent anti-cancer therapy and then respond will not be included as responders in the ORR. Note that for this analysis palliative radiotherapy is not considered as a subsequent anti-cancer therapy, nor is switching to carboplatin after 1 or more doses of cisplatin for patients on the SOC arm stratified to cisplatin eligible per IVRS.

For the definition of ORR (per RECIST 1.1, as assessed by BICR) used in the Interim 1, see Section 5.1.2

3.2.2.5 Duration of response

Duration of response (DoR) (per RECIST 1.1 as assessed by investigator) will be defined as the time from the date of first documented response until the first date of documented progression or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the RECIST 1.1 PFS endpoint.

The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of CR or PR. If a patient does not progress following a response, then

their DoR will be censored at the PFS censoring time. DoR will not be defined for those patients who do not have documented response.

For the definition DoR (per RECIST 1.1, as assessed by BICR) used in the Interim 1, see Section 5.1.2

3.2.2.6 Disease control rate

Disease control rate (DCR) at 6 or 12 months is defined as the percentage of patients who have a best objective response (BoR) of CR or PR, or who have demonstrated SD for a minimum interval of 24 or 48 weeks, respectively (-7 days, i.e., 161 or 329 days, respectively), following the start of study treatment.

DCR will be determined programmatically based on RECIST 1.1 using investigator data and all data up until the first progression event. This will use all data up until the progression event that is used for the analysis.

For the definition of DCR (per RECIST 1.1, as assessed by BICR) used in the Interim 1, see Section 5.1.2

3.2.2.7 Change in tumor size

For supportive purposes, percentage change from baseline in tumour size will be derived at each scheduled tumour assessment visit (i.e., week 8, week 16 etc. hereafter referred to as week X for convenience). Best percentage change from baseline in tumour size will also be derived as the biggest decrease or the smallest increase in tumour size from baseline.

Tumor size is defined as the sum of the longest diameters of the target lesions for the investigator data based on RECIST 1.1. Target lesions are measurable tumor lesions. Baseline for RECIST 1.1 is defined to be the last evaluable assessment prior to starting treatment. The change in target lesion tumour size at week X will be obtained for each patient by taking the difference between the sum of the target lesions at week X and the sum of the target lesions at baseline. To obtain the percentage change in target lesion tumour size at week X the change in target lesion tumour size is divided by the sum of the target lesions at baseline and multiplied by 100 (i.e. (week X - baseline) / baseline * 100). More details on target lesions and measurements can be found in Section 3.1.

3.2.2.8 Time from randomization to second progression or death

Time from randomization to second progression or death (PFS2) will be defined as the time from the date of randomization to the earliest of the progression event subsequent to first subsequent therapy or death (ie, date of PFS2 event or censoring – date of randomization +1). The date of the first progression will be programmatically determined from investigator assessed data (See Section 3.2.2.1 for details). The date of second progression will be recorded by the Investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: objective radiological progression by RECIST, progression by disease specific biomarker, symptomatic progression, new or worsening of soft tissue/visceral or bone metastases, or death. RECIST assessment will not be collected for assessment of PFS2. The date of the PFS2 assessment and investigator opinion of progression status (progressed or non-progressed) at each assessment will be recorded in the eCRF.

Second progression status will be reviewed in line with scheduled follow-up (Table 6 and 7 in the CSP) following of the progression event used for PFS (the first progression) and status recorded.

The analysis of PFS2 should include all randomized patients. Patients alive and for whom a second disease progression has not been observed should be censored at the earliest of: date of study termination, date last known alive, DCO or, if a patient has not had a first subsequent therapy; the date last known not to have received a first subsequent therapy (FST-FS censoring date). Subsequent therapy does not include Durvalumab or Tremelimumab re-treatment.

However, if the patient experiences a second progression or dies immediately after two or more consecutive missed visits, the patient will be censored at the time of the last assessment prior to the two missed visits.

Best objective response

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in Appendix E in the CSP. It is the best response a patient has had following randomization during their time in the study up until RECIST progression or the last evaluable assessment in the absence of RECIST progression.

Categorization of BoR will be based on RECIST (Appendix E in the CSP) using the following response categories: CR, PR, SD, PD, and NE. In order to have SD as BoR, the duration of SD should be at least 8 weeks minus 1 week, ie, at least 49 days (to allow for an early assessment within the assessment window), after randomization.

Best objective response will be determined programmatically based on RECIST using all investigator data up until the first progression event. For patients whose progression event is death, BoR will be calculated based upon all evaluable RECIST assessments prior to death.

For patients who die with no evaluable RECIST assessments, if the death occurs ≤17 weeks (i.e., 16 weeks +7 days) after randomization, then BoR will be assigned to the progression (PD) category, if the death occurs >17 weeks (i.e., 16 weeks +7 days) after the date of randomization then BoR will be assigned to the NE category.

Progression events that have been censored due to them being ≥18 weeks after the last evaluable assessment will not contribute to the BoR derivation. Patients, who discontinue randomized treatment without progression, receive a subsequent anti-cancer therapy (note that for this analysis palliative radiotherapy is not considered a subsequent anti-cancer therapy), and then respond, will not be included in BoR calculation.

3.2.3 Exploratory Endpoints

3.2.3.1 First subsequent therapy free survival

As a supportive summary to PFS, first subsequent therapy free survival (FST-FS) is defined as the time from the date of randomization to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomized treatment, or death (i.e. date of first subsequent cancer therapy/death or censoring – date of randomization + 1). Palliative Radiotherapy is not considered as a subsequent anti-cancer therapy. Any patient not known to

have had a first subsequent anti-cancer therapy will be censored at the last date that the patient was known not to have received a first subsequent anti-cancer therapy (obtained from the CAPRX1 and CAPRXR1 form). If a patient terminated the study for reason other than death before first subsequent therapy, these patients will be censored at the earlier of their last known to be alive and termination date. Patients not receiving randomized treatment would have FST-FS calculated in the same way, i.e. time from date of randomization to the subsequent therapy.

3.2.3.2 Second subsequent therapy free survival

As a supportive summary to PFS, second subsequent therapy free survival (SST-FS) is defined as the time from the date of randomization to the earlier of start date of the second subsequent anti-cancer therapy after discontinuation of randomized treatment, or death (i.e. date of second subsequent cancer therapy/death or censoring – date of randomization + 1). Palliative Radiotherapy is not considered a subsequent anti-cancer therapy. Any patient not known to have had a second subsequent anti-cancer therapy will be censored at the last date that the patient was known not to have received a second subsequent anti-cancer therapy (obtained from the CAPRX1 and CAPRXR1 form). If a patient terminated the study for reason other than death before second subsequent therapy, these patients will be censored at the earlier of their last known to be alive and termination date. Patients not receiving randomized treatment would have SST-FS calculated in the same way, i.e. time from date of randomization to the second subsequent therapy or death.

3.3 Patient-reported outcome (PRO) variables

Patient reported outcome (PRO) questionnaires will be assessed using the FACT-BL questionnaire, PGIC, PRO-CTCAE, and EQ-5D-5L. All items/questionnaires will be scored according to published scoring guidelines or the developer's guidelines, if published guidelines are not available. All PRO analyses will be based on the FAS, unless otherwise stated.

3.3.1 FACT-BL

The FACT-BL is a disease-specific 39-item questionnaire included for the purpose of assessing health-related quality of life (HRQoL) and bladder cancer-specific symptoms. It is a well-established measure of HRQoL/health status commonly used in bladder cancer clinical studies. The FACT-BL was developed specifically for patients with advanced bladder cancer and has been found to be reliable and valid in this population (Cella et al 1993).

The FACT-BL consists of 5 subscales: Physical Well-Being (PWB; 7 items), Functional Well-Being (FWB; 7 items), Emotional Well-Being (EWB; 6 items), Social Well-Being (SWB; 7 items), and Additional Concerns or Bladder Cancer Subscale (BlCS) specific to bladder cancer (12 items). The BlCS assesses 5 domains: urinary function (3 items), bowel function (2 items), sexual function (2 items; 1 item is not applicable to women), body image (1 item), weight loss/appetite (2 items), and the care of ostomy appliance or urinary diversion (2 items) (Cella et al 1993).

All FACT-BL questions are scored on a 5-point Likert scale from 0 to 4 (0 being not at all and 4 being very much). Negatively stated items are reversed by subtracting the response from 4.

After reversing proper items (all 7 items in PWB, GE1, GE3, GE4, GE5, GE6, BL1, C2, BL2, C5, BL3, C8 and C9), all subscale items are summed to a total, which is the subscale score. For total scores, subscales, symptoms index, and individual item scores, the higher the score, the better the HRQoL/symptom. Thus, a score of 0 is a severely symptomatic patient, and the highest possible score is an asymptomatic patient.

The sum of the FACT-G subscales (PWB, FWB, EWB, and SWB) gives the FACT-G total score. All the 5 subscales (PWB, FWB, EWB, SWB, and BlCS) are summed as the FACT-BL total score, while the sum of PWB, FWB and BICS constitutes the FACT-BL TOI, which is an efficient summary index of physical/functional outcomes used as a PRO endpoint in clinical trials because it is responsive to change in physical/functional outcomes. NFBlSI-18 is based on the scores of 16 items (GP4, C2, BL1, GP3, GE6, GE1, C6, BL5, GF5, GP2, GP1, GP6, C3, GP5, GF3, GF7) available in the FACT-BL TOI (Jensen et al 2013). Because 2 items in NFBlSI-18, "I feel weak all overall" and "I feel light-headed (dizzy)", are not collected in FACT-BL, these 2 items will be considered as missing in NFBlSI-18 calculation. Fatigue will be based on the question of "I have a lack of energy" and pain will be based on the question of "I have pain", according to GP1 and GP4 in PWB, respectively.

In this study, the change from baseline in the following total/index scores will be evaluated as secondary endpoints: FACT-BL TOI (refer to as TOI), FACT-BL Total score, NFBlSI-18 score, fatigue and pain. The change from baseline in the individual subscales (PWB, FWB, EWB, SWB, and BlCS), and the FACT-G Total score may also be examined as exploratory analyses.

Scores for the FACT-BL will be derived using the developer instructions/manual (see Appendix F of the CSP).

Handling of missing data

If there are missing items, subscale scores can be prorated. This can be done on the scoring guide or by using the formula below:

Prorated subscale score=[Sum of item scores] × [N of items in subscale] ÷ [N of items answered]

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the items were answered (a minimum of 4 of 7 items, 4 of 6 items, etc.). The total score is then calculated as the sum of the unweighted subscale scores. The total score representing a FACT scale is acceptable as an indicator of patient quality of life as long as *overall item response rate* is greater than 80% (at least 22 of 27 FACT-G items completed; at least 32 of 39 FACT-BL items completed). Thus, FACT-BL total score and FACT-G total score should only be computed when the patient responds to greater than 80% of items in the scales [this requirement should not to be confused with individual subscale item response rate, which allows a subscale score to be prorated for missing items if greater than 50% of items are answered].

In addition, a total score should only be calculated if all of the component subscales have valid scores. For individual subscale item response rate, a subscale score is prorated for missing items if greater than 50% of items are answered. If at least 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing. The reason for any missing questionnaire will be identified and recorded. If there is evidence that

the missing data are systematic, missing values will be handled to ensure that any possible bias is minimized. For the "Additional Concerns" subscale (BlCS) and the symptoms index, the procedures for scoring are the same as described above for the FACT-G. Again, over 50% of the items (eg, 7 of 12 items) must be completed in order to consider each subscale score valid.

3.3.2 PGIC

The response options of the PGIC are scored as follows: Very Much Improved (+3), Much Improved (+2), Minimally Improved (+1), No Change (0), Minimally Worse (-1), Much Worse (-2) and Very Much Worse (-3). Data from the PGIC will be summarized using FAS.

3.3.3 PRO-CTCAE

Data from the PRO-CTCAE will be summarized using FAS. The number (%) of patients with each level of response for each PRO-CTCAE item at baseline and over time will be summarized.

3.3.4 Health state utility (EQ-5D-5L)

The EQ-5D is a standardised measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal.

Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care.

The EQ-5D-5L index comprises 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). For each dimension, respondents select which statement best describes their health on that day from a possible 5 options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems, and extreme problems). A unique EQ-5D health state is referred to by a 5-digit code allowing for a total of 3125 health states. For example, state 11111 indicates no problems on any of the 5 dimensions. These data will be converted into a weighted health state index by applying scores from EQ-5D value sets elicited from general population samples (the base case will be the United Kingdom valuation set, with other country value sets applied in scenario analyses). Where EQ-5D-5L values sets are not available, the EQ-5D-5L to EQ-5D-3L crosswalk will be applied (Oemar and Oppe 2013).

In addition to the descriptive system, respondents also assess their health on the day of assessment on a visual analogue scale, ranging from 0 (worst imaginable health) to 100 (best imaginable health). This score is reported separately. The evaluable analysis set will comprise the FAS (ITT analysis set).

3.3.5 PRO Compliance Rates

Summary measures of overall compliance and compliance over time will be derived for the FACT-BL questionnaire, PGIC, PRO-CTCAE, and EQ-5D-5L respectively. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.

- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For patients that have progressed, the date of progression plus 6 months (180 days) will be used to assess whether the patient is still under HRQoL follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.
- Evaluable questionnaire = questionnaire with at least one subscale score that can be determined.
- Completed questionnaire = questionnaire with all questions completed
- Overall patient compliance rate is defined for each randomized treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire (i.e. randomized patients) multiplied by 100.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire at the time point (as defined above), divided by number of patients still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires (per definition above), divided by the number of received questionnaires. Completion rate over time will be calculated separately for each visit, including baseline, as the number of completed questionnaires (per definition above), divided by the number of received questionnaires. Finally, patient disposition of PRO assessments over time will be computed cumulatively at each visit using tables and bar charts. Descriptive summaries for patient/form disposition will include patients expected to provide PRO assessments and patients unexpected to provide PRO assessment due to death, disease progression and other reasons respectively at each visit.

3.4 Safety

Safety and tolerability will be assessed in terms of adverse events (AEs) (including serious adverse events [SAEs]), deaths, laboratory data, vital signs, electrocardiograms (ECGs) and exposure). These will be collected for all patients.

Safety data from the patients on the immunotherapy agents (MEDI4736 or MEDI4736+tremelimumab) or SoC will be summarized in the main presentations (see Section 4.1). ‘On treatment’ will be defined as assessments between date of start dose and 90 days following last dose of the immunotherapy agents (i.e., the last dose of MEDI4736 or MEDI4736+tremelimumab) or the last dose of the Standard of Care agents; For ECGs, vital signs and thyroid tests, “On treatment” will be defined as assessments between date of start dose and 30 days following last dose of the study treatment. For the majority of AE summaries the period of time after the administration of subsequent therapy will not be considered ‘on treatment’ (see further Section 4.2.14).

The Safety analysis set will be used for reporting of safety data.

3.4.1 Adverse events (AEs)

AEs and SAEs for all treatment arms will be collected from the time the informed consent is signed through 90 days after the last dose of the last study treatment and including the follow-up period. The Medical Dictionary for Regulatory Activities (MedDRA) (using the latest or current MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 4.03). A treatment emergent adverse event (TEAE) is an AE with an onset date or a pre-existing AE worsening following the first dose of study treatment through to 90 days after the last dose of the study medication. For the MEDI4736+tremelimumab arm and SoC arm, in the event of the components being administered separately then date of first dose/last dose will be considered as the earliest/latest dosing date of either component.

Other significant adverse events

During the evaluation of the AE data, an AstraZeneca medically qualified expert will review the list of AEs that were not reported as SAEs and 'Discontinuation of Investigational Product due to Adverse Events' (DAEs). Based on the expert's judgment, significant AEs of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered other significant adverse events (OAEs) and reported as such in the CSR. A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs.

Examples of these are marked hematological and other laboratory abnormalities, and certain events that lead to intervention (other than those already classified as serious) or significant additional treatment.

AEs of special interest

Some clinical concepts (including some selected individual preferred terms) have been considered "AEs of special interest" (AESI) to the MEDI4736 program. AESIs for MEDI4736 ± tremelimumab combination therapy include, but are not limited to, events with a potential inflammatory or immune-mediated mechanism and that may require more frequent monitoring and/or interventions such as steroids, immunosuppressants, and/or hormone replacement therapy. These AESIs are being closely monitored in clinical studies with MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy. A listing of the preferred terms in each grouping will be provided prior to data base lock. The AESI grouping may include but is not limited to the following: Adrenal insufficiency, Diarrhoea, Colitis, Select hepatic events, Infusion related/ Hypersensitivity/Anaphylactic reactions, Pneumonitis, Hyperthyroidism, Hypophysitis, Hypothyroidism, Dermatitis, Rash, Select pancreatic events, Select renal events, Other rare events of a potential immune-mediated nature. New AESI categories may be added as appropriate per periodic safety review.

Immune-mediated Adverse Events (imAE)

imAE will be identified from both AEs of special interest (AESIs) and AEs of possible interest (AEPIs) based on programmatic rules that consider interventions involving systemic steroid therapy, immunosuppressant use, and/or endocrine therapy (which, in the case of AEPIs, occurs after first considering an Investigator's causality assessment and/or an

Investigator's designation of an event as immune-mediated). Further details are provided in an imAE Charter.

In addition, the Sponsor may perform medical review of those AESIs and classify them as imAEs or not imAEs via an independent manual adjudication process.

3.4.2 Treatment exposure

Exposure for the immunotherapy agents and SoC as follows:

Total (or intended) exposure of MEDI4736 or tremelimumab

- Total (or intended) exposure = the earliest of (last dose date where dose > 0 mg +27, death date or DCO) – first dose date + 1

Total (or intended) exposure for SoC

- Total (or intended) exposure for cisplatin in 28 day cycle = the earliest of (last dose date where dose > 0 mg +27, death date or DCO) – first dose date + 1
- Total (or intended) exposure for cisplatin in 21 day cycle = the earliest of (last dose date where dose > 0 mg +20, death date or DCO) – first dose date + 1
- Total (or intended) exposure for carboplatin in 21 day cycle = the earliest of (last dose date where dose > 0 mg +20, death date or DCO) – first dose date + 1
- For gemcitabine:

- If it is CxD1 (cycle x Day 1) in 21 day cycle, then total (or intended) exposure = the earliest of (last dose date where dose > 0 mg +6, death date or DCO) – first dose date + 1
- If it is CxD8 (cycle x Day 8) in 21 day cycle, then total (or intended) exposure = the earliest of (last dose date where dose > 0 mg +13, death date or DCO) – first dose date + 1
- If it is CxD1 or CxD8 in 28 day cycle, then total (or intended) exposure = the earliest of (last dose date where dose > 0 mg +6, death date or DCO) – first dose date + 1
- If it is CxD15(cycle x Day 15) in 28 day cycle, then total (or intended) exposure = the earliest of (last dose date where dose > 0 mg +13, death date or DCO) – first dose date + 1

Actual exposure of study treatment

- Actual exposure = intended exposure – total duration of dose interruptions, where intended exposure will be calculated as above.

Dose reductions are not permitted per Section 6.7 of the CSP for the immunotherapy agents (MEDI4736 or MEDI4736+tremelimumab). The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Exposure will also be measured by the number of cycles received. For all two choices of SoC regimen (cisplatin + gemcitabine or carboplatin + gemcitabine), a cycle corresponds one dose of cisplatin or carboplatin, and for each immunotherapy agent a cycle corresponds to one dose of MEDI4736. If a cycle is prolonged for any reason, this should still be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

Patients who permanently discontinue during a dose interruption: if a decision is made to permanently discontinue study treatment in-between cycles or during a cycle delay then the date of last administration of study medication recorded will be used in the programming.

3.4.3 Dose intensity

Dose intensity will be derived for the MEDI4736+tremelimumab and MEDI4736 monotherapy group only. Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to treatment discontinuation.

Relative dose intensity (RDI) will be defined as follows for MEDI4736 and tremelimumab:

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule. When accounting for the calculation of intended cumulative dose 3 days may be added to the date of last dose to reflect the protocol allowed window for dosing.

3.4.4 Laboratory data

Laboratory data will be collected throughout the study, from screening to the follow-up visits as described in the CSP. Blood and urine samples for determination of haematology and clinical chemistry will be collected as described in Section 5.2.1 of the CSP. For the definition of baseline and the derivation of post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 3.4.8 below will be used.

Change from baseline in haematology and clinical chemistry variables will be calculated for each post-dose visit on treatment. CTC grades will be defined at each visit according to the CTC grade criteria using local or project ranges as required, after conversion of lab result to corresponding preferred units. The following parameters have CTC grades defined for both high and low values: Potassium, Sodium, Magnesium, Glucose and Corrected calcium, so high and low CTC grades will be calculated for these analytes.

Corrected calcium will be derived during creation of the reporting database using the following formulas:

Corrected calcium (mmol/L) = Total calcium (mmol/L) + ([40 – albumin (G/L)] x 0.02)

Absolute values will be compared to the project reference range and classified as low (below range), normal (within range or on limits of range) and high (above range).

The maximum or minimum on-treatment value (depending on the direction of an adverse effect) will be defined for each laboratory parameter as the maximum (or minimum) post-dose value at any time.

Project reference ranges will be used throughout for reporting purposes. The denominator used in laboratory summaries of CTC grades will only include evaluable patients, in other words those who had sufficient data to have the possibility of an abnormality.

For example:

- If a CTCAE criterion involves a change from baseline, evaluable patients would have both a pre-dose and at least 1 post-dose value recorded

- If a CTCAE criterion does not consider changes from baseline, to be evaluable the patient need only have 1 post dose-value recorded.

3.4.5 ECGs

ECG data obtained up until the 30 days from date of last dose of study treatment will be used for reporting. For derivation of post baseline visit values considering visit window and to handle multiple records present in any visit window, derivation rules as described in Section 3.4.8 below will be used.

At each time point the Investigator's assessment of the ECG will be collected locally. The data from this review will be stored for analysis if necessary at the end of the study. If it is necessary to analyse this data then QTcF (Fridericia) will be calculated programmatically using the reported ECG values (RR and QT).

QTcF = QT/RR^ (1/3) where RR is in seconds

For triplicate ECGs, the mean of the three ECG assessments will be used to determine the value at that time point.

3.4.6 Vital signs

Vital signs data obtained up until the 30 days from date of last dose of study treatment will be used for reporting. Change from baseline in vital signs variables will be calculated for each post-dose visit on treatment. For derivation of post baseline visit values considering visit window and to handle multiple records, derivation rules as described in Section 3.4.8 below will be used.

The denominator in vital signs data should include only those patients with vital sign data in safety analysis set.

3.4.7 Concomitant medication

Any medications taken by the patient at any time between the date of the first dose (including the date of the first dose) of study treatment up to the date of last dose of study treatment + 90 days in the study will be considered as concomitant medication. Any medication that started prior to the first dose of the study treatment and ended after the first dose or is ongoing will be considered as both prior and concomitant medication.

Allowed and disallowed concomitant medications will be presented by ATC classification and generic term.

3.4.8 General considerations for safety assessments

Time windows will need defining for any presentations that summarise values by visit. The following conventions should also apply:

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarized. Inclusion within the time window should be based on the actual date and not the intended date of the visit.
- All unscheduled visit data should have the potential to be included in the summaries.

The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day.

For example, the visit windows for vital signs data for MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy (with 4 weeks between scheduled assessments) are:

Day 1, visit window 1 to 3

Day 29, visit window 4 to 42

Day 57, visit window 43 to 70

Day 85, visit window 71 to 98

Day 113, visit window 99 to 126

Day 141, visit window 127 to 154

Day 169, visit window 155 to 182

Day 197, visit window 183 to 210

Day 225, visit window 211 to 238

Day 253, visit window 239 to 266

Day 281, visit window 267 to 294

Day 309, visit window 295 to 322

Day 337, visit window 323 to 350

Note: Due to the differing assessment schedules the visit windows will be different for the different study treatments and endpoints.

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).
- Listings should display all values contributing to a time point for a patient.
- For visit based summaries:

  If there is more than one value per patient within a time window then the closest value to the scheduled visit date should be summarized, or the earlier in the event the values are equidistant from the nominal visit date. The listings will highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all post baseline values collected are used including those collected at unscheduled visits regardless of whether or not the value is closest to the scheduled visit date.

  To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group visit data should only be summarized if the number of observations is greater than the minimum of 20 and > 1/3 of patients dosed.

- For summaries at a patient level, all values should be included, regardless of whether they appear in a corresponding visit based summary, when deriving a patient level statistic such as a maximum.
- Initial treatment and re-treatment will be combined into one treatment period. The first dose date will be the earliest dosed date, and the last dose date will be the latest dosed date, regardless of whether it is in initial treatment or re-treatment period.
- Baseline will be defined as the last non-missing measurement prior to dosing with study treatment. For laboratory data, any assessments made on day 1 will be considered pre-dose. Alternatively, if two visits are equally eligible to assess patient status at baseline (e.g., screening and baseline assessments both on the same date prior to first dose with no washout or other intervention in the screening period), the average will be taken as a baseline value. For non-numeric laboratory tests where taking an average is not possible then the best value would be taken as baseline as this is the most conservative. In the scenario where there are two assessments on day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline. Where safety data are summarized over time, study day will be calculated in relation to date of first treatment.

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e., below the lower limit of quantification) or > x (i.e., above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings.

3.5 Biomarker Variables

PD-L1 expression status (high, low/negative) is defined in Table 2. .

3.6 Pharmacokinetic and Immunogenicity variables

Analyses to evaluate the pharmacokinetics and immunogenicity of MEDI4736 and tremelimumab will be performed by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

3.6.1 Population pharmacokinetics and exposure-response/safety analysis

A population PK model may be developed using a non-linear mixed-effects modeling approach. The impact of physiologically-relevant patient characteristics (covariates) and disease on PK may be evaluated. The relationship between the PK exposure and the effect on safety and efficacy endpoints may be evaluated. The results of such an analysis, if conducted, will be reported separately from the main CSR, and therefore are not within the remit of the statistical analysis plan (SAP).

The PK, pharmacodynamic (PDx), demographic, safety, and efficacy data collected in this study may also be combined with similar data from other studies and explored using population PK and/or PK-PDx methods. Details of these analyses do not fall within the scope of this SAP.

3.6.2 Pharmacokinetic analysis

The PK analyses will be performed at AstraZeneca or appointed CRO. The actual sampling times will be used in the PK calculations. PK concentration data and summary statistics will be tabulated. PK parameters will be determined from raw data. The following PK parameters will be determined after the first and steady-state doses: peak and trough concentration (as data allow).

3.6.3 Immunogenicity analysis

Immunogenicity results will be analyzed descriptively by summarizing the number and percentage of patients who develop detectable ant-drug antibodies (ADAs) against MEDI4736 and tremelimumab. The immunogenicity titer and presence of neutralizing ADAs will be reported for samples confirmed positive for the presence of ADAs. Summaries will be based upon all patients from the safety analysis set.

The effect of immunogenicity on PK, pharmacodynamics, efficacy, and safety will be evaluated, if the data allow, but will be reported in a separate report and therefore are not within the remit of the SAP.

3.7 Health Resource Use

To investigate the impact of treatment and disease on health care resource, the following exploratory variables may be captured:

- Planned and unplanned hospital attendances beyond trial protocol mandated visits (including physician visits, emergency room visits, day cases and admissions)
- Primary sign or symptom for hospital/inpatient/emergency room visit
- Length of hospital stay
- Length of any time spent in an intensive care unit (ICU)

The length of hospital stay will be calculated as the difference between the date of hospital discharge (or death date) and the start date of hospitalisation or start of study drug if the start of study drug is after start date of hospitalisation (length of hospital stay = end date of hospitalisation – start date of hospitalisation + 1). Patients with missing discharge dates will be calculated as the difference between the last day with available data and the start date of hospitalisation. The length of ICU stay will be calculated using the same method.

4. ANALYSIS METHODS

In patients with UC, the following formal statistical analysis of OS (as co-primary endpoints) will be performed:

- H_0: No difference between MEDI4736 + tremelimumab combination therapy and SoC
- H_1: Difference between MEDI4736 + tremelimumab combination therapy and SoC

In patients with PD-L1-High UC, the following formal statistical analysis of OS (as co-primary endpoint) will be performed:

- H_0: No difference between MEDI4736 monotherapy and SoC
- H_1: Difference between MEDI4736 monotherapy and SoC

The study has been sized to characterize the OS benefit of MEDI4736 + tremelimumab combination therapy versus SoC in patients with UC, and the OS benefit of MEDI4736 monotherapy versus SoC in patients with PD-L1-High UC.

4.1 General principles

Descriptive statistics will be used for all variables, as appropriate, and will be presented by treatment arm. Continuous variables will be summarized by the number of observations, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized by frequency counts and percentages for each category. Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment arm. For log transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum.

Unless otherwise specified, efficacy and PRO data will be summarized and analyzed based on the FAS, PK data will be summarized and analyzed based on the PK analysis set, and safety data will be summarized on the safety analysis set.

All outputs will be summarized by treatment arm for all randomized patients (ITT) or Safety analysis set and where required, for all randomized patients within the PD-L High or PD-L1-Low/Negative subgroup.

For continuous data the mean and median will be rounded to 1 additional decimal place compared to the original data. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.

For PK data the geometric mean and CV will be presented to 4 significant figures (sf), minimum and maximum will be presented to 3 sf and n will be presented as an integer.

For categorical data, percentages will be rounded to 1 decimal place.

Post the DCO for final analysis of OS, data may be collected for a longer period with intent to analyze long-term OS and safety data (see section 7.8 of CSP). Any additional long-term analysis may be further clarified through an addendum to the SAP.

SAS® version 9.2 or above will be used for all analyses.

Baseline

In general, for efficacy and PRO endpoints the last observed measurement prior to randomization will be considered as the baseline measurement. However, if an evaluable assessment is only available after randomization but before the first dose of randomized treatment then this assessment will be used as baseline. Cisplatin eligibility status at baseline per eCRF will be based on CISELOM eCRF. For safety endpoints the last observation before the first dose of study treatment will be considered as the baseline measurement unless

otherwise specified. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose. Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The % change from baseline will be calculated as (post-baseline value - baseline value) / baseline value x 100.

4.2 Analysis methods

Results of all statistical analysis will be presented using a 95% confidence interval (CI) and 2-sided p-value, unless otherwise stated.

The following table (Table 8) details which endpoints are to be subjected to formal statistical analysis, together with pre-planned sensitivity analyses, making it clear which analysis is regarded as primary for that endpoint. Note, all endpoints will be performed in all randomized patients (ITT analysis set), unless otherwise indicated.

Table 8 Pre-planned statistical and sensitivity analyses to be conducted

Endpoints analyzed	Notes
Overall survival	Co-primary analysis using stratified log-rank tests: o MEDI4736 + tremelimumab combination therapy versus SoC (ITT population) o MEDI4736 monotherapy versus SoC (PD-L1 High population) Sensitivity analysis: o A Kaplan-Meier plot of time to censoring where the censoring indicator of the primary analysis is reversed – attrition bias o Stratified log-rank tests, using stratification factors (cisplatin eligibility status and visceral metastasis status) at baseline per eCRF Secondary analyses using stratified log-rank tests: o MEDI4736 monotherapy versus SoC (ITT population) o MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-Low/Negative population) o MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (PD-L1-Low/Negative population) o MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (ITT population) o MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-High population) o MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set for PD-L1 High population) o MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set) o MEDI4736 + tremelimumab combination therapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set) Secondary analyses of OS for IA1 only: o Median OS and 95% CI (MEDI4736 cisplatin ineligible population) Exploratory analysis of OS o HR and 95% CI of HR for MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (PD-L1 High population) o HR and 95% CI of HR for MEDI4736 monotherapy versus SoC (PD-L1-Low/Negative population) o Stratified Max-combo test based on adaptive procedure involving selection of best test statistics with log-rank (G0,0) and the Fleming-Harrington (FH) test o Exploratory analysis using bTMB cuts-off of 16 and 20 mut/mb

Endpoints analyzed	Notes
Progression free survival	Secondary analysis using stratified log-rank test (based on investigator data according to RECIST 1.1): o MEDI4736 + tremelimumab combination therapy versus SoC (ITT population) o MEDI4736 monotherapy versus SoC (PD-L1 High population) o MEDI4736 monotherapy versus SoC (ITT population) o MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-Low/Negative population) o MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (PD-L1-Low/Negative population) o MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (ITT population) o MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-High population) o MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set for PD-L1 High population) o MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set) o MEDI4736 + tremelimumab combination therapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set) Secondary analysis (based on BICR data according to RECIST 1.1) for IA1 only: o Median PFS and 95% CI (MEDI4736 cisplatin ineligible population) Exploratory analyses using stratified log-rank test (based on investigator data according to RECIST 1.1): o MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy versus SoC in patients with UC, PD-L1 high and PD-L1 low/negative by cisplatin eligibility status at baseline per eCRF. o Exploratory analysis using bTMB cuts-off of 16 and 20 mut/mb
Proportion of patients alive at 24 months	Kaplan-Meier estimates of survival at 24 months
Proportion of patients alive and progression-free at 12 months	Kaplan-Meier estimates of patients alive and progression-free at 12 months
Objective response rate	Logistic regression using investigator data (RECIST 1.1) ORR and 95% CI using BICR data (MEDI4736 cisplatin ineligible population) for IA1 only

Endpoints analyzed	Notes
Duration of response	Analysis following the method described by Section 4.2.3 using investigator data (RECIST 1.1) Analysis following method described by Section 5.1 using BICR data (MEDI4736 cisplatin ineligible population) for IA1 only
Disease Control Rate	Summarized by treatment arm n (%)
Time from randomization to second progression	Stratified log-rank test
Best Objective response	N (%) using investigator N (%) using BICR (MEDI4736 cisplatin ineligible population) for IA1 only
Change in tumor size	Waterfall plots of the best percentage change in tumor size by treatment arm
Change from baseline FACT-BL TOI, NFBlSI-18 score, FACT-BL Total score, FACT-BL subscales, FACT-G Total score	Average change from baseline using a Mixed Model Repeated Measurements (MMRM) analysis
Time to improvement in fatigue, and Time to deterioration in pain	Stratified log-rank test
EQ-5D-5L (health state utility values and Visual Analog Scale)	Summary statistics for health state utilities and visual analogue scale, including change from baseline.

BICR Blinded Independent Central Review; EQ-5D-5L EuroQol 5-dimension, 5-level health state utility index; FACT-BL Functional Assessment of Cancer Therapy - Bladder Cancer; FACT-G Functional Assessment of Cancer Therapy - General; FACT-BL TOI Functional Assessment of Cancer Therapy - Bladder Cancer Trial Outcome Index; HR Hazard ratio; ITT Intent-to-Treat; MMRM Mixed Model Repeated Measurements; PD-L1 Programmed cell death 1; RECIST Response Evaluation Criteria In Solid Tumors; SoC Standard of care.

4.2.1 Multiple testing strategy

In order to strongly control the Type I error at 5% 2-sided, a MTP with gatekeeping strategy will be used across the co-primary endpoints for OS (MEDI4736+tremelimumab versus SoC in ITT population, and MEDI4736 monotherapy versus SoC in PDL1 High population) and key secondary endpoints. If the higher level hypothesis in the MTP is rejected for superiority, the following hypotheses will then be tested as shown in Figure 3.

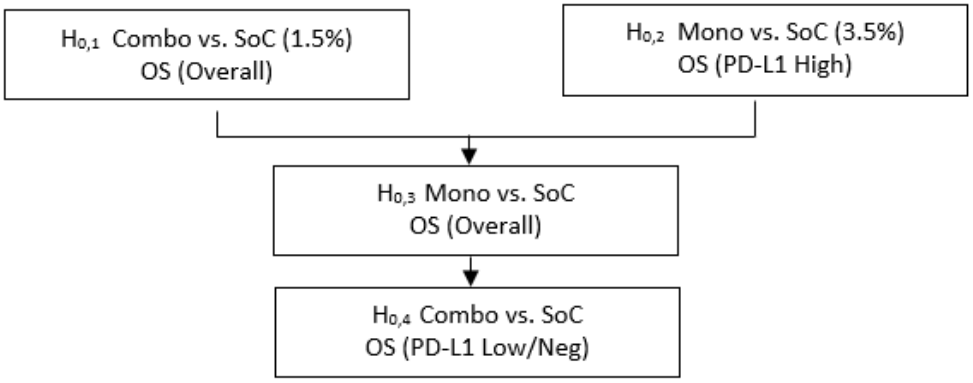
Hypotheses will be tested using a multiple testing procedure with an alpha-exhaustive recycling strategy (Burman et al 2009). With this approach, hypotheses will be tested in a

pre-defined order by first splitting the 5% alpha into 1.5%, and 3.5% for OS for MEDI4736 + tremelimumab combination therapy versus SoC (ITT population), and OS for MEDI4736 monotherapy versus SoC (PD-L1-High population), as outlined in Figure 3.

According to alpha (test mass) splitting and alpha recycling, the test mass that becomes available after each rejected hypothesis is recycled to secondary hypotheses not yet rejected. Since OS is tested at multiple timepoints (i.e., 1 interim analyses and final analysis), the OS tests for the same comparison/population (i.e., shown in 1 box in the MTP) will be considered as 1 test family. As long as 1 test in the family can be rejected, the family is rejected; thus, the assigned total alpha to the family can be recycled to the next MTP level. This testing procedure stops when the entire test mass is allocated to non-rejected hypotheses. Implementation of this pre-defined ordered testing procedure, including recycling, will strongly control type I error at 5% (2-sided), among all key hypotheses. Figure 3 shows the multiple testing framework for the co-primary endpoints and key secondary endpoints.

The details on how the alpha will be spent/controlled are outlined below:

Figure 3 Multiple testing procedures for controlling the type I error rate



Combo MEDI4736 + tremelimumab combination therapy; Mono MEDI4736 monotherapy; OS, overall survival; PD-L1, Programmed cell death ligand 1; SoC Standard of care.

1. Test $H_{0,1}$ and $H_{0,2}$ at level 1.5% and 3.5%, respectively.
   A. If neither of the 2 tests is statistically significant, accept $H_{0,1}$ and $H_{0,2}$, and stop procedure.
   B. If $H_{0,2}$ is not statistically significant at 3.5% level, but $H_{0,1}$ is statistically significant at 1.5% level, then accept $H_{0,2}$ and reject $H_{0,1}$, and continue testing $H_{0,3}$ at 1.5% level
      a) If $H_{0,3}$ is not statistically significant at 1.5% level, then accept $H_{0,3}$, and stop the procedure
      b) If $H_{0,3}$ is statistically significant at 1.5% level, then reject $H_{0,3}$ and continue test $H_{0,4}$ at 1.5% level

C. If $H_{0,1}$ is not statistically significant at 1.5% level, but $H_{0,2}$ is statistically significant at 3.5% level, then accept $H_{0,1}$ and reject $H_{0,2}$, and continue testing $H_{0,3}$ at 3.5% level

   a) If $H_{0,3}$ is not statistically significant at 3.5% level, then accept $H_{0,3}$, and stop the procedure

   b) If $H_{0,3}$ is statistically significant at 3.5% level, then reject $H_{0,3}$ and continue test $H_{0,4}$ at 3.5% level

D. If both $H_{0,1}$ and $H_{0,2}$ are statistically significant at 1.5% and 3.5% respectively, then reject $H_{0,1}$ and $H_{0,2,}$ and continue test $H_{0,3}$ at 5% level

   a) If $H_{0,3}$ is not statistically significant at 5% level, then accept $H_{0,3}$, and stop the procedure

   b) If $H_{0,3}$ is statistically significant at 5% level, then reject $H_{0,3}$ and continue test $H_{0,4}$ at 5% level

2. Test $H_{0,4}$ at alpha level as defined from Step 1 (1.5%, 3.5% or 5%)

   A. If $H_{0,4}$ is not statistically significant, accept $H_{0,4}$ and stop the procedure.

   B. If $H_{0,4}$ is statistically significant, reject $H_{0,4}$ and stop the procedure.

Both OS co-primary endpoints will be tested at 1 planned OS interim time-point and a final time-point. The alpha level allocated to OS will be controlled at the interim and primary timepoints by using the Lan DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the alpha level applied at the interim depends on the proportion of information available.

A separate Lan DeMets spending function that approximates an O'Brien Fleming approach will also be applied to each of the remaining endpoints in the MTP to enable testing at the OS interim and final timepoint (depending on the result for the co-primary endpoint). Adjustments to the alpha as a result of these interim analyses are discussed in Section 5.2. For the co-primary OS endpoint in the ITT population (MEDI4736 + tremelimumab combination therapy versus SoC), the OS interim analysis for superiority will occur when approximately 80% of the target OS analysis events have occurred across the MEDI4736 + tremelimumab combination therapy and SoC treatment arms. If exactly 80% of the target OS events are available at the time of the interim analysis of OS (i.e., 440/550 deaths have occurred), with an overall 2-sided alpha level of 1.5%, the 2-sided alpha to be applied at the OS interim analyses would be 0.56% . The 2-sided alpha to be applied for the final OS analysis would be 1.33%.

For the co-primary OS endpoint in the PD-L1-High population (MEDI4736 monotherapy versus SoC), the OS interim analysis for superiority will occur when approximately 80% of the target OS events have occurred across the MEDI4736 monotherapy and SoC treatment arms. If exactly 80% of the target OS events are available at the time of the interim analysis (i.e., 262/327 deaths have occurred), with an overall 2-sided alpha level of 3.5%, the 2-sided alpha to be applied at the OS interim analysis would be 1.58% .The 2-sided alpha to be applied for the final OS analysis would be 3.03%.

4.2.2 Primary endpoint - Overall survival

OS in the ITT population and PD-L1 high subgroup will be analyzed using a stratified log-rank test adjusting for cisplatin eligibility (i.e., eligible or ineligible), PD-L1 tumor expression (High versus Low/Negative, for OS in the ITT population) and visceral metastasis (presence or absence of lung and/or liver metastasis). The effect of MEDI4736 + tremelimumab versus SoC treatment in the ITT population will be estimated by the HR together with its corresponding 98.5% CI, 95% CI and p-value. The effect of MEDI4736 monotherapy versus SoC treatment in the PD-L1 High population will be estimated by the HR together with its corresponding 96.5% CI, 95% CI and p-value. The p-value will be based on stratified log-rank test.

The HR and its confidence interval will be estimated from a stratified Cox Proportional Hazards model *(*with ties = Breslow and the stratification variables included in the strata statement) and the CI calculated using a profile likelihood approach. The covariates in the statistical modeling will be based on the values entered into IVRS at randomization, even if it is subsequently discovered that these values were incorrect.

Summaries of the number and percentage of patients who have died, those still in survival follow-up, those lost to follow up, and those who have withdrawn consent will be provided along with the median OS for each treatment. Kaplan Meier plots of OS will be presented by treatment arm.

The assumption of proportionality will be assessed initially only with regard to the primary treatment comparisons. Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time-dependent covariate to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect will be described by presenting piecewise HR calculated over distinct time-periods. In such circumstances, the HR can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found, this may be a result of treatment-by-covariate interactions, which will be investigated.

Sensitivity Analyses

The following sensitivity analysis will only be performed for the primary comparison of OS in MEDI4736 + tremelimumab combination therapy versus SoC (Full analysis set) and OS in MEDI4736 monotherapy versus SoC (PD-L1 High analysis set) :

- Attrition bias will be assessed by producing a Kaplan-Meier plot of the time to censoring where the censoring indicator of the OS analysis is reversed.
- Stratified log-rank test. The stratification factors of visceral metastases and cisplatin eligibility status will be based on the baseline information per eCRF.
  Note: Sites are blinded to PD-L1 status. Hence, PD-L1 status at baseline is not reported on eCRF.

Exploratory analyses

Exploratory analyses of OS adjusting for the impact of subsequent immunotherapy or other systemic anti-cancer therapies may be performed, if a sufficient proportion of patients switch.

Methods such as Rank Preserving Structural Failure Time (Robins and Tsiatis 1991), Inverse Probability of Censoring Weighting (Robins 1993) and other methods in development may be explored. The decision to adjust and the final choice of methods will be based on a blinded review of the data and the plausibility of the underlying assumptions. Baseline and time-dependent characteristics may be explored, and summaries of baseline characteristics may be summarized by treatment arm, splitting between those that have and haven't switched at the time of the analyses. Further detail will be provided in the Payer Analysis Plan. These analyses are intended to support reimbursement appraisals.

Subsequent therapies received after discontinuation of treatment will be summarized and listed by treatment group. Patients who subsequently received an immunotherapy agent or entered an immunotherapy trial will be summarized and listed by treatment arm according to line of subsequent therapy, i.e. immediately after immunotherapy or as a later line.

Exploratory analyses of OS will be conducted for subjects in the 3i score low risk population. The 3i score (Immune Immediacy Index) low risk population consists of patients identified by the 3i score as having low risk of early mortality. The 3i score is based on six key routine laboratory parameters at baseline including NLR, neutrophils, albumin, lactate dehydrogenase, gamma glutamyltransferase and aspartate aminotransferase and tumor type. All six parameters are required in the model. The cut-off point of the 3i score is 0.649, subgroups are defined as 3i score low risk (≤0.649) and high risk (>0.649).

The stratified max-combo test will be conducted as an exploratory analysis on the OS data in the primary analysis set, to test for treatment differences in the case of nonproportional hazards. The analysis will be based on adaptive procedure involving selection of best test statistics with log-rank (G0,0) and the Fleming-Harrington (FH) test (G0,1, G1,0, and G1,1) with alpha correction (Duke-Margolis, 2018).

Subgroup Analyses

The following subgroup analyses (but not limited to) will be conducted to compare OS in MEDI4736 + tremelimumab versus SoC treatment (Full analysis set), and OS in MEDI4736 monotherapy versus SoC treatment (PD-L1 High analysis set):

- Sex (male versus female)
- Age at randomization (<65 versus ≥65 years of age)
    - This will be determined from the date of birth (BIRTHDAT in the DM module) and date of randomization (RND_DAT in the IE module) on the eCRF at screening, or AGE in DM module if AGE is available but BIRTHDAT is completely or partial missing; Patients with a partial date of birth (ie for those countries where year of birth only is given) will have an assumed date of birth of 1st Jan [given year]). Patients with a missing age value will be included using the mean age (overall FAS) and categorised accordingly.
- Visceral metastases at baseline per eCRF (presence versus absence of lung and/or liver metastasis) based on CISELOM CRF.
- Eligibility for cisplatin containing chemotherapy at baseline per eCRF (eligible or ineligible) (see the Appendix for details)

- Race (White versus non-White)
- Region (East Europe, West Europe , America, APEC) (see the Appendix for details)
- Smoking (Never, Former, Current)
- ECOG (0 versus >=1)
- Prior Adjuvant or Neo-adjuvant systemic chemotherapy (Yes versus No)
- Prior Bacillus Calmette-Guerin therapy (Yes versus No)
- Hemoglobin at baseline (<10 versus ≥10 g/dL)
- Primary tumor site (Upper tract [renal pelvis or Ureter] versus lower tract [bladder or urethra])
- Histology: (Transitional Cell Carcinoma versus Transitional Cell Carcinoma – Other [Transitional Cell Carcinoma With Squamous Differentiation, Transitional Cell Carcinoma With Glandular Differentiation, Transitional Cell Carcinoma With Variant Histology])
- Number of Bellmunt risk Factors (0, 1, 2 or 3)
    - The three risk factors are Hemoglobin < 10g/dL, ECOG ≥ 1, and the presence of liver metastasis based on CISELOM CRF).
- Number of Bajorn Risk factors (0 versus ≥ 1)
    - The two risk factors are: Eastern Cooperative Oncology Group (ECOG) score ≥ 2 (which corresponds to the Karnofsky Performance Status score of < 80 (ESMO 2008) and the presence of Bajorn -defined visceral metastasis (liver, lung or bone) at baseline based on DISEXT CRF.
- Metastatic site at baseline
    - Liver involvement (Yes versus No) based on CISELOM CRF.
- Lymph node only versus visceral metastasis (non-lymph node metastasis, including liver, lung, bone, soft tissue metastasis or others – a broader definition than protocol specified based on DISEXT CRF)

If there are too few patients in the certain categories of the subgroup, a combination of some categories may be applied. If there are too few death events available for a meaningful analysis of a particular subgroup (it is not considered as appropriate to present analyses where there are less than 20 events in a subgroup), the relationship between that subgroup and OS will not be formally analysed. In this case, only descriptive summaries will be provided.

For each subgroup, the HR (MEDI4736 + tremelimumab versus SoC in Full analysis set), HR (MEDI4736 monotherapy versus SoC in PD-L1 High analysis set) and the corresponding 95% CI will be calculated from a single model that contains treatment and the factor (only the factor that determines the subgroup). A forest plot, including the HR and 95% CI, will also be presented. No adjustment to the significance level for testing of the subgroup and sensitivity

analyses will be made since all these analyses will be considered as supportive of the analysis of OS.

Secondary Analysis

A secondary analysis of OS will be performed to compare the following:

- MEDI4736 monotherapy versus SoC (ITT population)
- MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-Low/Negative population)
- MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (PD-L1-Low/Negative population)
- MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (ITT population)
- MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1 High population)
- MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set for PD-L1 High population)
- MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set)
- MEDI4736 + tremelimumab combination therapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set)

This analysis will be performed using a stratified log-rank test adjusting for the same group of stratification factors as applicable. The HR, CI of HR, and p-value will be estimated using the same approach as specified above for the primary analysis of OS.

The significance level of OS in MEDI4736 monotherapy versus SoC (ITT population) and MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-Low/Negative population) will be determined by the MTP; a 5% significance level will be used for the remaining secondary analyses.

For OS analysis regarding MEDI4736 cisplatin ineligible population at IA1, see Section 5.1.3

Other cutoffs such as 16mut/MB and 20mut/MB will be used for bTMB exploratory analysis.

4.2.3 Progression-free survival

A secondary analysis of PFS based on the programmatically derived RECIST 1.1 using the investigator data will be performed to compare the following:

- MEDI4736 + tremelimumab combination therapy versus SoC (ITT population)
- MEDI4736 monotherapy versus SoC (ITT population)
- MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (ITT population

- MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-High population)
- MEDI4736 monotherapy versus SoC (PD-L1-High population)
- MEDI4736 + tremelimumab combination therapy versus SoC (PD-L1-Low/Negative population)
- MEDI4736 + tremelimumab combination therapy versus MEDI4736 monotherapy (PD-L1-Low/Negative population)
- MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set for PD-L1 High population)
- MEDI4736 monotherapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set)
- MEDI4736 + tremelimumab combination therapy versus SoC in bTMB high (≥ 24 mut/Mb) (bTMB analysis set)

These analyses will be performed using the same methodologies as described for the primary OS endpoints. 95% CI of HR will be estimated for these secondary analyses.

For PFS analysis regarding MEDI4736 cisplatin ineligible population at IA1, see Section 5.1.3

Kaplan-Meier plots of PFS will be presented by treatment arm. Summaries of the number and percentage of patients experiencing a PFS event and the type of event (RECIST 1.1 or death) will be provided along with median PFS for each treatment.

In addition, as a sensitivity analysis, patients who take subsequent therapy prior to progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy.

Other cutoffs such as 16mut/MB and 20mut/MB will be used for bTMB exploratory analysis.

The treatment status at progression of patients at the time of analysis will be summarized in the ITT population. This will include the number (%) of patients who were on treatment at the time of progression, the number (%) of patients who discontinued study treatment prior to progression, the number (%) of patients who have not progressed and were on treatment or discontinued treatment. This will also provide distribution of number of days prior to progression for the patients who have discontinued treatment.

Additional supportive summaries/graphs

In addition, in the ITT population, the number of patients prematurely censored will be summarized by treatment arm, and may also together with baseline prognostic factors, among the prematurely censored patients. A patient would be defined as prematurely censored if they had not progressed (or died in the absence of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval plus 2 weeks (10 weeks) prior to the DCO date.

Additionally, summary statistics will be given for the number of days from censoring to data cut-off for all censored patients, and for the number of weeks between the time of progression and the last evaluable RECIST assessment prior to progression for all patients with PFS event.

A summary of the duration of follow-up will be summarized using median time from randomization to date of censoring (date last known to have not progressed) in censored (not progressed) patients only, presented by treatment group.

Summaries of the number and percentage of patients who miss two or more consecutive RECIST assessments and the number of patients who miss one RECIST assessment will be presented for each treatment group.

All of the collected RECIST 1.1 data will be listed for all randomized patients. In addition, a summary of new lesions (i.e., sites of new lesions) will be produced.

Exploratory Analyses

If data allows, an exploratory analysis of PFS based on BICR assessments according to irRECIST criteria may be performed.

4.2.4 Overall survival at 24 months

Overall survival at 24 months (OS24) and the corresponding 95% CI will be presented (using the Kaplan-Meier technique) and presented by treatment arm. Note: 24 months equates to study day 731.

4.2.5 Alive and progression free at 12 months

The APF12 (where 12 months equates to study day 366) and the corresponding 95% CI will be summarized.

4.2.6 Objective response rate

The ORR will be based on the programmatically derived RECIST using the investigator data. The ORR will be compared between MEDI4736 + tremelimumab combination therapy versus SoC (ITT population) and MEDI4736 monotherapy versus SoC (PD-L1 high population) using logistic regression models adjusting for the same factors as the primary endpoint (cisplatin eligible status, PD-L1 tumor status and visceral metastasis). The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour MEDI4736+tremelimumab combination therapy or MEDI4736 monotherapy versus SoC) together with its associated profile likelihood CI (e.g. using the option 'LRCI' in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model).

Summaries will be produced that present the number and percentage of patients with a tumor response (CR/PR). Overall visit response data will be listed for all patients (i.e., the FAS). For each treatment arm, best objective response (BoR) will be summarized by n (%) for each category (CR, PR, SD, PD, and NE). No formal statistical analyses are planned for BoR.

For the analysis method of ORR and BoR used in the Interim 1, see Section 5.1.3

4.2.7 Duration of response

Descriptive data will be provided for the DoR in responding patients, including the associated Kaplan-Meier curves (without any formal comparison of treatment arms or p-value attached).

For the analysis method of DoR used in the Interim 1, see Section 5.1.3

4.2.8 Disease control rate

The DCR will be summarized (i.e., number of patients).

For the analysis method of DCR used in the Interim 1, see Section 5.1.3

4.2.9 Change in tumor size

Tumour size will also be presented graphically using waterfall plots for each treatment arm, to present each patient's best percentage change in tumour size as a separate bar, with the bars ordered from the largest increase to the largest decrease. Reference lines at the +20% and -30% change in tumour size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively. Additional waterfall plots showing percentage change in tumour size at specific time-points may be produced if it is felt that these are warranted to provide greater clarity.

The best percentage change will be presented in tables for each treatment arm in ITT analysis set.

The above outputs will be programmed for the investigator assessment according to RECIST 1.1.

For the analysis method of change in tumour size used in the Interim 1, see Section 5.1.3

4.2.10 Time from randomization to second progression or death (PFS2)

Time from randomization to second progression or death (PFS2) in the FAS will be analyzed using the same methodology as described in Section 4.2.3 and stratifying for the same covariates. Medians and 95% CI of PFS2 and Kaplan–Meier plots may be presented to support the analysis.

4.2.11 FST-FS and SST-FS

If data allows, for supportive purposes, the median and 95% CI of FST-TS and SST-FS will be estimated basing on Kaplan-Meier technique in FAS. No multiplicity adjustment will be applied as these are viewed as supportive endpoints

4.2.12 Healthcare resource use (HOSPAD)

Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) will be reported by treatment group, for planned and unplanned hospital attendances beyond trial protocol mandated visits, the length of hospital stay, and length of stay in ICU, as well as primary sign or symptom for hospital/inpatient/emergency room visit. To support submissions to payers, additional analyses may be undertaken and these will be outlined in a separate Payer Analysis Plan., which will be reported outside of the CSR.

4.2.13 Patient reported outcomes

4.2.13.1 FACT-BL

A separate family of endpoints is defined in order of importance (highest to lowest) for the secondary endpoints of NFBlSI-18 (a bladder symptoms index), FACT-BL TOI, FACT BL Total score, fatigue and pain. Descriptive statistics including line graphs of absolute value and change from baseline score for these secondary endpoints at each visit will be presented where appropriate.

An alpha of 5% is allocated to the statistical testing of these endpoints.

For NFBlSI-18, FACT-BL TOI,FACT Total score, FACT-BL subscales, FACT-G Total score, the mean change in score from baseline will be analyzed for each measure using a mixed model for repeated measures (MMRM) and estimation will be based on Restricted maximum likelihood method (REML). The model for change from baseline will include treatment, the 3 stratification factors (cisplatin eligibility, PD-L1 and visceral metastasis), visit and treatment by visit interaction as explanatory variables and the appropriate baseline FACT-BL value as a covariate. Treatment, the 3 stratification factors, visit and treatment by visit interaction will be fixed effects in the model. A random intercept term will also be included. The treatment by visit interaction will remain in the model regardless of significance. To determine the cut-off point for the MMRM analysis, if at least one of the following conditions is met at a visit, then that visit and the visits after will not be included in the MMRM model: (1) compliance rate at a visit is < 50% in any treatment arm; or (2) less than 20 patients at a visit in any treatment arm. An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom.

If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, and autoregressive.

The adjusted mean change from baseline estimates (obtained from LSMEANS-Statement as the calculated least square means adjusted for the random component of the model) and corresponding 95% CIs will be presented by visit for each treatment group. Appropriate method to estimate effect size may also be computed. Corresponding plots over time will be presented as applicable.

Improvement in fatigue is defined as at least 1 point improvement using GP1 of FACT-BL from baseline value, and is confirmed by repeat assessment not less than 14 days after the assessment when the improvement was first observed and with no deterioration (i.e. at least 1 point deterioration). Deterioration in pain is defined as at least 1 point deterioration in GP4 of FACT-BL from baseline value, and is confirmed by repeat assessment not less than 14 days after the assessment when the deterioration was first observed and with no improvement (i.e. at least 1 point improvement). If a patient has an initial deterioration in pain, then died within two scheduled pain assessments using FACT-BL, the initial deterioration of pain is considered as confirmed. Time to deterioration analysis may be repeated for functioning and HRQoL based on the following subscales and total scores as necessary: FACT-BL TOI, NFBlSI-18 score, FACT-BL Total score, FACT-G subscales (PWB, FWB, EWB, SWB), BlCS, FACT-G Total score. The threshold for deterioration will be defined as decrease in scores as listed: FACT-BL total (≥11), TOI (≥9), NFBlSI-18 (≥8), FACT-G total (≥8), FACT-G subscales (≥3) and BlCS (≥4). These thresholds are based on mUC patients (Degboe et al 2019)

As time to event endpoints, time to improvement in fatigue and time to deterioration /worsening in pain will be analyzed using the same methodologies described for OS analysis. The analysis will exclude patients whose baseline value is “Not at all” for fatigue, “Very much” for pain or they do not have any valid baseline evaluation. If patients did not have any confirmed improvement in fatigue or confirmed deterioration in pain, they will be censored at

the last FACT-BL assessment date. If a patient had missed at least 2 consecutive assessments prior to an initial improvement in fatigue or deterioration in pain being observed, the patient will be censored at the last evaluable assessment prior to the improvement/deterioration. assessment. The definition of 2 consecutive missed visits will equate to 18 weeks. Time to deterioration analysis maybe replicated for functioning and HRQoL based on FACT-BL total and subscale scores as appropriate.

Proportion of patients with confirmed improvement in fatigue and proportion of patients with confirmed deterioration in pain will be performed using the same methodologies as described for ORR in Section 4.2.6. Patients in ITT population will be excluded from estimation of proportion of patients with confirmed improvement in fatigue or deterioration in pain if their baseline value is "Not at all" for fatigue, "Very much" for pain or they don't have any valid baseline or post baseline assessment. The logistic regression analysis maybe replicated for functioning and HRQoL based on FACT-BL total and subscale scores as appropriate.

Clinically meaningful change in symptoms, functioning and HRQoL from baseline maybe further explored using a cumulative distribution function (CDF) and a probability density function (PDF) contingent on study results. These curves may be generated for the total and subscales scores (FACT-BL TOI, NFBlSI-18, FACT-BL Total, FACT-G total, PWB, FWB, EWB, SWB, and BlCS) separately at Week 8, Week 16, Week 32 and Week 48.

Exploratory descriptive item analysis based on responses at each visit as appropriate may be summarized for fatigue, pain and overall impact of treatment side effects (FACT-BL item GP5, "I am bothered by my side effects of treatment).

If missing data is substantial, multiple imputation approaches may be explored where appropriate.

4.2.13.2 PGIC

PGIC data will be presented using summaries and descriptive statistics based on the FAS.

4.2.13.3 PRO-CTCAE

Data from the PRO-CTCAE will be summarized using FAS. The number (%) of patients with each level of response for each PRO-CTCAE item at baseline and over time will be summarized. Further summaries to explore the data (i.e. the severity of symptoms) may be produced.

4.2.13.4 EQ-5D-5L

Descriptive statistics will be calculated for each scheduled visit/time point in the study, for each trial arm. These will report the number of patients, the number of EQ-5D questionnaires completed at each visit, the number and proportion responding to each dimension of the EQ-5D-5L. Additionally summary statistics (e.g. n, mean, median, SD, min, max) may be reported for the EQ-5D index score and the EQ-VAS score, and the change from baseline for the EQ-5D index score and the EQ-VAS score.

Graphical plots of the mean EQ-5D index score and EQ-VAS score, including change from baseline, by scheduled visits in the study may be produced. To support submissions to payers, additional analyses may be undertaken and these will be outlined in a separate Payer Analysis Plan, which will be reported outside of the CSR.

4.2.14 Safety data

Safety and tolerability data will be presented by treatment arm using the safety analysis set. Safety data will be summarized only. No formal statistical analyses will be performed on the safety data.

Data from all cycles of treatment will be combined in the presentation of safety data. AEs (both in terms of MedDRA preferred terms and CTCAE grade) will be listed individually by patient. The number of patients experiencing each AE will be summarized by treatment arm and CTCAE grade. Additionally, data presentations of the rate of AEs per person-years at risk will be produced.

Other safety data will be assessed in terms of clinical chemistry, hematology, vital signs, and ECGs. Exposure to MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy, and SoC (cisplatin + gemcitabine or carboplatin + gemcitabine) will be summarized. Dose delays/interruptions in MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy, and SoC will also be summarized.

The following sections describe the planned safety summaries. However, additional safety tables may be required to aid interpretation of the safety data.

Adverse Events

All AEs, both in terms of current MedDRA preferred term and CTCAE grade, will be summarized descriptively by count (n) and percentage (%) for each treatment group. The current MedDRA dictionary will be used for coding. The AE summaries, unless otherwise stated, will be based on treatment-emergent AEs. Any AE occurring before study treatment (i.e. before the administration of the first dose on Study Day 1) and without worsening after initial of study treatment will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as ‘pre-treatment’. However, any AE occurring before the administration of the first dose on Study Day 1 that increases in severity after the first dose will be regarded as treatment emergent and thus will be included in the majority of summary tables.

AEs observed up until 90 days following discontinuation of the immunotherapy agents (i.e., the last dose of MEDI4736 or MEDI4736+tremelimumab) and the Standard of Care agent or until the initiation of the first subsequent therapy following discontinuation of treatment (whichever occurs first) will be used for reporting of the AE summary tables. This will more accurately depict AEs attributable to study treatment only as a number of AEs up to 90 days following discontinuation of the immunotherapy agents and the Standard of Care agent are likely to be attributable to subsequent therapy.

However, to assess the longer term toxicity profile, some of the AE summaries may also be produced containing AEs observed up until 90 days following discontinuation of the immunotherapy agents and the Standard of Care agent (i.e. without taking subsequent therapy into account).

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved) and investigator’s assessment of severity and relationship to study drug. Frequencies and percentages of patients reporting each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from on the episode level summaries).

Summary information (the number and percent of patients by system organ class and preferred term separated by treatment group) will be tabulated for:

- All AEs
- All AEs possibly related to study medication (as determined by the reporting investigator)
- AEs with CTCAE grade 3 or 4
- AEs with CTCAE grade 3 or 4, possibly related to study medication (as determined by the reporting investigator)
- Most common AEs with CTCAE grade 3 or 4
- AEs with outcome of death
- AEs with outcome of death possibly related to study medication (as determined by the reporting investigator)
- All SAEs
- All SAEs possibly related to study medication (as determined by the reporting investigator)
- AEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication, possibly related to study medication (as determined by the reporting investigator)
- AEs leading to dose interruption of study medication
- Other significant AEs
- Other significant AEs possibly related to study medication (as determined by the reporting investigator)
- Immune mediated AEs (as determined by the reporting investigator)
- Infusion reaction AEs (as determined by the reporting investigator)
- Infection AEs

An overall summary of the number and percentage of patients in each category will be presented, as will an overall summary of the number of episodes in each category. In addition, a truncated AE table of most common AEs and another table showing most common AEs with CTCAE grade 3 or 4, showing all events that occur in at least 5% of patients overall will be summarized by preferred term, by decreasing frequency. This cut-off may be modified after review of the data. When applying a cut-off (i.e., x %), the raw percentage should be compared to the cut-off, no rounding should be applied first (i.e., an AE with frequency of 4.9% will not appear if a cut-off is 5%). Summary statistics showing the time to onset and the duration of the first AE will also be presented as appropriate.

Each AE event rate (per 100 patient years) will also be summarized by preferred term within each system organ class. For each preferred term, the event rate is defined as the number of

patients with at least 1 event divided by the total treatment duration (days) summed over patients and then multiplied by 365.25 x 100 to present in terms of per 100 patient years.

Summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, system organ class, preferred term and treatment group.

In addition, all AEs will be listed.

Deaths

A summary of all deaths and deaths on-treatment or within 90 days of last dose will be provided.

Adverse events of special interest (AESI)

Preferred terms used to identify adverse events of special interest based on most recent AESI preferred terms (currently including but not limited to Adrenal insufficiency, Diarrhoea, Colitis, Select hepatic events, Infusion related/ Hypersensitivity/Anaphylactic reactions, Pneumonitis, Hyperthyroidism, Hypophysitis, Hypothyroidism, Dermatitis, Rash, Select pancreatic events, Select renal events, Other rare events of a potential immune-mediated nature) will be listed before DBL and documented in the Study Master File. Grouped summary tables of certain MedDRA preferred terms will be produced.

AESI summaries by grouped term and preferred term for the safety analysis set to be provided are listed below.

- All AESI (Any Grade and CTCAE grade 3 or 4)
- AESI by Highest Severity
- Treatment related AESI (Any Grade and CTCAE grade 3 or 4)
- Treatment related AESI by Highest Severity
- Serious AESI
- AESI Resulting in Permanent Discontinuation
- AESI Resulting in Dose Delay
- AESI and Treatment related AESI Resulting in Death
- AESI by outcome

Immune-mediated Adverse events (imAEs)

The imAEs will be summarized in the same manner as for the summaries for AESI described above. The additional analyses include but not limited to, time to first onset imAE and resolution of imAE of Grade 3 or 4. See further details in the imAE Charter with respect to derivation rules.

In addition, the following analyses regarding systemic steroid use for imAE will be provided by imAE group. See further details in the imAE Charter with respect to derivation rules associated with duration and time to systemic steroid use for imAE. The following analysis except for duration of steroid use will be produced by imAE group for both all systemic steroids and high dose systemic steroids.

- Starting steroid dose
- Time to first steroid dose

Summary of long term tolerability

To assess long term tolerability, provided that there are a sufficient number of patients with events to warrant it, prevalence plots, life table plots and cumulative incidence plots may be presented for the most common AESI grouped terms as well as AEs of maximum CTCAE grade 3 or 4 and any other events considered important after review of the safety data, provided there are ≥ 10 events.

A prevalence plot provides information on the extent to which the events may be an ongoing burden to patients. The prevalence at time t after first dose of study treatment is calculated as the number of patients experiencing the event divided by the number of patients receiving study treatment or in safety follow-up at time t; generally, t is categorised by each day after dosing. The prevalence will be plotted over time and presented for each treatment group separately. Multiple occurrences of the same event are considered for each patient but a patient is only counted in the numerator whilst they are experiencing one of the occurrences of the event. These plots will only be produced for AESIs that have ≥10 events.

A life table plot may be used to describe the time to onset of the event and specifically when patients are at most risk of first experiencing the event. The hazard, or in other words, the probability of having an AE in a specified time period (e.g. 0-1 months, 1-3 months, 3-6 months, etc.) given that the patient reaches that time period without having an event is plotted for each time period. The hazard is calculated as the number of events per interval divided by total follow-up in the interval. These plots will only be produced for AESIs that have ≥10 events.

A cumulative incidence plot is a plot of the raw cumulative incidence and cumulative incidence function over time with the treatment groups presented on separate plots. The raw cumulative incidence is the actual probability that a patient will have experienced their first occurrence of the event by a given time point. The cumulative incidence function estimates the cumulative incidence if the data cut-off had not been imposed and all patients had completed safety follow-up (Pintilie M. 2006). These plots will only be produced for AESIs that have ≥10 events.

Laboratory assessments

Post baseline data obtained up until the 90 days following discontinuation of immunotherapy agents (i.e., the last dose of MEDI4736 or MEDI4736+tremelimumab) or following discontinuation of the Standard of Care agent or until the initiation of the first subsequent therapy following discontinuation of treatment (whichever occurs first) are considered as “on-study” and will be used for reporting. This will more accurately depict laboratory toxicities attributable to study treatment only as a number of toxicities up to 90 days following discontinuation of immunotherapy agents or the Standard of Care agent are likely to be attributable to subsequent therapy. For thyroid test, ‘on-study’ is defined as post baseline data obtained up until 30 days following discontinuation of the study treatment or until the

initiation of first subsequent therapy following discontinuation of treatment (whichever occurs first).

However, to assess the longer-term toxicity profile, summaries of laboratory data will also be produced containing data collected up until 90 days following discontinuation of the immunotherapy agents or Standard of Care agent (i.e., without taking subsequent therapy into account).

A small selection of summaries of laboratory data may also be produced containing data from initiation of the first subsequent therapy following discontinuation of study treatment until 90 days following discontinuation of immunotherapy agents or the Standard of Care agent (ie summarising the laboratory data collected on patients taking subsequent therapy during the safety collection follow-up window post discontinuation of study treatment). These outputs will only be produced if the number of laboratory toxicities observed warrant the inclusion of such outputs for interpretational purposes. Any data post 90 days last dose for immunotherapy agents or Standard of Care agents will not be summarized.

Data summaries will be provided in preferred units

Shift tables for laboratory values by worst CTC grade will be produced, and for specific parameters separate shift tables indicating hyper- and hypo- directionality of change from baseline will be produced. The laboratory parameters for which CTC grade shift outputs will be included but not limited are:

- Haematology: Haemoglobin, Leukocytes, Lymphocytes, absolute count-hypo and hyper, Neutrophils, absolute count, Platelets
- Clinical chemistry: ALT, AST, ALP, Total bilirubin, Albumin, Magnesium – hypo and – hyper, Sodium – hypo and – hyper, Potassium – hypo and – hyper, Corrected calcium – hypo and – hyper, Glucose –hypo and – hyper, Creatinine

For the parameters with no CTCAE grading that are listed in the CSP, shift tables from baseline to worst value on-treatment may be provided.

Liver Enzyme Elevations and Potential Hy's law

The following summaries will include the number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study

  ALT ≥ 3x –<= 5x, > 5x – <=8x , >8 x – <=10x, >10x – <=20x and > 20x Upper Limit of Normal (ULN) during the study

  AST ≥ 3x – <=5x, > 5x – <=8x , >8 x – <=10x, >10x – <=20x and > 20x ULN during the study

  Total bilirubin ≥2x – <= 3x, >3x– <= 5x and >5x ULN during the study

- Narratives will be provided in the CSR for patients who have ALT ≥ 3x ULN plus Total bilirubin ≥ 2x ULN or AST ≥ 3x ULN plus Total bilirubin ≥ 2x ULN at any visit.

Liver biochemistry test results over time for patients with elevated ALT or AST (ie ≥ 3x ULN), and elevated total bilirubin (ie ≥ 2x ULN) (at any time) will be plotted. Individual

patient data where ALT or AST (ie ≥ 3x ULN) plus total bilirubin (ie ≥ 2x ULN) are elevated at any time will be listed also.

Plots of ALT and AST vs. total bilirubin by treatment group will also be produced with reference lines at 3xULN for ALT, AST, and 2xULN for Total bilirubin. In each plot, Total bilirubin will be in the vertical axis.

ECGs

Post baseline ECG data obtained up until the safety follow-up are considered as "on-study". . 'On-study' is defined as post baseline data obtained up until 30 days following discontinuation of the study treatment or until the initiation of first subsequent therapy following discontinuation of treatment (whichever occurs first). Overall evaluation of ECG is collected at each visit in terms of normal or abnormal, and the relevance of the abnormality is termed as "clinically significant" or "not clinically significant". A listing of the ECG data will be produced.

Vital signs

Post baseline vital sign data obtained up until the safety follow-up are considered as "on-study" and will be included in the summary tables. 'On-study' is defined as post baseline data obtained up until 30 days following discontinuation of the study treatment or until the initiation of first subsequent therapy following discontinuation of treatment (whichever occurs first). Summaries of systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate, temperature, respiratory rate and weight will be presented.

Time to Subsequent Therapy from discontinuation of study treatment

Descriptive summaries will be produced for time to subsequent therapy from discontinuation of study treatment. These summaries are supportive of the adverse event and laboratory data outputs.

ECOG performance status

All Eastern Cooperative Oncology Group (ECOG) performance status will be summarized over time for the ITT analysis set.

4.2.15 Pharmacokinetic data

Summaries of pharmacokinetic concentration data will be provided for all evaluable patients in the PK analysis set.

4.2.16 Immunogenicity analysis

Summaries of immunogenicity data will be provided of the number and percentage of patients who develop detectable anti-MEDI4736 and anti-tremelimumab antibodies based on the safety analysis set. The immunogenicity titre and neutralizing ADA data will be listed for samples confirmed positive for the presence of anti-MEDI4736 antibodies and/or anti-tremelimumab antibodies.

The effect of immunogenicity on PK, PDx, efficacy and safety will be evaluated if data allows. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

4.2.17 PK/PDx relationships (MEDI4736 monotherapy and MEDI4736 + tremelimumab)

If the data are suitable, the relationship between PK exposure and efficacy/safety parameters may be investigated graphically or using an appropriate data modelling approach. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

4.2.18 Biomarker data

If applicable, the relationship of exploratory biomarkers, which may include but is not limited to tissue tumor mutation burden, to OS, PFS, ORR and DoR will be presented for a subset of patients in the ITT analysis set who are evaluable for each biomarker.

Summaries and analyses for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the CSR in a separate report. These outputs will be produced by AstraZeneca/MedImmune Biomarker group or designee.

4.2.19 Demographic and baseline characteristics data

The following will be summarized for all patients in the FAS (unless otherwise specified) by treatment group:

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations
- Inclusion in analysis sets
- Demographics (age, age group[<50, >=50-< 65, ≥ 65 - < 75 and >= 75 years], sex, race and ethnicity)
- Patient characteristics at baseline (height, weight, weight group)
- Patient recruitment by region, country and centre
- Previous disease-related treatment modalities
- Previous chemotherapy prior to this study
- Disease characteristics at baseline (ECOG performance status, primary tumour location, histology type, and overall disease classification)
- Extent of disease at baseline
- TNM classification at diagnosis
- Medical history (past and current)
- Relevant surgical history
- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer and radiation therapy

- Nicotine use, categorised (never, current, former)

The medications will be coded following AZ standard drug dictionary / WHO Drug dictionary as applicable.

4.2.20 Treatment exposure

The following summaries related to study treatment will be produced for the safety analysis set by actual treatment group:

- Total exposure of each treatment group.
- Actual exposure of each treatment group.
- Total number of cycles received for each treatment group.
- Number of, reasons for, and duration of dose delays/interruptions of MEDI4736, MEDI4736 plus tremelimumab and SoC. Dose interruptions will be based on investigator initiated dosing decisions. In addition, interruptions due to AEs and due to reasons other than AEs will be summarized separately.
- Number of infusions received.
- RDI (relative dose intensity) of MEDI4736 and tremelimumab.

For patients on study treatment at the time of the OS analysis, the DCO date will be used to calculate exposure.

4.2.21 Subsequent Therapy

Subsequent therapies received after discontinuation of study treatment will have summaries produced by treatment group.

5. INTERIM ANALYSES

Two interim analyses will be performed. The first interim analysis (Interim 1) will focus on ORR and DoR in patients who are not cisplatin eligible and who are treated with MEDI4736 monotherapy (see Section 5.1). The next interim analysis will focus on the co-primary OS endpoints (see Section 5.2).

5.1 Interim analysis focusing on ORR and DoR

This interim analysis (Interim 1) will focus on ORR in patients who are not eligible for cisplatin treatment at baseline and are treated in the MEDI4736 monotherapy arm. The data cut-off will take place at least 6 months after the last patient was randomized in the global cohort.

This interim analysis will be used for potential interactions with regulatory agencies regarding future development of MEDI4736 monotherapy in patients who are not eligible for cisplatin treatment. OS in the UC population and PD-L1 high UC population will remain the co-primary endpoints of this study. The evaluation of cis-ineligible patients in MEDI4736 monotherapy group will be considered as a secondary analysis outside of the MTP. Hence, there are no plans to stop the study early based on the interim results in the cis-ineligible population and thus no formal statistical adjustments are planned, and no alpha adjustment on

the primary analyses and secondary analyses in the MTP due to the evaluation of the cis-ineligible patients is planned.

In order to maintain the integrity of this ongoing study, the interim results will be reviewed by the IDMC against a pre-defined set of criteria. If the IDMC recommends that interim analysis 1 indicates a favorable benefit: risk profile of MEDI4736 monotherapy in cisplatin-ineligible patients, the interim results will be shared with a separate unblinded internal (AstraZeneca) committee, consisting of Sponsor personnel who are not involved in the DANUBE study conduct. The details of IDMC process and the interim analysis decision criteria will be detailed in the IDMC Charter/addendum. Finally, in order to further maintain DANUBE study trial conduct, the supportive analysis as well as preparation of the supplemental BLA submission will be performed by a separate team of individuals whom are not involved in the day to day conduct of the DANUBE study. Results will not be communicated externally from AstraZeneca, except with regulators for a potential discussion on registration in the cis-ineligible population. Additionally, no further amendments to the protocol will be allowed after the Interim 1 analysis.

Details of this interim analysis, such as analysis sets, endpoint definitions (if different from those used in Section 3) and analysis methods are specified in the following sections as needed.

5.1.1 Analysis dataset

Definition of MEDI4736 cisplatin ineligible population and cisplatin ineligible safety analysis set can be found in Section 2.1.

Sample size: in the case of 130 patients in the MEDI4736 cisplatin ineligible population, the maximum width between the observed ORR and its lower limit of the exact 95% CI will be no more than 9%.

5.1.2 Analysis endpoints

For this interim analysis, ORR (per RECIST 1.1 as assessed by the BICR) is defined as the number (%) of patients with a confirmed overall response of CR or PR and will be based on MEDI4736 cisplatin ineligible population. A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit. Therefore, data obtained up until progression, or the last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Any patient who discontinues treatment without progression, receives a subsequent therapy and then responds will not be included as responders in the ORR.

Duration of response (per RECIST 1.1 as assessed by the BICR) will be defined as the time from the date of first documented response (which is subsequently confirmed) until the first date of documented progression or death in the absence of disease progression. The end of response should coincide with the date used for the PFS endpoint (per RECIST 1.1 as assessed by BICR). DoR will not be defined for those patients who do not have a documented response.

The DCR (per RECIST 1.1 as assessed by the BICR) at 6 or 12 months is defined as the percentage of patients who have a BoR of CR or PR in the first 6 months or 12 months,

respectively, or who have demonstrated SD for a minimum interval of 24 or 48 weeks (-7 days, ie, 161 or 329 days, respectively), following the start of treatment.

Time to response (per RECIST1.1 by the BICR) is defined as the time from the date of randomization to the first documented response (which is subsequently confirmed). TTR will not be defined for those patients who do not have a documented response.

Progression free survival (per RECIST 1.1 as assessed by the BICR) and OS will also be evaluated as secondary endpoints. Change in tumour size (per RECIST 1.1 as assessed by the BICR) will also be presented.

ORR, DoR, DCR, TTR and PFS as well as change in tumour size will also be obtained using the algorithm described above for the RECIST1.1 site investigator tumor data.

PK and ADA will be assessed in patients who are in the MEDI4736 cisplatin ineligible population with the corresponding samples taken.

Safety will be evaluated in the Cisplatin ineligible safety analysis set

5.1.3 Analysis methods

ORR will be estimated with a 95% exact CI by Clopper-Pearson method. The primary analysis will be based on the programmatically derived ORR based on BICR assessments, and using all scans regardless of whether they were scheduled or not (see section 3.1.4). The primary analysis population for ORR will be the MEDI4736 cisplatin ineligible population.

A sensitivity analysis excluding patients who do not have measurable disease at baseline per BICR will be presented.

An analysis of ORR using the results of the programmatically derived RECIST site investigator tumor data from all scans will also be conducted as a sensitivity analysis to confirm the results of the primary analysis.

Summaries will be produced that present the number and percentage of patients with a tumor response (CR/PR). The number (%) of patients with a confirmed response and the number (%) of patients with a single visit response (ie, an unconfirmed response) will also be presented.

Kaplan Meier plots of DoR based on the BICR assessment of RECIST will be presented. Median DoR will also be summarized. Only patients who have a response will be included in this summary table. Duration of response will also be analyzed based upon the site investigator tumor data.

The DCR based upon the BICR assessment of RECIST will be summarized (ie, number of patients [%]). Disease control rate will also be summarized based upon the site investigator tumor data.

Descriptive summary statistics (ie, minimum, maximum, and median) will also be presented for TTR based on BICR assessments. TTR will also be summarised based upon the site investigator tumor data.

Kaplan-Meier plots of PFS (per BICR assessment) will be presented. Summaries of the number and percentage of patients experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided along with median PFS. The proportion of patients alive and

progression free at 3, 6, 9 and 12 months will be summarized (using the Kaplan-Meier curve). This analysis will be repeated for site investigator data.

Kaplan-Meier plots of OS will be presented. Summaries of the number and percentage of patients who have died, are still in survival follow-up, are lost to follow-up and have withdrawn consent will be provided along with median OS. The proportion of patients alive at 6, 9 and 12 months will be summarized (using the Kaplan-Meier curve).

Waterfall plots and summary table of best percentage change in tumour size (per BICR assessment) will be presented. This analysis will be repeated for site investigator data.

5.1.4 Subgroup analysis

In order to assess the consistency, subgroup analyses of ORR, DoR and safety will be conducted by the factors specified in Section 4.2.2 and Section 4.2.14 as appropriate, as well as by PD-L1 status. It is not considered as appropriate to present analyses where a subgroup has less than 20 events/responses combined for MEDI4736 monotherapy and SoC, and/or less than 5 events/responses in each treatment arm.

5.2 Interim analyses of overall survival endpoint

Both OS co-primary endpoints will be tested at Interim 2 and a final timepoint. The alpha level allocated to OS will be controlled at the interim and primary timepoints by using the Lan DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the alpha level applied at the interim depends on the proportion of information available.

The OS interim analysis will be conducted at the time when approximately 80% of the final OS analysis events have occurred across the MEDI4736 + tremelimumab combination therapy and SoC treatment arm (440 events, 66% maturity); AND across the MEDI4736 monotherapy and SoC treatment arm in PD-L1 High population (262 events, 65% maturity), projected approximately 16 months after the last patient being randomized.

The two key secondary endpoints that are in the lower levels of the MTP (MEDI4736 monotherapy versus SoC in terms of OS in patients with UC, MEDI4736 + tremelimumab compared to SoC in terms of OS in patients with PD-L1-low/negative UC) may also be tested at Interim 2 and the final analysis. The alpha level allocated to these tests will be based on O'Brien Fleming approach and depend on achieving a statistically significant results from the higher level tests in the MTP

Detail examples for alpha allocation for the tests in the MTP can be found in Appendix C.

6. CHANGES OF ANALYSIS FROM PROTOCOL

Analysis in Protocol	Analysis in SAP	Rationale of change
The OS24 and APF12 will be summarized (using the Kaplan-Meier curve) and presented by treatment arm. The HR and CI of HR will be also presented for OS24 and APF12	OS24 and APF12, and the corresponding CI will be estimated (using the Kaplan-Meier method) and presented by treatment arm	Kaplan-Meier estimate of OS24 and APF12 and the corresponding confidence interval are considered sufficient to characterize OS24 and APF12.
For EQ-5D-5L, average change from baseline using a MMRM analysis	Descriptive statistics, such as mean, median, SD, min, max values at each visit and change from baseline, will be calculated for each scheduled visit/time point in the study, for each trial arm.	Descriptive information at each visit/time point in the study is considered sufficient.

7. REFERENCES

Bellmunt et al 2012

Bellmunt J, von der Maase H, Mead GM, Skoneczna I, De Santis M, Daugaard G, et al. Randomized Phase III study comparing paclitaxel/cisplatin/gemcitabine and gemcitabine/cisplatin in patients with locally advanced or metastatic urothelial cancer without prior systemic therapy: EORTC Intergroup Study 30987. J Clin Oncol 2012;30:1107-13.

Berry et al 1991

Berry G, Kitchin RM, Mock PA. A comparison of 2 simple hazard ratio estimators based on the logrank test. Stat Med 1991;10:749-55.

Burman et al 2009

Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni based multiple tests. Stat Med 2009;28:739-61.

Cella et al 1993

Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993;11:570-9.

Collett 2003

Collett D. Modelling survival data in medical research: 2nd ed. Chapman and Hall/CRC; 2003.

De Santis et al 2012

De Santis M, Bellmunt J, Mead G, Kerst JM, Leahy M, Maroto P, et al. Randomized Phase II/III trial assessing gemcitabine/carboplatin and methotrexate/carboplatin/vinblastine in patients with advanced urothelial cancer who are unfit for cisplatin-based chemotherapy: EORTC Study 30986. J Clin Oncol 2012;30:191-9.

Duke Margolis Public Workshop 2018

Public Workshop: Oncology Clinical Trials in the Presence of Non‐Proportional Hazards, The Duke‐Margolis Center for Health Policy Feb. 2018.

Degboe et al 2019

Degboe A, Ivanescu C, Rohay J, Turner R, Cella D. Validity and performance of the Functional Assessment of Cancer Therapy-Bladder (FACT-Bl) among advanced urothelial cancer patients. Support Care Cancer (2019). https://doi.org/10.1007/s00520-019-04709-0

Ellis et al 2008

Ellis S, Carroll KJ, Pemberton K. Analysis of duration of response in oncology trials. Contemp Clin Trials 2008;29:456-65.

ESMO 2008

Performance scales: Karnofsky & ECOG scores. OncologyPRO - http://oncologypro.esmo.org/Guidelines-Practice/Practice-Tools/Performance-Scales

Fleischer et al 2011

Fleischer F, Gaschler-Markefski B, Bluhmki E. How is retrospective independent review influenced by investigator-introduced informative censoring: a quantitative approach. Stat Med 2011 Dec 20;30(29):3373-86.

Gail and Simon 1985

Gail M, Simon R. Test for qualitative interactions between treatment effects and patient subsets. Biometrics 1985;41:361–72.

Jensen et al 2013

Jensen SE, Beaumont JL, Jacobsen PB, Abernethy A, Syrjala KL, Cella D. Measuring priority symptoms in advanced bladder cancer: development and initial validation of a brief symptom index. J Support Oncol 2013;11:86–93.

Klein et al 2007

Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. Stat Med 2007;26:4505–19.

Lan and DeMets 1983

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

Nishino et al 2013

Nishino, M, Giobbie-Hurder, A., Gargano M, et al. "Developing a Common Language for Tumor Response to Immunotherapy: Immune-Related Response Criteria Using Unidimensional Measurements." Clinical Cancer Research 2013; 19:3936-3943

Oemar and Oppe 2013

Oemar M, Oppe M. EQ-5D-3L User Guide: Basic information on how to use the EQ-5D-3L instrument V5.0 (October 2013). Available from: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-3L_UserGuide_2013_v5.0_October_2013.pdf. Accessed 07 January 2014.

Pintilie M 2006

Competing risks: A practical perspective. Wiley.

Robins 1993

Robins JM. Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers. Proceedings of the Biopharmaceutical Section, American Statistical Association 1993; 24-33.

Robins and Tsiatis 1991

Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structure failure time models. Communications in Statistics 1991;20(8):2609-2631

Selke and Siegmund 1983

Selke T, Siegmund D. Sequential analysis of the proportional hazards model. Biometrika 1983;70:315-26.

Sun and Chen 2010

Sun X, Chen C. Comparison of Finkelstein's method with the conventional approach for interval-censored data analysis. Stat Biopharm Res 2010;2;97-108.

Whitehead and Whitehead 1991

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

Wolchok et al 2009

Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. Guidelines for the evaluation of immune therapy activity in solid tumours: immune-related response criteria. Clin Cancer Res 2009 Dec 1;15:7412-20.

8. APPENDIX

Appendix A. Region

Region	Country
East Europe	Poland, Russia, Israel, Greece, Turkey
West Europe	Austria, Belgium, Denmark, France, Germany, Italy, Netherlands, Portugal, Spain, United Kingdom, Australia
America	United States, Mexico, Canada, Brazil
APEC	China, Japan, South Korea, Taiwan

Appendix B. Cisplatin ineligible per eCRF

Cisplatin ineligible per eCRF is defined as meeting at least one of the following criteria:

- Creatinine clearance (CrCl) used for IVRS stratification from screening period and prior to randomization <60 mL/min (calculated by Cockcroft-Gault equation or by measured 24-hour urine collection)
- Common Terminology Criteria for Adverse Events (CTCAE) Grade ≥2 audiometric hearing loss
- CTCAE Grade ≥2 peripheral neuropathy
- New York Heart Association Class III or higher heart failure

Appendix C. Significant level (alpha) to the tests in the MTP

The alpha allocated at IA2 and Final analysis (FA) will depend on the actual number of death events observed at these timepoints. Hence, the alphas in the tables below are examples for different possible scenarios on the assumption that 80% of the target events occur at IA2.

- Alpha allocated to the 2nd level test in the MTP (Mono vs SoC in ITT) assume 80% target events occur at IA2

	Time point	Combo vs. SoC in ITT For the 1st level test in MTP	Mono vs. SoC in PD-L1 high For the 1st level test in MTP	Overall alpha recycled For the 2nd level test in MTP	Alpha[a] For the 2nd level test in MTP
Neither of primary endpoints is significant [b]					
1	IA2	Not sig (p≥0.56%)	Not sig(p≥1.58%)	0	0
	FA	Not sig(p≥1.33%)	Not sig(p≥3.03%)	0	0
Only mono vs. SoC in PD-L1 high is significant [b]					

	Time point	Combo vs. SoC in ITT For the 1st level test in MTP	Mono vs. SoC in PD-L1 high For the 1st level test in MTP	Overall alpha recycled For the 2nd level test in MTP	Alpha[a] For the 2nd level test in MTP
2	IA2	Not sig (p≥0.56%)	Sig(p<1.58%)	3.5%	1.58%
	FA	Not sig(p≥1.33%)	NA	3.5%	3.03%
3	IA2	Not sig (p≥0.56%)	Not sig(p≥1.58%)	0	0
	FA	Not sig(p≥1.33%)	Sig(p<3.03%)	3.5%	3.03%
Only combo vs. SoC in ITT is significant [b]					
4	IA2	Sig(p<0.56%)	Not sig(p≥1.58%)	1.5%	0.56%
	FA	NA	Not sig(p≥3.03%)	1.5%	1.33%
5	IA2	Not sig (p≥0.56%)	Not sig(p≥1.58%)	0	0
	FA	Sig (p<1.33%)	Not sig(p≥3.03%)	1.5%	1.33%
Both primary endpoints are significant [b]					
6	IA2	Sig(p<0.56%)	Sig(p<1.58%)	5%	2.44%
	FA	NA	NA	5%	4.29%
7	IA2	Not sig (p≥0.56%)	Not sig(p≥1.58%)	0	0
	FA	Sig (p<1.33%)	Sig(p<3.03%)	5%	4.29%
8	IA2	Not sig (p≥0.56%)	Sig(p<1.58%)	3.5%	1.58%
	FA	Sig (p<1.33%)	NA	5%	4.29%
9	IA2	Sig(p<0.56%)	Not sig(p≥1.58%)	1.5%	0.56%
	FA	NA	Sig(p<3.03%)	5%	4.29%

Not sig = not significant; Sig = significant

a. Assume 80% target events occur at IA2 .
b. Only alpha from primary endpoints known to be statistically significant at that particular timepoint will be recycled to the timepoint level 2 test.

- Level 3 of the MTP (Combo vs SoC in PD-L1 Low/Neg**)** will be handled in a similar way as level 2 in the MTP (Mono vs SoC in ITT).