AstraZeneca

Statistical Analysis Plan

Study Code	D933RC00001
Edition Number	6.0
Date	22 Mar 2024

A Phase III, Randomized, Open-Label, Multi-Center, Global Study to Determine the Efficacy and Safety of Durvalumab in Combination with Gemcitabine+Cisplatin for Neoadjuvant Treatment Followed by Durvalumab Alone for Adjuvant Treatment in Patients with Muscle-Invasive Bladder Cancer (NIAGARA)

A Phase III, Randomized, Open-Label, Multi-Center, Global Study to Determine the Efficacy and Safety of Durvalumab in Combination with Gemcitabine+Cisplatin for Neoadjuvant Treatment Followed by Durvalumab Alone for Adjuvant Treatment in Patients with Muscle-Invasive Bladder Cancer (NIAGARA)

PPD

PPD (AstraZeneca) Date

A Phase III, Randomized, Open-Label, Multi-Center, Global Study to Determine the Efficacy and Safety of Durvalumab in Combination with Gemcitabine+Cisplatin for Neoadjuvant Treatment Followed by Durvalumab Alone for Adjuvant Treatment in Patients with Muscle-Invasive Bladder Cancer (NIAGARA)

PPD

PPD (AstraZeneca)

Date

TABLE OF CONTENTS PAGE

LIST OF TABLES

TABLE OF FIGURES

LIST OF ABBREVIATIONS

The following is a list of abbreviations used in this document. Each individual SAP would contain its own set of abbreviations.

Abbreviation or special term	Explanation
ADAs	Antidrug antibodies
AE	Adverse event
AESI	Adverse event of special interest
AUA	American Urological Association
BDRM	Blinded data review meeting
BICR	Blinded independent central review
BMI	Body mass index
CI	Confidence interval
CIS	Carcinoma in situ
CR	Complete response
CrCl	Creatinine clearance
CRF	Case report form
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
ctDNA	Circulating tumor DNA
DBL	Database lock
DCO	Data cut-off
DFS	Disease-free survival
DSS	Disease-specific survival
EAU	European Association Urology
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EFS	Event-free survival
EFS24	Proportion of patients alive and event free at 24 months

EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer – Quality of Life Questionnaire – 30 Core
EQ-VAS	EuroQoL-visual analogue score
EQ-5D-5L	EuroQoL five dimensions, five level
FAS	Full Analysis Set
G+C	Gemcitabine plus cisplatin
HOSPAD	Hospital resource use module
HR	Hazard ratio
HRQoL	Health related quality of life
IC+	Immune cell with staining
ICP	Immune cell present
ICU	Intensive care unit
IDMC	Independent Data Monitoring Committee
IHC	Immunohistochemistry
imAE	Immune-mediated adverse event
IPD	Important protocol deviation
ITT	Intention-to-treat
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
KM	Kaplan-Meier
Ln	Natural logarithm or logarithm to the base e
MDSC	Myeloid derived suppressor cell
MedDRA	Medical Dictionary for Regulatory Activities
MFS	Metastasis-free survival
MIBC	Muscle-invasive bladder cancer
MMRM	Mixed model repeated measurement
MTP	Multiple testing procedure
NC	Not calculable
NCI	National Cancer Institute
NMIBC	Non-muscle-invasive bladder cancer
OAB	Overactive bladder
OAE	Other significant adverse events
OS	Overall survival
OS5	Proportion of patients alive at 5 years

pCR	Pathologic complete response
PD	Progressive disease
PD-L1	Programmed death ligand 1
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PID	Percentage intended dose
PH	Proportional hazards
PK	Pharmacokinetics
PR	Partial response
PRO	Patient reported outcome
PRO CTCAE	Patient-reported outcomes version of the CTCAE
RDI	Relative dose intensity
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS®	A commercially available integrated system of software products, commonly used for reporting and analysis of Clinical Studies
SD	Standard deviation
SD	Stable disease
SOC	System organ class
TEAE	Treatment emergent AE
TURBT	Transurethral resection of bladder tumor
ULN	Upper limit of normal
UTI	Urinary tract infection
WHO	World Health Organization

AMENDMENT HISTORY

Section # and Name	Date	Description of change	In line with CSP? Y (version) / N / NA	Rationale
Section 4.2.3.2 Dual primary endpoint – pCR: Subgroup Analysis Section 4.2.3.3 Dual primary endpoint – EFS: Subgroup Analysis	22Mar2024	Addition of subgroup analysis TC1 and TC25	NA	To assess the consistency of the treatment effect within subgroup TC1 and TC25
Section 4.2 Analysis Methods (Table 13) Section 4.2.3.3 Dual primary endpoint – EFS: Sensitivity Analysis	22Mar2024	Addition of sensitivity analysis for EFS	NA	To assess the impact of PD-L1 on efficacy.
Section 4.2.3.3 Dual primary Endpoint -EFS: Sensitivity Analysis	16Feb2024	Update the Interval censored analysis – evaluation time bias	Y (v7.0)	To align with TA-SAP V5.0
Section 3.5.1.1	16Feb2024	Removal of two missed visit rule from PRO analysis	Y (v7.0)	To align with CSP
Section 3.4.2 Dual primary Endpoint -EFS	16Feb2024	Removal of two missed visit rule from primary endpoint	Y (v7.0)	To align with CSP
Section 4.2 Analysis Methods (Table 13) Section 4.2.3 2 Dual primary Endpoint -EFS: Sensitivity Analysis	16Feb2024	Addition of sensitivity analysis Using a Kaplan-Meier plot of time to censoring where the censoring indicator of the EFS analysis is reversed to assess attrition bias Analysis where subjects who take subsequent anti-cancer therapy prior to the EFS event will be censored	Y (v7.0)	To align with CSP

		at their last evaluable assessment prior to taking the subsequent therapy Analysis using the 2 missed visit censoring rule		
Section 4.2.6 Safety	01Nov2023	Change the defination for post-surgery period, adjuvant period, and overall period	Y (v6.0)	To align with CSP
Section 3.4.2 Dual primary Endpoint -EFS	19Sep2023	Clarified that for the purpose of EFS, the date of is considered as disease assessments and NE is not considered a missed visit	Y (v6.0)	Clarify intent of CSP
Section 3.4.2 Dual primary Endpoint -EFS	19Sep2023	Added Table 6 and Table 7 Definition of 2 missed RECIST visits	NA	To align with TA SAP
Section 3.4.2 Dual primary Endpoint -EFS	19Sep2023	Added if a subject is known to have died where only a partial death date is available, then the date of death will be imputed as described in 3.4.3.3.”	NA	Clarify the death date where only a partial death date is available
Section 4.2.2 Subgroup analysis	19Sep2023	Updated ”Pathologic lymph node metastasis at cystectomy” to”Pathologic lymph node metastasis at baseline”	NA	Performed for FAS
Section 1.1 Table 1 Objectives and Endpoints, Section 3.6 Health Care Resource Use Variables, Section 4.2.6 Health Care Resource Use (HOSPAD)	30Jun 2023	Removal of calculation of healthcare resource use (HOSPAD) Delete Section 3.6, Section 4.2.6	Y (v6.0)	To align with CSP
Section 1.3 Number of Patients, Section 4.2.1 Multiplicity, Section 5 INTERIM ANALYSIS EFS	30Jun 2023	Changed number of events for final EFS analysis and also added calendar-based assessment timepoints for	Y (v6.0)	To align with CSP

		EFS IA2 and final analysis (FA). In accordance with these changes, the study power and critical value were also updated, as were the information fraction values at the interim analysis The calculation method of the timing of targeted events has been changed. A blinded event prediction has been used to provide a more accurate prediction		
Section 3.1 Derivation of RECIT visit responses	30Jun 2023	For adjuvant baseline assessments, remove "relative to the date of randomization"	Y (v6.0)	To align with CSP
Section 3.4.2 Dual primary Endpoint -EFS	30Jun 2023	Updated text to clarify EFS censoring rules 1. At the time of radical cystectomy patients will now be censored if the general 2 missed visit rule applies rather than considering only the scan immediately after cystectomy 2. Patients with no evaluable visits or baseline disease assessment prior to neoadjuvant treatment will not be censored at day 1 if they die within 112 days of randomization (rather than 2 visits)	Y (v6.0)	To align with CSP
Section 3.4.2 Dual primary Endpoint -EFS Section 4.1 Baseline	30Jun 2023	Added "If an adjuvant baseline scan is not recorded, it will be considered that no lesions	NA	Clarify the adjuvant baseline assessment

		are present following surgery (however, this will not count as a completed visit for the purposes of the 2 or more consecutive missed visit assessment)"		when the scan is missing
Section 4.2 Analysis Methods Table 11, Section 4.2.3 Sensitivity Analyses for Primary Endpoint	30Jun 2023	Added sensitivity analysis for missing adjuvant baseline scan using alternative censoring rules.	NA	To assess the impact of such data
Section 4.2.6 Safety	30Jun 2023	Added safety data could be presented by study period Defined the study period i.e., neoadjuvant period, post-surgery period, adjuvant period and overall period	NA	To enable assessment of the safety in the different periods
Section 5 INTERIM ANALYSES	30Jun 2023	Additional interim analysis of OS at EFS IA2 Predicted OS events at the time of EFS IA2 have been added and predicted OS events at the EFS final analysis have been revised	Y (v6.0)	To align with CSP
Section 4.2.1 Multiplicity, Section 5 INTERIM ANALYSES	30Jun 2023	Text has been added to state that the actual alpha level will be based on the observed number of events and as such the final analysis alpha will be derived using the generalized Haybittle-Peto method	Y (v6.0)	To align with CSP
Section 2.1.6 Table 2 Summary of Outcome Variables and Analysis Population Section 3.4.3.9 Pathological downstaging (pDS) rate Section 4.2 Table 11 Formal Statistical Analyses to be	8Aug 2023	Removed pDS	Y (v6.0)	To align with CSP pDS is same with Proportion of patients who achieve <P2

Conducted and Pre-Planned Sensitivity Analysis Section 4.2.4.9 pDS rate				
Section 1.1 Table 1 Objectives and Endpoints Section 1.3 Number of patients Section 2.1 Definition of Analysis Sets and Table for Summary of Outcomes and Analysis Population Section 3.4 Efficacy Variable Section 4.2 Analysis Methods	25Oct 2022	Updated study objectives and study population (Primary analysis for pCR and EFS will be performed on the ITT population instead of patients in the adequate renal function cohort) Reference to the adequate renal function population has been removed from all the secondary objectives	Y (v5.0)	To align with CSP
Section 1.2 Study Design	25Oct 2022	Add text to clarify "Noncystectomy extension phase"	Y (v6.0)	To align with CSP
Section 2.1 Definition of Analysis Sets	25Oct 2022	Updated text to clarify Cystectomy population. Defined PD-L1 high analysis set	Y (v5.0)	To align with CSP
Section 3.1.1.1.2 Baseline	25Oct 2022	Added text to clarify that if all TL measurements are missing, then the TL visit response is NE	NA	Clarify the TL visit response when all TL measures are missing
Section 3.4 Efficacy Variable Section 5 Interim Analysis	25Oct 2022	Regarding pCR and EFS endpoints, updated "coprimary" endpoint(s) language to "dual primary" endpoint(s)	Y (v5.0)	To align with CSP
Section 3.4.2 Dual primary endpoint-EFS	25Oct 2022	Updated text to clarify window for post-cystectomy scan,	Y (v5.0)	To align with CSP

Section 1.1 Table 1 Objectives and Endpoints Section 1.3 Number of patients Section 3.4 Efficacy Variable Section 4.2 Analysis Methods	25Oct 2022	Updated text to clarify additional survival endpoint (OS) and pDS rate as secondary endpoints	Y (v5.0)	To align with CSP Data considered to be clinically relevant to summarize
Section 4.2 Analysis Methods	25Oct 2022	Updated text to reflect population for primary analysis and also modified multiple testing plan and schedule for interim analyses Added language describing COVID-19 assessment for patients receiving durvalumab	Y (v5.0)	To align with CSP
Section 5 Interim Analysis	25Oct 2022	Updated text to reflect population for primary analysis and also modified multiple testing plan and schedule	Y (v5.0)	To align with CSP

1. STUDY DETAILS

This Statistical Analysis Plan (SAP) has been prepared in accordance with the Clinical Study Protocol (CSP) version 6.0 (19 June 2023) for study D933RC00001 (NIAGARA). Full details of the study design, rationale, patient selection, enrolment, plan and timing of data collection, etc. can be found therein.

1.1. Study Objectives

Table 1 Objectives and endpoints

Primary objectives:	**Endpoints/variables:**
To assess the efficacy of durvalumab + G+C combination therapy (neoadjuvant)/durvalumab alone (adjuvant) (Arm 1) compared to G+C combination therapy (neoadjuvant)/no adjuvant treatment (Arm 2) in terms of pCR and EFS in MIBC patients	pCR using assessments per central pathology review EFS using assessments per BICR or by central pathology review if a biopsy is required for a suspected new lesion
Secondary objectives:	**Endpoints/variables:**
To assess the efficacy of Arm 1 versus Arm 2 in terms of EFS at 24 months in MIBC patients	EFS24 using assessments per BICR or by central pathology review if a biopsy is required for a suspected new lesion
To assess the efficacy of Arm 1 compared to Arm 2 in terms of pathologic response at radical cystectomy and EFS in MIBC patients	pCR using assessments per local pathology review Proportion of patients who achieve <P2 per local pathology review EFS using assessments per local Investigator or local biopsy review if a biopsy is required for a suspected new lesion EFS24 using assessments per local Investigator or local biopsy review if a biopsy is required for a suspected new lesion
To assess the efficacy of Arm 1 versus Arm 2 in MIBC patients	Metastasis-free survival and disease-specific survival per Investigator assessments or local biopsy review if a biopsy is required for a suspected new lesion Overall survival (OS) OS at 5 years Disease-free survival in patients who undergo radical cystectomy Proportion of patients who undergo radical cystectomy PFS2 as defined by local standard clinical practice

To assess the efficacy of Arm 1 versus Arm 2 in terms of pCR and EFS in MIBC patients in the PD-L1-High subgroup	pCR using assessments per central pathology review EFS using assessments per BICR or by central pathology review if a biopsy is required for a suspected new lesion
To assess disease-related symptoms, physical function, and other HRQoL in Arm 1 versus Arm 2 using the EORTC QLQ-C30 questionnaire	Adjusted mean change from baseline and time to definitive clinically meaningful deterioration in EORTC QLQ-C30 scale/item scores (prioritized domains: fatigue and pain, physical functioning, global health status/quality of life)
To assess the PK of durvalumab when used in combination with G+C	Serum concentration of durvalumab and non-compartmental PK parameters (such as peak and trough concentrations, as data allow; sparse sampling)
To investigate the immunogenicity of durvalumab when used in combination with G+C	Presence of ADAs for durvalumab (confirmatory results: positive or negative)
Safety objective:	**Endpoints/variables:**
To assess the safety and tolerability profile of Arm 1 versus Arm 2 in MIBC patients	AEs, laboratory findings, vital signs, and ECGs
Exploratory objectives:	**Endpoints/variables:**
To assess patient-reported treatment-related symptoms or tolerability of Arm 1 versus Arm 2 using PRO-CTCAE	PRO-CTCAE (items pre-selected based on systemic treatment arms) – descriptive summary of responses
To assess overall health status and overall severity of disease-related symptoms in patients in Arm 1 versus Arm 2 using the PGIC and PGIS questionnaires, respectively	PGIC and PGIS – descriptive summary of responses
To explore the impact of treatment and disease state on health state utility using the EQ-5D-5L	The EQ-5D-5L health state utility index will be used to derive health state utility based on patient reported data
To evaluate tumor-based biomarkers and associations with efficacy parameters, potentially including, but not limited to, microsatellite stability, tumor mutational burden, and other immune-related biomarkers	Association of tumor-based assessments with efficacy and clinical parameters
To evaluate circulatory-based and urine-based biomarkers and associations with efficacy parameters, including, but not limited to, ctDNA	Association of ctDNA, whole blood gene expression, and urine biomarkers with efficacy and clinical parameters

ADA Antidrug antibody; AE Adverse event; BICR Blinded Independent Central Review; ctDNA Circulating tumor DNA; ECG Electrocardiogram; EFS Event-free survival; EFS24 Proportion of patients alive and event free at 24 months using local pathology or BICR; EORTC QLQ-C30 European Organization for Research and Treatment of Cancer 30-item Core Quality of Life Questionnaire; EQ-5D-5L EuroQol 5-dimension, 5-level health state utility index; G+C gemcitabine plus cisplatin; HRQoL Health-related quality of life; MIBC Muscle-invasive bladder cancer; OS Overall survival; pCR Pathologic complete response; PD-L1 Programmed cell death-ligand 1; PFS2 Time from the date of randomization to the earliest date of progression which occurs on subsequent therapy following an EFS event or death;

PGIC Patient Global Impression of Change; PGIS Patient Global Impression of Severity; PK Pharmacokinetics; PRO-CTCAE Patient-reported outcomes version of the Common Terminology Criteria for Adverse Events.

1.2. Study Design

This is a Phase III, randomized, open-label, multi-center, global study to determine the efficacy and safety of durvalumab (MEDI4736) in combination with Gemcitabine+Cisplatin (G+C) for neoadjuvant treatment followed by durvalumab alone for adjuvant treatment in patient with MIBC.

This study will randomize approximately 1050 patients globally in a 1:1 ratio to receive durvalumab + G+C combination therapy every 3 weeks (q3w) (Arm 1) or G+C combination therapy q3w (Arm 2) for 4 cycles of neoadjuvant chemotherapy prior to radical cystectomy. Following radical cystectomy and during adjuvant therapy, patients in Arm 1 will receive durvalumab monotherapy every 4 weeks (q4w) for 8 additional cycles, and patients in Arm 2 will receive no adjuvant treatment.

Neoadjuvant therapy
Patients with adequate renal function (CrCl ≥ 60 mL/min)

- Arm 1: Day 1: durvalumab 1500 mg IV, cisplatin 70 mg/m^2, gemcitabine 1000 mg/m^2; Day 8: gemcitabine 1000 mg/m^2; every 21 days for 4 cycles.
- Arm 2: Day 1: cisplatin 70 mg/m^2, gemcitabine 1000 mg/m^2; Day 8: gemcitabine 1000 mg/m^2; every 21 days for 4 cycles.

Patients with borderline renal function (CrCl ≥ 40 mL/min to < 60 mL/min)

- Arm 1: Day 1: durvalumab 1500 mg IV, cisplatin 35 mg/m^2, gemcitabine 1000 mg/m^2; Day 8: gemcitabine 1000 mg/m^2, cisplatin 35 mg/m^2; every 21 days for 4 cycles.
- Arm 2: Day 1: cisplatin 35 mg/m^2, gemcitabine 1000 mg/m^2; Day 8: gemcitabine 1000 mg/m^2, cisplatin 35 mg/m^2; every 21 days for 4 cycles.

In scenarios when patients are unable to complete the intended 4 cycles of chemotherapy prior to radical cystectomy, patients will be permitted to receive less than 4 cycles of chemotherapy, at the discretion of the Investigator and upon discussion with AstraZeneca.

Noncystectomy extension phase:
Patients in either treatment arm who fulfil the necessary criteria may enter the noncystectomy extension phase after consultation and approval by AstraZeneca. Patients enrolled into Arm 1 who enter the noncystectomy extension phase may be administered durvalumab 1500 mg (as monotherapy) every 28 days for a maximum of 8 doses (corresponding to a maximum exposure of 12 months) or until study-specific discontinuation criteria is met. For patients who enter the noncystectomy extension phase and subsequently undergo a radical cystectomy, further treatment should be discussed and agreed upon with AstraZeneca.

Adjuvant therapy (regardless of renal status)

- Arm 1: Day 1: durvalumab 1500 mg IV; every 28 days for 8 cycles.

- Arm 2: No adjuvant treatment.

Adjuvant therapy is recommended to begin as soon as the patient recovers from radical cystectomy and within 120 days after and no earlier than 42 days after radical cystectomy. Cycle 1 Day 1 of the adjuvant treatment phase for patients in Arm 2 is recommended to occur as soon as the patient recovers from the radical cystectomy (no earlier than 42 days and no later than 120 days after radical cystectomy).

Randomization will be stratified by:

- Tumor stage (T2N0) versus >T2N0 [including T2N1, T3 and T4a])
- Renal function (adequate renal function versus borderline renal function)
- PD L1 status (high versus low/negative).

Patients will provide a tumor tissue sample at screening to determine PD-L1 status for stratification. Crossover from Arm 1 to Arm 2 will not be permitted.

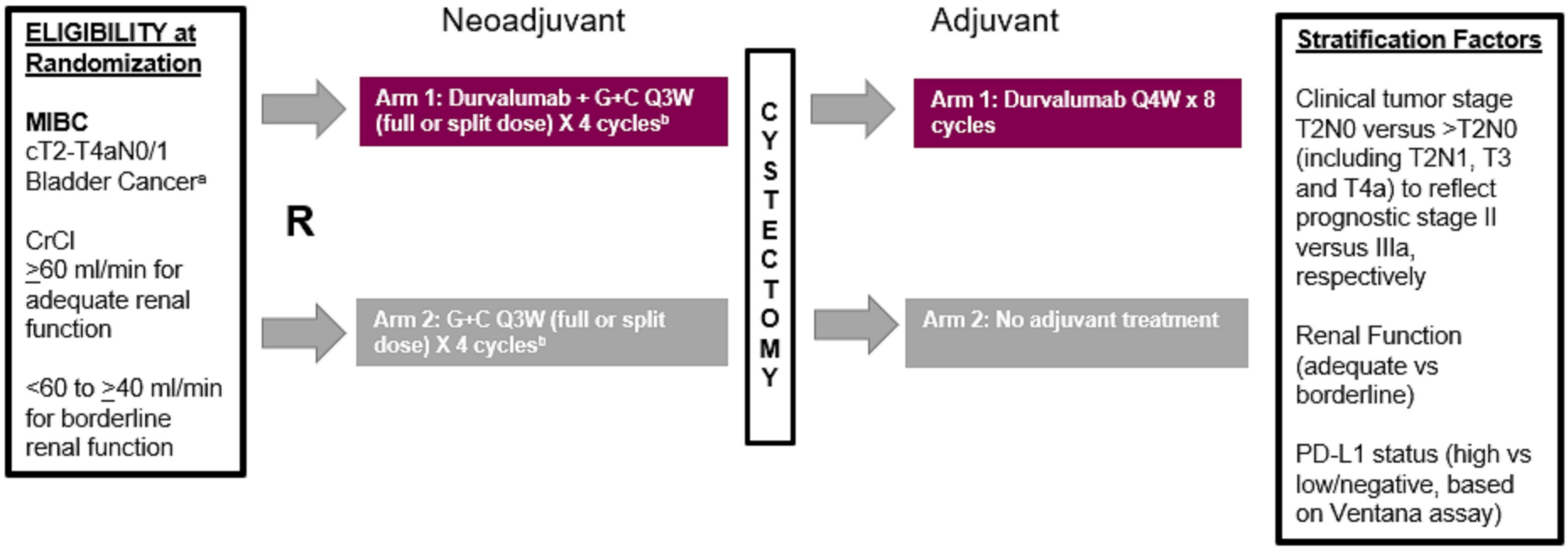
Note

For more information related to the screening period, the treatment, duration of treatment, progression during treatment, progression during follow-up and the survival follow-up period, please see the synopsis of the CSP.

For the summary of the overall study design, please see Figure 1 below.

Doses and treatment regimens are described in section 6.1.2 of the CSP. Assessments will be conducted as indicated in Table 1, Table 2, Table 3, and Table 4 in the CSP.

Figure 1 Overall study design



[a] Enrolment of patients with T2N0 disease is limited to approximately 40% of the targeted global population (for both treatment arms).

[b] Patients with borderline renal function will receive split-dose G+C and will be limited to up to 20% of the targeted global population.

CrCl Creatinine clearance; Durva Durvalumab; G+C Gemcitabine and cisplatin; q3w Every 3 weeks; q4w Every 4 weeks; MIBC Muscle-invasive bladder cancer; PD-L1 Programmed cell death-ligand 1; R Randomization.

1.3.Number of Patients

This study will plan to enroll and screen approximately 1400 patients at approximately 150 sites in order to randomize approximately 1050 eligible patients in a 1:1 ratio to receive either durvalumab + G+C combination therapy (neoadjuvant)/durvalumab alone (adjuvant) in Arm 1 or G+C combination therapy (neoadjuvant)/no adjuvant treatment in Arm 2.

The study is sized to characterize the pCR rate and EFS benefit of Arm 1 versus Arm 2 in MIBC patients who have not received prior systemic chemotherapy.

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 ⩽22 months] following the start of the study is assumed to be $(t/22)^2$.

The final analysis of the dual primary pCR will be performed approximately 6 months after the last patient is randomized to the study.

The final analysis of the dual primary EFS will be performed when approximately 451 EFS events in ITT have occurred across the 2 arms (43% maturity) or June 2025, whichever occurs first.

Arm 1 versus Arm 2 (pCR in ITT)

It is assumed that the pCR for patients in Arm 2 is 35% (Grossman et al 2003). Under the alternative hypothesis, pCR is assumed to be 50% for Arm 1. With 525 patients in each arm, the study will have at least 95% power to demonstrate a statistically significant difference at a 2-sided alpha level of 0.1%. The smallest treatment effect that could be observed as being statistically significant at the time of pCR analysis is 45% in Arm 1. This translates to an increase of approximately 10% from that in Arm 2, where pCR is assumed to be 35%.

Arm 1 versus Arm 2 (EFS in ITT)

The assumed EFS treatment effect under the alternative hypothesis is an average HR of 0.733 for Arm 1 versus Arm 2. This is based on the following assumptions:

- An exponential model was assumed for EFS such that in patients who are assigned to the Arm 1, the overall median EFS is 38.0 months and the EFS rate at 24 months is 64.5%.

- For Arm 2, an exponential model was assumed for EFS such that in patients who are assigned to Arm 2, the overall median EFS is 27.8 months and the EFS rate at 24 months is 55%.

Based on a blinded event prediction, an estimated 451 EFS events (43% maturity) are expected to be observed at 45 months after the date of the last patient randomized. With 451 EFS events, the study will have at least 90% power to demonstrate a statistically significant difference at a 2-sided overall alpha level of 4.90%, allowing two interim analyses to be conducted at

approximately 67% and 91% of the target events. The smallest treatment difference that could be statistically significant will be an average HR of 0.82.

Arm 1 versus Arm 2 (OS in ITT)

The assumed OS treatment effect under the alternative hypothesis is an average HR of 0.76 for Arm 1 versus Arm 2. This is based on the following assumptions:

- An exponential model was assumed for OS such that in all patients who are assigned to Arm 1, the overall median OS is 8.6 years (103 months), and the OS rate at 5 years is 66.8%.
- For Arm 2, an exponential model was assumed for OS such that in all patients who are assigned to Arm 2, the overall median OS is 6.5 years (78 months) and the OS rate at 5 years is 58.7%.

The final analysis of OS based on approximately 428 OS events for the comparison of Arm 1 versus Arm 2 (41% maturity, 428/1050), from ITT, is expected to occur 5 years (60 months) after the last patient is randomized to the study and will provide at least 80% power to demonstrate a statistically significant difference in OS at a 2-sided alpha level of 4.9%, allowing for two interim analyses (only formerly tested if EFS is positive per the MTP) to be conducted at approximately 67% and 74% of the target events. The smallest treatment difference that could be statistically significant will be an HR of 0.82.

Arm 1 versus Arm 2 (OS5 in ITT)

The statistical model assumptions for OS5 in the ITT of each arm are stated above.

The analysis of OS5 that is performed at the time of the final analysis of OS will provide at least 77% power to demonstrate a statistically significant difference in OS5 at a 2-sided alpha level of 4.9%.

2. ANALYSIS SETS

2.1. Definition of Analysis Sets

2.1.1. Full Analysis Set (Intention to treat [ITT])

The full analysis set (FAS) will include all randomized patients. The FAS will be used for all efficacy analyses (including PROs) and the treatment arms will be compared based on 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 arm to which they were randomized.

2.1.2. Cystectomy population

The Cystectomy population will include all patients in FAS who undergo radical cystectomy and disease free at adjuvant baseline. Unless otherwise specified, the analysis set will be used for DFS only. Treatment arms will be compared on the basis of randomized study treatment, regardless of the treatment actually received.

2.1.3. 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 Ventana PD-L1 (SP263) Assay at baseline by IVRS.

2.1.4. Safety Analysis Set

The safety analysis set will consist of all patients who received at least 1 dose of the study treatment. Safety data will not be formally analysed but summarized using the safety analysis set according to the treatment received; that is, erroneously treated patients for e.g., those randomized to treatment A but actually given treatment B will be summarized according to the treatment they actually received.

2.1.5. PK Analysis Set

All patients who receive at least 1 dose of IP per the protocol for whom any post-dose data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analyses will be included in the PK analysis set. The PK analysis set will be defined by AstraZeneca/Medimmune, the Study Physician, Pharmacokineticist, and Statistician prior to any analyses being performed.

2.1.6. ADA Analysis Set

All patients who receive at least 1 dose of IP per the protocol for whom baseline and at least one non-missing post-baseline ADA result are available will be included in the ADA analysis set.

Table 2 below presents the summary of outcome variables and analysis populations.

Table 2 Summary of Outcome Variables and Analysis Populations

Outcome Variable	Population
Efficacy Data	
pCR rate	Full analysis set (ITT population) PD-L1 High analysis set
EFS	Full analysis set (ITT population) PD-L1 High analysis set
Proportion of patients who achieve <P2, EFS24, Metastasis-free survival (MFS),	Full analysis set (ITT population)

Outcome Variable	Population
Disease-specific survival(DSS), OS, OS5, PFS2, Proportion of patients who undergo cystectomy,	
PROs	Full analysis set (ITT population)
DFS	Cystectomy population
Study Population/Demography Data	
Demography characteristics (e.g., age, sex etc.)	FAS (ITT)
Baseline and disease characteristics	FAS (ITT)
Medical and surgical history	FAS (ITT)
PK Data	
PK data	PK Analysis Set
ADA Data	
ADA data	ADA Analysis Set
Safety Data	
Exposure	Safety Analysis Set
AEs	Safety Analysis Set
Laboratory measurements	Safety Analysis Set
Vital signs	Safety Analysis Set
ECOG performance status	Safety Analysis Set
Physical examinations	Safety Analysis Set
ECG	Safety Analysis Set

AE Adverse event; DFS Disease-free survival; DSS Disease-specific survival; ECOG Eastern Cooperative Oncology Group; EFS Event-free survival; EFS24 Proportion of patients alive and event free at 24 months; ITT Intent-to-treat; MFS Metastasis-free survival; OS5 Proportion of patients alive at 5 years; pCR Pathologic complete response; PFS2 The time from the date of randomization to the earliest date of progression which occurs on subsequent therapy following an EFS event, or death; PK Pharmacokinetic; PRO Patient-reported outcome.

2.2. Protocol Deviations

The following general categories will be considered important protocol deviations (IPDs) and will be programmatically derived from the electronic case report form (eCRF) data. These will be listed and discussed in the clinical study report (CSR) consistent with the guidelines in *Table1* of the *Non-compliance Handling Plan*.

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.

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 errors in treatment dispensing (with regard to continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader and/or statistician.

The important protocol deviations will be listed and summarised by randomised treatment arm. Deviation 1 will lead to exclusion from the safety analysis set. None of the other deviations will lead to patients being excluded from the analysis sets described in section 2.1 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK). A per-protocol analysis excluding patients with specific IPDs is not planned; however, a sensitivity analysis for deviation bias may be performed on the EFS endpoint excluding patients with deviations that may affect the efficacy of the trial therapy if > 10% of patients in either treatment group had one or more important protocol deviations.

The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of database lock and will be documented prior to the primary analysis being conducted.

In addition to the programmatic determination of the deviations above, other study deviations captured from the CRF module for inclusion/exclusion criteria will be tabulated and listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

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 tumor stage, renal function (adequate or borderline), and PD-L1 status (high or low/negative) based on the IVRS will be summarized and/or listed.

3 PRIMARY AND SECONDARY VARIABLES

This study will evaluate the dual primary endpoints of pCR and EFS. Efficacy assessments of pCR (primary) will be derived using central pathology review of the radical cystectomy sample. The efficacy assessments of EFS (primary) and EFS24 will be derived using BICR assessments according to RECIST 1.1 or by central pathology review if a biopsy is required for a suspected new lesion.

Additional secondary objectives will be OS, OS5, PFS2 as defined by local standard clinical practice, and proportion of patients who undergo radical cystectomy. The proportion of patients who achieve <P2 will be derived (by AstraZeneca) using local pathology assessments of radical cystectomy samples.

3.1 Derivation of RECIST visit responses

For all subjects, the radiological efficacy will be assessed by RECIST version 1.1 (see clinical study protocol [CSP] Appendix F).

A “Neoadjuvant Baseline” radiological tumor assessments are to be performed no more than 28 days before the date of randomization and ideally as close as possible to randomization (see CSP Table 1).

A post-neoadjuvant/ pre-radical cystectomy follow-up scan must be performed upon completion of neoadjuvant chemotherapy prior to surgery.

An “Adjuvant Baseline” scan should be collected 42 days (±14 days) after radical cystectomy and ideally should be performed as close as possible and prior to the first date of adjuvant phase (see CSP Table 2).

In most instances, no lesions will be observed on the Adjuvant Baseline scans and ‘No Evidence of Disease’ will be recorded for the Adjuvant Baseline RECIST assessment; however, if any radiological observable tumors exist, a new selection of Target and Non-Target lesions are recorded. A follow-up scan should be performed at least 4 weeks later, as an assessment using RECIST 1.1 criteria and then every 12 weeks, thereafter. The use of an earlier scan is in place to allow early confirmation of a metastatic lesion. Additionally, a new lesion can be evaluated pathologically at any time, when feasible, to confirm metastatic disease.

Adjuvant tumor assessments occur every 12 weeks ± 7 days, after the date of cystectomy for the first 24 months, then every 24 weeks ±7 days for 36 months, and then every 52 weeks (annually) thereafter until unequivocal progression, the end of study, death, study discontinuation, or Sponsor decision, whichever comes first.

Where possible or feasible, radiological progression should be biopsy proven. If a new lesion is equivocal and biopsy cannot be performed, treatment should continue and the lesion should be assessed at a subsequent scan no earlier than 6 weeks later or at the next scheduled imaging visit to determine if it becomes unequivocal.

If measurable, in order for a previously equivocal new lesion to become unequivocal at a subsequent scan, the long axis diameter of the previously new equivocal nodal lesion should show an increase of at least 5 mm. If the event of progression is confirmed on the subsequent follow-up scan, the date of progression corresponds to the first evidence of progression. Other imaging modalities (e.g., bone scan, MRI scan) may be required to define progression in equivocal cases.

During adjuvant treatment, the imaging schedule must be followed regardless of any delays in dosing.

For patients who do not have radical cystectomy and therefore will not have an "Adjuvant Baseline" scan or adjuvant treatment (see CSP Table 4), scans will be conducted every 12 weeks ±7 days after the date of the pre- cystectomy scan, and these follow-up scans will use the original "neoadjuvant" screening scan as the baseline scan for RECIST 1.1 assessments.

If an unscheduled assessment is performed, and the subject has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled imaging visits. This schedule is to be followed in order to minimize any unintentional bias caused by some subjects being assessed at a different frequency than other subjects.

For patients who discontinue treatment due to toxicity or other reasons in the absence of confirmed objective recurrence, tumor assessments should continue according to the schedules of assessments.

The following sections pertain to site investigator data and the programmatic derivation of visit response.

Please note, Section 3.1 defines general RECIST 1.1 criteria.

3.1.1 Investigator RECIST 1.1-based assessments: Target lesions

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

A subject can have a maximum of five measurable lesions recorded at baseline with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as TLs. Lymph nodes, in any location, are collectively considered as a single organ, with a maximum of 2 lymph nodes as TLs. A bilateral or multi-lobular organ is considered as a single organ. If more than one baseline scan is recorded then measurements from the one that is closest and prior to 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 TLs should be identified as NTLs at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

If subjects do not have measurable disease at entry (i.e. no TLs), but have non-measurable disease, are enrolled in the study, evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions (see Section 3.1.3 for further

details). If a subject does not have measurable disease at baseline then the TL visit response will be not applicable (NA).

If no TLs and no NTLs are recorded at a visit, both the TL and NTL visit response will be recorded as NA and the overall visit response will be no evidence of disease (NED). If a new lesion is observed, then the overall visit response will be PD.

Table 3 TL visit responses (RECIST 1.1)

Visit responses	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 ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 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 if any of the TLs at follow-up were not assessed or not evaluable (e.g. missing anatomy) or had a lesion intervention at this visit. Note: If the sum of diameters meets the PD criteria, PD overrides NE as a TL response.
Not applicable (NA)	No TLs are recorded at baseline.

3.1.1.1.1 Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to one 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%.

3.1.1.1.2 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 longest diameter (LD) recorded.

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

If all TL measurements are missing, then the TL visit response is NE. Overall visit response will also be NE, unless there is a progression of non-TLs or new lesions, in which case the response will be PD.

3.1.1.1.3 Lymph nodes

For lymph nodes, if the size 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.

3.1.1.1.4 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 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.

3.1.1.1.5 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 most cases.

3.1.1.1.6 TL too small to measure

If a TL becomes too small to measure, then this will be indicated as such on the eCRF and a value of 5mm will be entered into the database and used in TL calculations. However, a smaller value may be used if the radiologist has not indicated 'too small to measure' on the case report form and has 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.

3.1.1.1.7 Irradiated lesions/lesion intervention

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolization), 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 tumors:

- 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 subject would be assigned a TL response of PD.

- Step 3: If, after both steps, PD has not been assigned, then, if appropriate (i.e. if ≤ 1/3 of the TLs have missing measurements), the scaled sum of diameters calculated in step 2 should be used, and PR or SD then assigned as the visit response. Subjects 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 have a value of 0 (or <10mm for lymph nodes) 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 (as per step 2 above).

3.1.1.1.8 Scaling (applicable only for irradiated lesions/lesion intervention)

If > 1/3 of TL measurements are missing (because of intervention) then the TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (i.e. 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 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.

An 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 mm. The sum of the corresponding lesions at the nadir visit is 26.8 mm.

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

(26 / 26.8) * 29.3 = 28.4 mm

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.

3.1.1.1.9 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.

3.1.1.1.10 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 0 mm.

3.1.1.1.11 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 and new lesions

At each visit, the investigator should record an overall assessment of the NTL response. 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 4 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 subjects 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 tumor 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 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 tumor.

If a new lesion is equivocal, for example because of its small size, the treatment and tumor assessments should be continued until the previously new lesion has been assessed as

unequivocal and then the progression date should be declared using the date of the initial scan when the new lesion first appeared.

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 since baseline' 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. This scenario (i.e., whereby new lesion response is NE), should only occur in exceptional cases, as missing data for the new lesion field should always be queried.

Symptomatic 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.

Subjects with 'symptomatic progression' requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumor assessments where possible until objective disease progression is observed, or death occurs (whichever comes first).

3.1.3 Investigator RECIST 1.1-based assessments: Overall visit response

Table 5 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 5 Overall visit responses

Target lesions	Non-target lesions	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
NA	CR	No (or NE)	CR
NA	Non-CR/Non-PD	No (or NE)	SD
NA	NE	No (or NE)	NE

NA	NA	No (or NE)	NED

CR Complete response, NA Not applicable (only relevant if there were no NTLs at baseline), NE Not evaluable, NED No evidence of disease, PD Progressive disease, PR Partial response, SD Stable disease.

3.1.4 Blinded Independent Central Review of RECIST 1.1-based assessments

All images, including unscheduled visit scans, will be collected on an ongoing basis and sent to an AstraZeneca-appointed Contract Research Organization (CRO) for quality control and storage. Guidelines for image acquisition, de-identification, storage at the investigative site as source data, and transfer to the imaging CRO will be provided in a separate document. A BICR of images will be performed at the discretion of AstraZeneca. The results of these independent reviews will not be communicated to Investigators, and the results of Investigator RECIST 1.1 assessments will not be shared with the central reviewers. The management of patients will be based in part upon the results of the RECIST 1.1 assessment conducted by the Investigator. Further details of the BICR will be documented in the Independent Review Charter (also referred to as “Imaging Charter”).

3.2 Other tumor assessment

3.2.1 Local biopsy review

Where possible or feasible, suspected progression/recurrence events should be biopsy proven as soon as feasible. If a biopsy is performed and the histopathological assessment reveals the presence of recurrent tumor, progression will be recorded using the date of biopsy.

3.2.2 Local pathology review

Local pathology review of radical cystectomy specimen to assess the pathological stage will be based on American Joint Committee on Cancer tumor-node-metastasis classification of carcinomas of the urinary bladder.

3.2.3 Central biopsy review

Where possible or feasible, suspected progression/recurrence events should be biopsy proven as soon as feasible. If a biopsy is performed and the histopathological assessment reveals the presence of recurrent tumor, progression will be recorded using the date of biopsy.

3.2.4 Central pathology review

Central pathology review of radical cystectomy specimen to assess the pathological stage will be performed at the discretion of AstraZeneca. Guidelines for sample requirements will be provided in a separate document. A central pathology review will be based on American Joint Committee on Cancer tumor node metastasis classification of carcinomas of the urinary bladder.

3.3 Survival assessments

Assessments for survival must be made at Months 3, 6, and 9 (±1 week); Month 12 (±2 weeks); and then every 6 months thereafter (±2 weeks) following treatment discontinuation or adjuvant

phase study visits (see CSP Table 4). Survival information may be obtained via telephone contact with the patient or the patient's family, or by contact with the patient's current physician. The details of first and subsequent therapies for cancer, after discontinuation of treatment, will be collected.

In addition, patients on treatment or in survival follow-up will be contacted following the data cutoff for the primary analysis and all subsequent survival analyses to provide complete survival data. These contacts should generally occur within 7 days of the data cutoff. If patients are confirmed to be alive or if the death date is after the data cutoff date, these patients will be censored at the date of data cutoff. Death dates may be found by checking publicly available death registries, where allowed by local regulations.

3.4 Efficacy Variables

3.4.1 Dual primary endpoint – pCR

The dual primary pCR is the pCR assessment in MIBC patients per central pathology review.

pCR rate is defined as the proportion of patients whose pathological staging was T0N0M0 as assessed per central pathology review using specimens obtained via radical cystectomy following the neoadjuvant treatment. The denominator for pCR will be the number of patients in the FAS.

pCR will also be assessed per local pathology review.

3.4.2 Dual primary Endpoint – EFS

The dual primary EFS is the EFS assessment in MIBC patients per BICR or by central pathology review if a biopsy is required for a suspected new lesion.

EFS is defined as the time from randomization to the first recurrence of disease post radical cystectomy, time of first documented progression in patients who were medically precluded for radical cystectomy, or time of expected surgery in patients who refuse to undergo a radical cystectomy or failure to undergo a radical cystectomy in participants with residual disease, or the time of death due to any cause, whichever occurs first.

- A recurrence of disease includes local (pelvic) recurrence of UC, urinary tract recurrence of UC, or distant metastasis of UC. In the event that progression is confirmed via biopsy or subsequent scans (the confirmation of suspected new lesions initially identified in the scans if applicable), the date of recurrence will be the earliest date among the initial detection of radiological unequivocal new lesion, or the pathological confirmation of new lesion if biopsy is performed to confirm suspected new lesion post cystectomy, or the death due to any causes.

- Patients who are suspected of having microscopic disease (i.e., no evidence on imaging) or who have documented macroscopic disease (confirmed by imaging) at the completion of neoadjuvant therapy and who refuse to proceed with a radical

cystectomy, are declared as progressed, with EFS being declared at the time of expected surgery.

- For patients who fulfil criteria for a complete clinical response, refuse an initial radical cystectomy and are entered in a noncystectomy extension phase, EFS is defined as time to the first recurrence of disease following a delayed radical cystectomy (if performed). For patients who are medically precluded from or refuse a delayed radical cystectomy, EFS is confirmed at time of unequivocal progression.

EFS will be assessed using CT/MRI and pathology testing performed according to local standards and as clinically indicated. The EFS assessment will be done by BICR or by central pathology review if a biopsy is required for a suspected new lesion, and by local investigator or local biopsy review if a biopsy is required for a suspected new lesion.

Patients who take subsequent therapy prior to their last evaluable RECIST assessment or progression or death will not be censored at their last evaluable RECIST assessment prior to taking the subsequent therapy. Additionally, if the patient progresses or experiences recurrent disease or dies directly preceded by 2 or more consecutive missed visits, the patient will still be counted as having an EFS event. For both of these situations a sensitivity analysis will be performed.

Patients who have not progressed or experienced recurrent disease or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable disease assessment. For the purpose of EFS, the date of surgery is considered as disease assessment date.

If the patient has no evaluable visits or does not have baseline disease assessment (i.e., a baseline scan) prior to neoadjuvant treatment, they will be censored at Day 1 unless they die within 112 days of randomization. If an adjuvant baseline scan is not recorded, it will be considered that no lesions are presented following surgery.

The EFS time will always be derived based on assessment dates and not visit dates.

If a subject is known to have died where only a partial death date is available, then the date of death will be imputed as described in **3.4.3.3**.

3.4.3 Secondary Endpoints

3.4.3.1 Proportion of patients who achieve <P2

The proportion of patients who achieve <P2 is defined as the proportion of patients whose pathological staging at radical cystectomy was P0 (T0N0M0)/Pa/P1/Cis as assessed per central pathology review using specimens obtained via radical cystectomy following the neoadjuvant treatment.

The denominator for this endpoint is the number of patients in the FAS.

3.4.3.2 EFS24

EFS24 will be defined as the Kaplan-Meier estimate of EFS at 24 months after randomization, as assessed per BICR or by central pathology review if a biopsy is required for a suspected new lesion, and per local investigator or local biopsy review if a biopsy is required for a suspected new lesion.

3.4.3.3 Overall Survival and Overall Survival at 5 Years (OS5)

The OS5 will be defined as the KM estimate of OS at 5 years after randomization.

OS is defined as the time from the date of randomization until death due to any cause regardless of whether the patient withdraws from randomized therapy or receives another anti-cancer therapy (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 electronic case report form [eCRF]). In order to minimize confounding of OS, crossover is not permitted in this study.

Note: Survival calls will be made in the week following the date of Data Cut Off (DCO) for the analysis, and if patients are confirmed to be alive or if the death date is after 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.

Note, for any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment (since the SURVIVE module is only completed for patients off treatment if a survival sweep is not performed). The last date for each individual patient is defined as the latest among the following dates recorded on the case report forms (CRFs):

- AE start and stop dates
- Admission and discharge dates of hospitalization
- Study treatment date

• End of treatment date

• Laboratory test dates

• Date of vital signs

• Disease assessment dates on RECIST CRF

• Start and stop dates of alternative anticancer treatment

• Date last known alive on survival status CRF

• End of study date

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

a. For Missing day only – using the 1st of the month

b. For Missing day and Month – using the 1st of January

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

3.4.3.4 Metastasis-free survival (MFS)

MFS is defined as the time from date of randomization until the first recognition of distant metastases or death, whichever occurs first. Patients who were alive and free from metastases were censored at the time of the latest date of assessment from their last evaluable disease assessment.

3.4.3.5 Disease-specific survival (DSS)

DSS is defined as the time from the date of randomization until death due to bladder cancer (BC). Any patient not known to have died due to BC at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive.

3.4.3.6 Proportion of patients who undergo cystectomy

The proportion of patients who undergo cystectomy is defined as the proportion patients who undergo radical cystectomy after the neoadjuvant treatment. The denominator will be patients in the FAS.

3.4.3.7 Disease-free survival (DFS)

DFS is defined as the time from the date of radical cystectomy to the first recurrence of disease post radical cystectomy, or death due to any cause, whichever occurs first. DFS will be assessed in patients who undergo radical cystectomy and are disease free at adjuvant baseline visit per BICR assessment.

3.4.3.8 Time from randomization to subsequent progression or recurrence post-EFS event (PFS2)

Time from randomisation to second progression or death (PFS2) will be defined as the time from the date of randomisation to the earliest of the progression event subsequent to first subsequent therapy or death. 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 imaging, symptomatic progression, or death. 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.

If a patient was censored for EFS, that patient will also be censored for PFS2 at the same censoring date unless the patient died after being censored for EFS.

3.5 Patient-Reported Outcomes (PRO) Variables

All questionnaires will be scored according to published scoring guidelines or the developer's guidelines. All PRO analyses will be based on the FAS, unless otherwise stated.

3.5.1 EORTC QLQ-C30

The EORTC QLQ-C30 consists of thirty questions that can be aggregated into five functional scales (physical, role, cognitive, emotional, and social), three multi-item symptom scales (fatigue, pain, and nausea/vomiting), six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties), and a two-item global health status/QoL (global HRQoL) scale. None of the items is included in more than one scale.

For EORTC QLQ-C30, the primary or prioritized domains/endpoints include physical function, global HRQoL and fatigue and pain.

The EORTC QLQ-C30 scales/items will be scored according to the EORTC QLQ-C30 Scoring Manual (Fayers et al, 2001). An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales/symptom items, for each of the functional scales, and for the global HRQoL scale according to the EORTC QLQ-C30 Scoring Manual. Higher scores on the global HRQoL and functioning scales indicate better health status/function, whereas higher scores on symptom scales and individual symptom items represent greater symptom severity. The number of items and item range for each scale/item are displayed in **Table 6** below.

Table 6 Scoring the EORTC QLQ-C30

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

**Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range*=3.

Clinically Meaningful Changes

A minimum clinically meaningful change in an EORTC QLQ-C30 scale/item is defined as an absolute change in the score from baseline of ≥ 10 point increase or decrease for scales/items (Osoba et al, 1998). At each post-baseline assessment, the change in symptoms/functioning score from baseline will be categorized as improvement, no change or deterioration as shown in **Table 7** below.

Table 7 EORTC QLQ-C30 Clinically Meaningful Changes

Score	Change from baseline	Assessment period response
EORTC QLQ-C30 global health status/ QoL and functional scales	≥ 10 point increase	Improvement
	≥ 10 point decrease	Deterioration
	Otherwise	No change
EORTC QLQ-C30 symptom scales and items	≥ 10 point decrease	Improvement
	≥ 10 point increase	Deterioration
	Otherwise	No change

A patient's best overall response in symptoms, function, or global health status/QoL will be derived as the best response the patient achieved, based on evaluable PRO data collected during the study period. The criteria in **Table 8** will be used to assign a best response in symptoms or function or global health status/QoL.

Table 8 Best response in EORTC QLQ-C30

Overall score response	Criteria
Improved	Patient meets one of the following criteria: 1. Has visit response of "improvement" and subsequent visit response of "improvement" within 2 PRO assessment visits and at least 21 days apart. 2. Has 1 visit response of 'improvement' and no further assessments.
No Change	Patient does not qualify for an overall score response of 'improved' and meets 1 of the following criteria: 1. Has 2 consecutive visit responses of 'no change'. 2. Has 1 visit response of 'no change' and no further assessments. The two responses must be at least 21 days apart.
Deterioration	Patient does not qualify for an overall score response of 'improved' or 'no change' and meets 1 of the following criteria: 1. Has 2 consecutive visit responses of 'deterioration' at least 21 days apart. 2. Has 1 visit response of 'deterioration' and no further assessments. 3. Has 1 visit response of 'improvement', 'no change', or 'deterioration' followed by death within 2 PRO assessment visits.

Overall score response	Criteria
Other	Patient meets one of the following criteria: 1. Does not qualify for one of the above. 2. Has either no baseline or no post-baseline evaluable PRO assessment.
Missing	Patient has no baseline and no post-baseline evaluable PRO assessments.

For each subscale, if <50% of the items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items in the subscale (Fayers et al, 2001). 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.

3.5.1.1 Time to definitive/sustained clinically meaningful deterioration in health-related QoL, functioning or symptoms

Time to definitive/sustained clinically meaningful deterioration in global health status/QoL and functioning as measured by EORTC QLQ-C30 scales/items will be defined as the time from the date of randomization until the date of the first observation with ≥10-point decrease in score with no subsequent observations with <10-point decrease from baseline. Similarly, time to definitive clinically meaningful deterioration in symptoms as measured by EORTC QLQ-C30 scales/items will be defined as the time from the date of randomization until the date of the first observation with ≥10-point increase in score with no subsequent observations with <10-point increase from baseline. Sensitivity analysis will be performed by including death (by any cause) in the absence of a clinically meaningful deterioration, regardless of whether the patient discontinues study drug(s) or receives another anticancer therapy prior to global health status/QoL, function or symptom deterioration. Such death will be included as an event in the sensitivity analysis only if it occurs within 2 PRO assessment visits from the last available PRO assessment. Patients whose global health status/QoL, functioning, or symptoms (as measured by EORTC QLQ-C30) has not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the global health status/QoL, function, or symptom could be evaluated.
The at-risk population for the analysis of time to global health status/QoL or functioning deterioration is defined as the subset of the FAS having baseline scores of ≥10. The at-risk population for the analysis of time to symptom deterioration is defined as the subset of the FAS having baseline scores ≤90.

3.5.1.2 Symptom Improvement Rate

The symptom improvement rate will be defined as the number (and proportion) of patients with a best overall score response of "improved" in symptoms. The denominator will consist of a subset of the FAS who have a baseline symptom score ≥10 decrease.

3.5.1.3 HRQoL/Function Improvement Rate

The HRQoL/function improvement rate (hereafter function improvement rate) will be defined as the number and proportion of patients with a best overall response of "improved" in QoL or function. The denominator will consist of a subset of patients in the FAS who have a baseline HRQoL/function score ≤90.

3.5.2 PGIC

The response options of the Patient Global Impression of Change (PGIC) are scored as follows: Much Better (+3), Moderately Better (+2), A Little Better (+1), About the Same (0), A Little Worse (-1), Moderately Worse (-2), and Much Worse (-3). Data from the PGIC will be summarized using FAS.

3.5.3 PGIS

The response options of the Patient Global Impression of Severity (PGIS) are scored as follows: No symptoms (0), very mild (1), mild (2), moderate (3), severe (4), and very severe (5). Data from the PGIS will be summarized using FAS.

3.5.4 PRO-CTCAE

The PRO-CTCAE system has been developed by the National Cancer Institute (NCI). The PRO-CTCAE will only be administered in those countries where a linguistically validated version exists. PRO-CTCAE is an item bank of symptoms experienced by patients while undergoing treatment of their cancer. It was developed in recognition that collecting symptom data directly from patients using PRO tools can improve the accuracy and efficiency of symptomatic AE data collection. This was based on findings from multiple studies demonstrating that physicians and nurses underestimate symptom onset, frequency, and severity in comparison with patient ratings (Sprangers et al 1992; Litwin et al 1998; Basch et al 2009). To date, 78 symptoms of the CTCAE (version 4) have been identified to be amenable to patient reporting. These symptoms have been converted to patient terms (e.g., CTCAE term "myalgia" converted to "aching muscles"). For several symptoms, like fatigue and pain, additional questions are asked about symptom frequency, severity, and interference with usual activities. For other symptoms like rash, additional questions focus on the presence on the body. The items included in the PRO-CTCAE have undergone extensive qualitative review among experts and patients. Using cognitive testing methods, these items and the additional questions for some of the symptoms have been extensively evaluated by cancer patients, so that symptoms of interest are clear, comprehendible and measurable. Not all items are administered in any one clinical trial. The intention is to only ask patients to complete those items, which are considered relevant for the trial, site of cancer, and cancer treatment.

For this study, 16 items are considered relevant for this cancer treatment, i.e. problems with tasting food or drinks, decreased appetite, nausea, vomiting, constipation, diarrhea, shortness of breath, cough, rash, itching, numbness or tingling in hands or feet, dizziness, joint pain, muscle pain, fatigue, abdominal pain, , chills, constipation, cough, decreased appetite, diarrhea, dizziness, fatigue, headache, itching, joint pain, muscle pain, nausea, painful urination, rash, shortness of breath, urinary frequency, vomiting (Appendix H of the CSP).

3.5.5 Health State Utility (EQ-5D-5L)

The EQ-5D-5L is a standardized measure of health status developed by the EuroQoL Group in order to provide a simple, generic measure of health for clinical and economic appraisal (EuroQoL Group 1990). 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 EQ-5D descriptive system, respondents also assess their health on the day of assessment on a visual analogue scale (EQ-VAS), ranging from 0 (worst imaginable health) to 100 (best imaginable health). This score is reported separately to the EQ-5D-5L health states.

The evaluable population will comprise of a subset of the FAS who have completed the EQ-5D-5L questionnaire (5 questions and EQ-VAS) at baseline.

3.5.6 PRO Compliance

Summary measures of overall compliance and visit compliance will be derived for each PRO questionnaire. These will be based upon the following definitions:

- Expected questionnaire = questionnaire to be completed by a patient at a scheduled assessment time point e.g., 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 latest of progression and safety follow-up will be used to assess whether the patient is still under PRO 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 patients or forms.

- Received questionnaire = a completed questionnaire received from a patient at a scheduled visit and has a completion date and at least a one item domain completed.
- Evaluable = completed questionnaire received from a patient with enough responses to score at least one scale/domain.
- Overall PRO compliance rate is defined for each randomized treatment arm as the 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 multiplied by 100.

Compliance rate 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 at the time point. Similarly, the evaluability rate will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire (per definition above), divided by the number of patients from whom questionnaires were received.

3.5.7 Missing Values

Missing data will be handled based on QLQ-C30 scoring manuals.

3.6 Safety Variables

Safety and tolerability will be assessed in terms of adverse events (AEs) (including serious adverse events [SAEs]), exposure, laboratory data, vital signs, ECGs, deaths, dose intensity, WHO/ECOG performance status, and physical examination. These will be collected and summarized for all patients in the safety analysis set with the exception of ECOG which will be summarized for the full analysis set.

3.6.1 Adverse Events (AE)

AEs and SAEs will be collected from the time of the patient signing the ICF until the follow-up period is completed (90 days after the last dose of treatment (Arm 1) or adjuvant phase study visits (Arm 2)). If an event that starts post the defined safety follow-up period noted above is considered to be due to a late onset toxicity to study drug then it should be reported as an AE or SAE as applicable.

Supporting AE summaries for AEs and SAEs occurring from date of informed consent until 90 days after completion of adjuvant phase study visit, and for AEs and SAEs occurring from date of radical cystectomy to 90 days post radical cystectomy, may also be provided to support the AE summaries. The latest version of the Medical Dictionary for Regulatory Activities (MedDRA) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 5.0).

3.6.1.1 Treatment Emergent Adverse Events (TEAEs)

A treatment emergent adverse event (TEAE) is an AE with an onset date or a pre-existing AE

worsening (by investigator report of a change in intensity) following the first dose of study treatment up to and including min (date of last dose of study treatment + 90 days, day before the first dose of subsequent anti-cancer therapy).

Any AE occurring before any study intervention and without worsening after initial of study treatment will be referred to as 'pre-treatment'.

3.6.1.2 Adverse Events of Special Interest (AESI) and AEs of possible interest (AEPI)

Some clinical concepts (including some selected individual preferred terms and higher-level terms) have been considered "AEs of special interest" (AESI) and "AEs of possible interest" (AEPI) to the durvalumab program. All AESIs are being closely monitored in clinical studies using durvalumab alone, and durvalumab in combination with other anti-cancer agents.

AESIs are defined as AEs that with a likely inflammatory or immune-mediated pathophysiological basis resulting from the mechanism of action of durvalumab and requiring more frequent monitoring and/or interventions such as corticosteroids, immunosuppressants, and/or endocrine therapy. Endocrine therapies include standard endocrine supplementation, as well as treatment of symptoms resulting from endocrine disorders (for example, therapies for hyperthyroidism include beta blockers [e.g., propranolol], calcium channel blockers [e.g., verapamil, diltiazem], methimazole, propylthiouracil, and sodium perchlorate). In addition, infusion-related reactions and hypersensitivity/anaphylactic reactions are also considered AESIs.

AEPIs are defined as AEs that could have a potential inflammatory or immune-mediated pathophysiological basis resulting from the mechanism of action of durvalumab but are more likely to have occurred due to other pathophysiological mechanisms, thus, the likelihood of the event being inflammatory or immune-mediated in nature is not high and/or is most often or usually explained by the other causes. These AEs not routinely arising from an inflammatory or immune-mediated mechanism of action – typically quite general clinical terms that usually present from a multitude of other causes –are classified as AEPIs.

These AESIs and AEPIs have been identified as Pneumonitis, Hepatic events, Diarrhea/Colitis, Intestinal perforations, Adrenal Insufficiency, Type 1 diabetes mellitus, Hyperthyroid events, Hypophysis, Hypothyroid events, Thyroiditis, Renal events, Dermatitis/Rash, Pancreatic events, Myocarditis, Myasthenia gravis, Guillain-Barre syndrome, Myositis, Infusion/hypersensitivity reactions and Other rare/miscellaneous. Other categories may be added, or existing terms may be merged as necessary. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician has reviewed the AEs of interest and identified which MedDRA preferred terms contribute to each AESI/AEPI. A further review will take place prior to Database lock (DBL) to ensure any new terms not already included in the older MedDRA version are captured within the categories for the new higher MedDRA version. The list will be provided by AZ prior to database lock.

3.6.1.3 Immune-mediated adverse events

Immune-mediated adverse drug reactions (imAEs) will be identified from both AESIs and 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). Endocrine therapies include standard endocrine supplementation, as well as treatment of symptoms resulting from endocrine disorders (for example, therapies for hyperthyroidism include beta blockers [e.g., propranolol], calcium channel blockers [e.g., verapamil, diltiazem], methimazole, propylthiouracil, and sodium perchlorate). Infusion-related reactions and hypersensitivity/anaphylactic reactions are exceptions because they are common to monoclonal antibody drugs in general and occur due to a mechanism of action different from that for imAEs, thus, these events are not considered imAEs as defined in the Durvalumab imAE charter. Identification of imAEs will be performed by the Sponsor and further details are provided in the Sponsor Durvalumab 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.

3.6.1.4 Other Significant Adverse Events (OAEs)

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 AEs leading to discontinuation. Based on the expert's judgment, significant AEs of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered as 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.

3.6.2 Treatment Exposure

Table 9 Total and Actual Exposure

Study Treatment	Total (or intended) Exposure	Actual Exposure
Durvalumab	Neoadjuvant and Non-cystectomy extension phase total (or intended) exposure: Minimum of (last dose date where dose>0 mg + ZZ days, date of death, date of DCO) – first dose date in neoadjuvant treatment period +1.	Neoadjuvant phase total (or intended) exposure – total duration of dose delays in neoadjuvant phase + adjuvant phase total (or intended) exposure – total duration of dose delays in adjuvant phase, where Minimum of (last dose date

Study Treatment	Total (or intended) Exposure	Actual Exposure
	ZZ = 20 if last dose is in neoadjuvant phase, ZZ = 27 if last dose is non-cystectomy extension. Adjuvant phase total (or intended) exposure: Minimum of (last dose date where dose>0 mg + 27 days, date of death, date of DCO) – first dose date in adjuvant treatment period +1	where dose>0 mg + ZZ days, date of death, date of DCO) – first dose date in neoadjuvant treatment period +1. ZZ = 20 if last dose is in neoadjuvant phase, ZZ = 27 if last dose is non-cystectomy extension phase And the total duration of dose delays in adjuvant phase = sum of (date of delayed dose – date of previous dose – 28 days)
Cisplatin	**Patient with adequate renal function:** Minimum of (last dose date where dose>0 mg + 20 days, date of death, date of DCO) – first dose date+1 **Patient with borderline renal function:** If CxD1 in 21 day cycle: Minimum of (last dose date where dose>0 mg + 6 days, date of death, date of DCO) – first dose date+1 If CxD8 in 21 day cycle: Minimum of (last dose date where dose>0 mg + 13 days, date of death, date of DCO) – first dose date+1	Total (or intended) exposure – total duration of dose delays, where the total duration of dose delays = sum of (date of delayed dose – date of previous dose – 21 days)
Gemcitabine	If CxD1 in 21 day cycle: Minimum of (last dose date where dose>0 mg + 6 days, date of death, date of DCO) – first dose date+1 If CxD8 in 21 day cycle: Minimum of (last dose date where dose>0 mg + 13 days, date of death, date of DCO) – first dose date+1	Total (or intended) exposure – total duration of dose delays, where the total duration of dose delays = sum of (date of delayed dose – date of previous dose – 21 days)

Exposure to study drug(s), time on study, dose delays, and dose interruptions will be summarized for all treatment arms.

Total exposure, actual exposure, and time on study expressed in months will be summarized in tables. The duration in months will be calculated as follows:

$$\frac{Duration\ in\ days}{365.25 \div 12}.$$

Exposure will also be measured by the number of cycles received. A cycle in neoadjuvant phase corresponds to a period of 21 days and a cycle in adjuvant phase corresponds to a period of 28 days. If a cycle is prolonged due to toxicity, 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.

3.6.3 Dose Intensity

Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. RDI will be defined as follows:

$$RDI = 100\% \times \frac{d}{D},$$

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 should be added to the date of last dose to reflect the protocol allowed window for dosing. When deriving actual dose administered, the volume before and after infusion will also be considered.

3.6.4 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 hematology and clinical chemistry will be collected as described in section 8.8.2 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 4.2.7 below will be used.

Change from baseline in hematology and clinical chemistry variables will be calculated for each post-dose visit on treatment. Common toxicity criteria (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, calcium and corrected calcium. For these parameters, high and low CTC grades will be calculated.
Corrected calcium will be derived during creation of the reporting database using the following formula:

$$Corrected\ calcium\ (mmol/L) = Total\ calcium\ (mmol/L) + ([40 - albumin\ (g/L)] * 0.02)$$

Absolute values will be compared to the project reference ranges and classified as low (below the lower limit of reference range), normal (within reference range, upper and lower limit included) and high (above upper limit of reference 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 on treatment (defined as between the start of study treatment and up to and including the earlier of 90 days following the date of last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2) or the date of initiation of the first subsequent anti-cancer therapy).

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

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

3.6.5 ECGs

ECG will be recorded at screening and as clinically indicated throughout the treatment and disease assessment period. The following ECG variables will be collected in the eCRF: PR interval, QRS duration, QT interval and QTcF interval. All ECGs will be assessed by the investigator as to whether they are clinically significantly abnormal. (Any clinically significant abnormalities detected require triplicate ECG results, and triplicate ECG results are also required at screening.)

Post-baseline data obtained up until 30 days following discontinuation of study treatment (or the last adjuvant visit for Arm 2) or the date of initiation of subsequent anti-cancer therapy (whichever occurs first), will be considered as “on-treatment”. On-treatment results will be included in the summaries.

3.6.6 Vital Signs and Measurements

Post-baseline data obtained up until 30 days following discontinuation of study treatment (or the last adjuvant visit for Arm 2) or the date of initiation of subsequent anti-cancer therapy (whichever occurs first), will be considered as “on-treatment”. On-treatment results will be included in the summaries.

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 4.2.7 below will be used.

Body surface area is calculated at the beginning of each cycle in the neoadjuvant phase only, for the purpose of chemotherapy dose calculations.

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

3.6.7 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 (Arm 1) or last adjuvant visit for Arm 2) + 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.

3.6.8 Clavien-Dindo Assessment

Clavien-Dindo assessment will be utilized for grading surgical complications. The highest-grade complication, which occurs within 90 days after the radical cystectomy, will be recorded at the time specified in the SoAs (Table 2 and Table 3). Investigators will indicate which AE resulted in a surgical complication grade reported. The following classification will be used:

Grade	Definition
Grade 0	No event observed
Grade I	Any deviation from the normal post-operative course without the need for pharmacological treatment or surgical, endoscopic, and radiological interventions. Allowed therapeutic regimens are drugs as anti-emetics, antipyretics, analgesics, diuretics, electrolytes, and physiotherapy. This grade also includes wound infections opened at the bedside.
Grade II	Requiring pharmacological treatment with drugs other than such allowed for Grade I complications. Blood transfusions and total parenteral nutrition are also included.
Grade III	Requiring surgical, endoscopic, or radiological intervention
Grade IIIa	Intervention not under general anesthesia
Grade IIIb	Intervention under general anesthesia
Grade IV	Life-threatening complication (including CNS complications)a requiring IC/ICU management
Grade IVa	Single organ dysfunction (including dialysis)
Grade IVb	Multiorgan dysfunction
Grade V	Death of a patient

[a] Brain hemorrhage, ischemic stroke, subarrachnoidal bleeding, but excluding transient ischemic attacks.
Source: Dindo et al 2004
CNS central nervous system; IC intermediate care; ICU intensive care unit.

3.7 Pharmacokinetic and Immunogenicity Variables

Pharmacokinetic concentration data and immunogenicity data will be collected as per the protocol.

Analyses to evaluate the pharmacokinetics (PK) and immunogenicity of durvalumab will be performed by AstraZeneca/Medimmune Clinical Pharmacology group or designee.

3.8 Pharmacogenetic Variables

In the case of genetic data, only the date that the patient gave consent to participation in the genetic research and the date the blood sample was taken from the patient will be recorded in the eCRF and database. The genetic data generated from the study will be stored in the AstraZeneca Laboratory Information Management System (LIMS) database or other appropriate system. This database is a secure database that is separate from the database used for the main study. Some or all the dataset from the main study may be duplicated within the AstraZeneca LIMS database for exploratory genetic analysis. Data will be reported outside of the CSR (please see Appendix C of the CSP).

3.9 Biomarker Variables

PD-L1 expression status (high, low/negative) is defined in **Table 10** below. Exploratory analyses based on different definitions of PD-L1 expression may be performed based on emerging data.

PD-L1 status is determined by the percentage of tumor cells with any membrane staining above background or by the percentage of tumor-associated immune cells with staining (IC+) at any intensity above background. Percent of tumor area occupied by any tumor-associated immune cells (ICP) is used to determine IC+, which is the percent area of ICP exhibiting PD-L1 positive immune cell staining.

Table 10 PD-L1 Status Defined by VENTANA PD-L1 (SP263) Assay

Interpretation	Staining Description
PD-L1 High	≥ 25% of tumor cells exhibit membrane staining; OR ICP > 1% and IC+ ⩾ 25%; OR ICP =1% and IC+ = 100%.
PD-L1 Low/Negative	< 25% of tumor cells exhibit membrane staining; AND ICP > 1% and IC+ < 25%; AND ICP =1% and IC+ < 100%.

IC+ immune cell with staining; ICP immune cell present;

4 ANALYSIS METHODS

The formal statistical analysis will be performed to test the following hypothesis:

- H_0: No difference between durvalumab + G+C combination therapy (neoadjuvant)/durvalumab alone (adjuvant) (Arm 1) compared to G+C combination therapy (neoadjuvant)/no adjuvant treatment (Arm 2)
- H_1: Difference between durvalumab + G+C combination therapy (neoadjuvant)/durvalumab alone (adjuvant) (Arm 1) compared to G+C combination therapy (neoadjuvant)/no adjuvant treatment (Arm 2)

4.1 General Principles

The below mentioned general principles will be followed throughout the study:

- 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, upper and lower quartiles, minimum, and maximum. Categorical variables will be summarized by frequency counts and percentages for each category.
- For log-transformed data, it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum, and maximum.
- Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment arm.
- All outputs will be summarized by treatment arm for all randomized patients (FAS) or safety analysis set and where required.
- 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.
- SAS® version 9.4 or above will be used for all analyses.

Efficacy and HRQoL data will be summarized and analyzed based upon the FAS. Safety and treatment exposure data will be summarized based upon the safety analysis set. Study population and demographic data will be summarized based upon the FAS.

Baseline

In general, for efficacy and PRO endpoints, the last observed measurement prior to randomization will be considered 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. A “Neoadjuvant Baseline” radiological tumor assessments are to be performed no more than 42 days before the date of randomization and ideally as close as possible to randomization (see CSP Table 1). A post-neoadjuvant/ pre-radical cystectomy follow-up scan must perform upon completion of neoadjuvant chemotherapy

prior to surgery. An "Adjuvant Baseline" scan should be collected 42 days (±14 days) after radical cystectomy and ideally should be performed as close as possible and prior to the first date of adjuvant phase (see CSP Table 2).

The PRO endpoints are scheduled to be collected on the first day of randomized treatment; these data will be used as baseline provided they are collected on or before the first day of study treatment. For safety endpoints, the last observation before the first dose of study treatment will be considered 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:

$$post\ baseline\ value\ -\ baseline\ value.$$

The percentage change from baseline will be calculated as:

$$\frac{post\ baseline\ value\ -\ baseline\ value}{baseline\ value} \times 100.$$

Radiological efficacy will be assessed by RECIST 1.1. There will be 2 baseline assessments, the first for the neoadjuvant phase and the second for the adjuvant phase. A first "Neoadjuvant Baseline" scan should be collected during pre-randomization screening (Day -28 to -1) for disease staging and for use as a RECIST 1.1 baseline for the post-neoadjuvant/pre-radical cystectomy follow-up scans.

A second "Adjuvant Baseline" scan should be collected 42 days (±2 weeks) after radical cystectomy and ideally should be performed as close as possible and must be prior to the first date of adjuvant treatment. In most instances, no lesions will be observed on the Adjuvant Baseline scans and 'No Evidence of Disease' will be recorded for the Adjuvant Baseline RECIST assessment. If an adjuvant baseline scan is not recorded, it will be considered that no lesions are presented following surgery (however, this will not count as a completed visit for the purposes of the 2 or more consecutive missed visit assessment).

4.2 Analysis Methods

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

Table 11 details the endpoints that are subject to formal statistical analysis, together with pre-planned sensitivity analyses, making it clear which analysis is regarded as primary for that endpoint.

Table 11 Formal Statistical Analyses to be Conducted and Pre-Planned Sensitivity analyses

Endpoint Analyzed	Notes
pCR rate	Logistic regression adjusted for the stratification factors, odds ratio and the corresponding confidence interval, and p-value: Dual primary analysis using central pathology review: • Arm 1 versus Arm 2 (FAS) Secondary analysis using central pathology review: • Arm 1 versus Arm 2 (PD-L1 High Population) Secondary analysis using Investigator assessment: • Arm 1 versus Arm 2 (FAS population) • Arm 1 versus Arm 2 (PD-L1 High Population)

EFS	Stratified log-rank test to obtain the p-value, stratified Cox PH model to obtain the hazard ratio and the corresponding confidence interval: Dual primary analysis using BICR or by central pathology review if a biopsy is required for a suspected new lesion: • Arm 1 versus Arm 2 (FAS) Secondary analysis using BICR or by central pathology review if a biopsy is required for a suspected new lesion: • Arm 1 versus Arm 2 (PD-L1 High Population) Sensitivity analysis for primary and secondary using BICR or by central pathology review if a biopsy is required for a suspected new lesion: o Arm 1 versus Arm 2 (FAS): Excluding the PD-L1 stratification factor from the stratified log-rank test and stratified Cox PH model ▪ Subsequently add TC1 or TC25 separately (2 models) as categorical covariates in the model o Using a KM plot of time to censoring where the censoring indicator of the primary analysis is reversed – attrition bias (FAS) o Interval censored analysis – evaluation time bias (FAS) o Analysis where subjects who take subsequent anti-cancer therapy prior to the EFS event will be censored at their last evaluable assessment prior to taking the subsequent therapy – attrition bias (FAS) o Analysis using the 2 missed visit censoring rules – attrition bias (FAS) o Analysis using alternative censoring rules – no adjuvant baseline (FAS) o Sensitivity analysis to assess impact of COVID-19 deaths (FAS)

Endpoint Analyzed	Notes
	Secondary analysis per local Investigator or Investigator biopsy review if a biopsy is required for a suspected new lesion: • Arm 1 versus Arm 2 (FAS) • Arm 1 versus Arm 2 (PD-L1 High Population)

Endpoint Analyzed	Notes
EFS24	Kaplan-Meier estimates of event-free survival rate at 24 months by treatment Secondary analysis using BICR or by central pathology review if a biopsy is required for a suspected new lesion: • Arm 1 versus Arm 2 (FAS) Secondary analysis per local Investigator or Investigator biopsy review if a biopsy is required for a suspected new lesion: • Arm 1 versus Arm 2 (FAS)
Proportion of patients who achieve <P2	Logistic regression adjusted for the stratification factors (FAS)
Overall survival OS5	Stratified log-rank test for OS (FAS) Kaplan-Meier estimates of survival rate at 5 years by treatment arm (FAS)
Proportion of patients who undergo cystectomy	Point estimate and 95% CI (FAS)
DSS	Stratified log-rank test (FAS)
DFS	Stratified log-rank test (Cystectomy population)
MFS	Stratified log-rank test (FAS)
PFS2	Stratified log-rank test (FAS)
EORTC QLQ-C30 endpoints	Average change from baseline using a MMRM analysis (FAS)
Time to definitive/sustained clinically meaningful deterioration (EORTC QLQ-C30)	Stratified log-rank test (FAS)

BICR Blinded Independent Central Review; CI Confidence interval; DFS Disease-free survival; DSS Disease specific survival; EFS Event-free survival; EFS24 Proportion of patients alive and event free at 24 months; EORTC QLQ-C30 European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire; ITT Intent-to-treat; MFS Metastasis-free survival; MMRM Mixed-model for repeated-measures; OS5 Proportion of patients alive at 5 years; pCR Pathologic complete response. PFS2 The time from the date of randomization to the earliest date of progression which occurs on subsequent therapy following an EFS event, or death.

4.2.1 Multiplicity

To strongly control the type I error at the 5% 2-sided alpha level, a MTP with gatekeeping strategy will be used across the dual primary endpoints (pCR rate and EFS). If the higher level

hypothesis in the MTP is rejected for superiority, the following hypothesis will then be tested as shown in Figure 2.

Hypotheses will be tested using a multiple testing procedure with an alpha-exhaustive recycling strategy (Burman et al 2009). With this approach, hypothesis will be tested in a pre-defined order by first splitting the 5% alpha into 0.1% and 4.9% for pCR and EFS for Arm 1 versus Arm 2, respectively, in patients in the FAS as outlined in Figure 2.

All key hypotheses are defined as follows:

- $H_{0.1}$: pCR, Arm 1 vs Arm 2
- $H_{0.2}$: EFS, Arm 1 vs Arm 2
- $H_{0.3}$: OS, Arm 1 vs Arm 2
- $H_{0.4}$: OS5, Arm 1 vs Arm 2

The details on how the alpha will be spent/controlled in all the possible scenarios are outlined below:

1. Test $H_{0.1}$: pCR for Arm 1 vs Arm 2 (FAS) at 0.1%.
1.1. If $H_{0.1}$ is not significant at 0.1% level then accept H0.1 and go to step 2. If $H_{0.1}$ is significant at 0.1% level then reject $H_{0.1}$ and continue to test $H_{0.2}$ at 5% level.
1.2. If $H_{0.2}$ is not significant at 5% level then accept $H_{0.2}$ and stop the procedure. If $H_{0.2}$ is significant at 5% level then reject $H_{0.2}$ and continue to test $H_{0.3}$ at 5% level.
1.3. If $H_{0.3}$ is not significant at 5% level then accept $H_{0.3}$ and stop the procedure. If $H_{0.3}$ is significant at 5% level then reject $H_{0.3}$ and continue to test $H_{0.4}$ at 5% level.
1.4. If $H_{0.4}$ is not significant at 5% level then accept $H_{0.4}$ and stop the procedure. If $H_{0.4}$ is significant at 5% level then reject $H_{0.4}$.

2. Test $H_{0.2}$: EFS for Arm 1 vs Arm 2 (FAS) at 4.9%.
2.1. If $H_{0.2}$ is not significant at 4.9% level then accept $H_{0.2}$ and stop the procedure. If $H_{0.2}$ is significant at 4.9% level then reject $H_{0.2}$ and continue to test $H_{0.3}$ at 4.9% level.
2.2. If $H_{0.3}$ is not significant at 4.9% level then accept $H_{0.3}$ and stop the procedure. If $H_{0.3}$ is significant at 4.9% level then reject $H_{0.3}$ and continue to test $H_{0.4}$ at 4.9% level.
2.3. If $H_{0.4}$ is not significant at 4.9% level then accept $H_{0.4}$ and stop the procedure. If $H_{0.4}$ is significant at 4.9% level then reject $H_{0.4}$.

The primary endpoint of pCR (FAS) will be tested at 1 time point (final analysis of pCR), and the primary endpoint of EFS (FAS) will be tested at 3 time points: 2 interim analyses and 1 final analysis. The alpha level allocated to EFS will be controlled at the superiority interims 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. There will be two superiority interim analyses for EFS: the first when the pCR analysis is conducted and the second when approximately 410 EFS events (39% maturity,

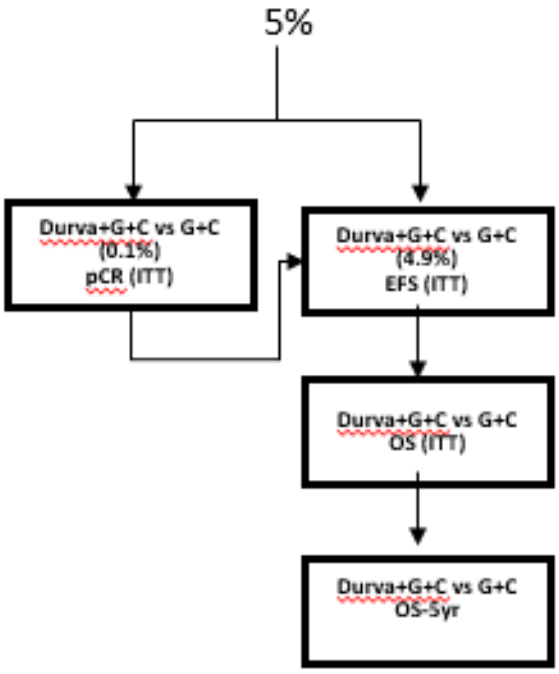
80% target events) have occurred across the 2 arms in the FAS, or in April 2024 whichever occurs first (approximately 31 months after the last patient is randomized).

The final analysis of EFS will be performed when 451 EFS (43% maturity) events in patients across the 2 arms in the FAS, or in June 2025, whichever occurs first (approximately 45 months after the last patient is randomized).

The first interim analysis has been performed with 301 events and the 2-sided alpha of 0.69% (calculated assuming 509 total events as stated in the previous protocol) has been spent. Applying this alpha spend at the first interim analysis and considering the revised total of 451 events, if exactly 91% of the target 451 events are available at the time of the second interim analysis, with overall 2-sided alpha level of 4.9%, the 2-sided alpha to be applied at the second interim analysis, and final analysis would be, 3.5%, and 3.9%, respectively.

However, the derivation of actual rejection boundary for any interim analysis will use the observed number of events at the interim analysis and the number of events planned for the final analysis. For the planned final analysis, the rejection boundaries will be derived based on the observed number of events and previous rejection boundaries using the generalized Haybittle-Peto method (SAS manual), exhausting any remaining alpha for the analysis.

Figure 2 Multiple Testing Procedures for Controlling the Type I Error Rate



Durva Durvalumab; EFS Event-free survival; G+C Gemcitabine+cisplatin; ITT Intention to treat; OS5 Proportion of patients alive at 5 years; pCR Pathologic complete response; vs versus.

4.2.2 Dual primary endpoint: pathologic complete response (pCR)

The pCR per central pathology review will be compared between two arms using logistic regression models adjusting for stratification factors (renal function [adequate vs borderline], tumor stage [T2 versus >T2] and PD-L1 status [high versus low/negative]) as covariates in the model based on patients in the FAS. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favor Arm 1) together with its associated profile likelihood 99.9% and 95% 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). The covariates in the statistical modeling will be based on the values entered in the IVRS at randomization.

If there are not enough responses for a meaningful analysis using logistic regression, then a CMH test will be presented.

Subgroup Analyses

Subgroup analyses will be conducted comparing pCR between arms in the following subgroups of patients in the FAS including, but not limited to:

- Sex (male versus female)
- 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])
- Prior Bacillus Calmette-Guerin therapy (Yes versus No)
- Age at randomization (<65 years versus ≥65 years)
    - This will be determined from the date of birth (BIRTHDAT in the DM module) and date of randomization (IERNDDATin 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 (i.e. for those countries where year of birth only is given) will have an assumed date of birth of 1st Jan [given year])
- Lymph node positive (N0 versus N1)
- Tumor stage (T2N0 versus >T2N0) at baseline per IVRS

- All visible tumor removed during the TUBRT procedure prior to study entry (Yes versus No)
- PD-L1 status (high, low/negative) per IVRS
- Race (white versus non-white)
- TC25 (TC≥ 25% versus TC<25%) and TC1 (TC≥1% versus TC<1%)

For each subgroup, the odds ratio and the corresponding 95% CI will be calculated from logistic regression models with treatment and the factor (only the factor that determines the subgroup). A forest plot, including the odds ratio 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.

If there are too few patients in the certain categories of the subgroup, a combination of some categories may be applied.

Secondary Analyses

An analysis of pCR rate per the central pathology review will also be based on the PD-L1 High population. The logistic regression model for the PD-L1 High population will be adjusted for tumor stage and renal function.

In addition, a secondary analysis, using pCR rate per local Investigator assessment, will be performed in patients in the FAS and the PD-L1 High population. The secondary analyses will use the same methodology as for the analysis described above.

4.2.3 Dual primary endpoint: Event-free survival (EFS)

The dual primary analysis of EFS will be performed to assess the efficacy between two arms in patients in the FAS per BICR or by central pathology review if a biopsy is required for a suspected new lesion.

The EFS will be analyzed using stratified log-rank test adjusted for the stratification (tumor stage, renal function, and PD-L1) for generation of the p-value .

The HR and its corresponding [1- adjusted alpha]x100% and 95% CI will be estimated from the stratified Cox proportional hazards model (with ties = Efron 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.

KM plots of EFS will be presented by treatment arm. Summaries of the number and percentage of patients experiencing an EFS event and the type of event (recurrence of disease post cystectomy, progression in patients who were precluded for cystectomy, or time of expected surgery in patients who refuse to undergo a radical cystectomy or failure to undergo a radical cystectomy in participants with residual disease, or death due to any cause in the absence of other EFS events). Those still in survival follow-up, those lost to follow up, and those who withdrew consent will be provided along with the median EFS for each treatment arm.

The assumption of proportionality will be assessed firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time-dependent covariate (i.e. treatment-by-time or treatment-by-ln(time) interaction) 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 (0-6m, 6-12m, etc.). 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. The stratified max-combo test will be considered as a sensitivity analysis on the EFS data in the FAS, to test for treatment differences as a robust test less reliant on the proportional hazards assumption.

Sensitivity Analyses for Primary Endpoint

Sensitivity analyses will be performed to assess possible evaluation-time bias and attrition bias for patients in the FAS.

Evaluation-time bias:

Two separate supplementary analyses will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled time points. In the first supplementary analysis, if the patient has progressed or experienced recurrent disease or died directly preceded by 2 or more consecutive missed visits, the patient will be censored at the time of the latest evaluable disease assessment prior to the consecutive missed visits (Note: NE visit is not considered a missed visit) (See section attrition bias 3).

In the second supplementary analysis, for these patients who missed two consecutive missed visits prior to EFS events, EFS will be interval censored, where the beginning of the interval is the last evaluable assessment where the patient was known to be event free and the end of the censoring interval is the time of the first assessment where the event was detected. For any other EFS events, will use the exact observation. In this analysis patients not experiencing EFS will be right-censored at the date of the last disease assessment. The HR and its 95% CI and p-value will be presented. The HR and CI are estimated from a proportional hazards model including stratification factors as categorical covariates in the model for interval censored data (Finkelstein 1986) using a piecewise constant hazard rate model for the baseline hazard function. A log-rank test for interval censored data using Sun's weights (Sun 1996) stratified by the same stratification

factors as in the primary analysis will also be undertaken to test for a difference in the EFS time between treatment groups. The EFS probabilities (including median) will be estimated using the EMICM algorithm (Wellner et al 1997) and will be plotted against time by treatment group.

Attrition bias 1:

Using a Kaplan-Meier plot of time to censoring where the censoring indicator of the EFS will be reversed.

Attrition bias 2:

Subsequent anti-cancer therapy: The primary analysis for EFS will be repeated, but the censoring rule will be modified so that patient who take subsequent therapy prior to EFS event will be censored at the last evaluable assessment prior to taking anti-cancer therapy.

Attrition bias 3:

Analysis using the 2 missed visit rule: The primary analysis for EFS will be repeated, but if the patient progresses or experiences recurrent disease or dies directly preceded by 2 or more consecutive missed visits, the patient will be censored at the time of the latest evaluable disease assessment prior to the consecutive missed visits.

Given that there is only one RECIST 1.1 visit scheduled prior to surgery, the censoring of subjects due to 2 or more consecutively missed post-baseline RECIST 1.1 visits will only occur after the date of post-surgery. The two missed visits will be calculated from the date of surgery and defined as shown in **Table 12**. Based on the RECIST 1.1 assessment schedule for subjects with surgical intervention, the definition of 2 missed visits will change over time and is calculated as the protocolled time between 2 subsequent scans + the protocol allowed visit window for an early visit at the previous assessment + the protocol allowed visit window for a late visit at the expected assessment. For example, if the previous RECIST assessment is between study days 57 and 581 (i.e., week 8-83) then two missing visits will equate to 26 weeks since the previous RECIST assessment, allowing for early and late visits (i.e., 12 week interval + 1 weeks for an early assessment + 12 week interval + 1 week for a late assessment = 26 weeks).

For patients who are determined to be in complete clinical response and who enter into a noncystectomy extension phase with a plan for a potential delay cystectomy, the two missed visits will be calculated from the date of post neoadjuvant baseline scan and defined as shown in **Table 13**.

If an adjuvant baseline scan is not recorded, it will be considered that no lesions are presented following surgery, which is same with primary analysis for EFS. However, this will not count as a completed visit for the purposes of the 2 or more consecutive missed visit assessment.

Table 12 Definition of 2 missed RECIST visits for subjects with surgical intervention

Scheduled Assessments	Previous RECIST Assessment (Time from Surgery)	Two Missed Visits Window
No disease assessment after surgery, or assessment prior to Day 56	(<=Day 56)	12x2+1 = 25 weeks (175 days) after surgery*
Q12W ± 1 week for the first 24 months (up to week 96)	Week 8 - Week 83 (Day 57 – 581)	12x2+1+1 = 26 weeks (182 days)
	Week 83 - Week 95 (Day 582 – 665)	Average of 12 and 24 = 18 18x2+1+1 = 38 weeks (266 days)
Q24W ±1 weeks for 36 months (up to week 264)	Week 95 - Week 239 (Day 666 – 1673)	24x2+1+1 = 50 weeks (350 days)
	Week 239 - Week 263 (Day 1674 – 1841)	Average of 24 and 52 = 38 38x2+1 = 77 weeks (539 days)
Q52W thereafter	(≥ Day 1842)	52x2 = 104 weeks (728 days)

* Window, in this case only, is measured from date of surgery, as this is the first scan post surgery.

Table 13 Definition of 2 missed RECIST visits for patients who are determined to be in complete clinical response and who enter into a noncystectomy extension phase

Scheduled Assessments	Previous RECIST Assessment (Time from post neoadjuvant baseline scan)	Two Missed Visits Window
Q12w ± 1 week for the first 24 months (up to week 96)	<= Week 83 (<= Day 581)	12x2+1+1 = 26 weeks (182 days)
	Week 83 - Week 95 (Day 582 – 665)	Average of 12 and 24 = 18 18x2+1+1 = 38 weeks (266 days)
Q24w ±1 weeks for 36 months (up to week 264)	Week 95 - Week 239 (Day 666 – 1673)	24x2+1+1 = 50 weeks (350 days)
	Week 239 - Week 263 (Day 1674 – 1841)	Average of 24 and 52 = 38 38x2+1 = 77 weeks (539 days)
Q52w thereafter	(≥ Day 1842)	52x2 =104 weeks (728 days)

No adjuvant baseline bias:

Assessed by repeating the EFS analysis using the alternative censoring rules, i.e., if the patient has radical cystectomy and there is no scan within 120 days following the date of radical cystectomy and prior to the start of adjuvant treatment (Arm 1) or within the 120 days, regardless of timing relative to the first study visit (Arm 2), they will be censored at the date of radical cystectomy unless died within 120 days of radical cystectomy.

Sensitivity analyses will also be undertaken to assess the impact of taking account of PD-L1 on the overall result: An analysis of EFS in the FAS per BICR or by central pathology review if a biopsy is required for a suspected new lesion will be performed by removing the PD-L1 stratification factor from the stratified log-rank test and stratified Cox PH model (ie the model will be adjusted only for the stratification factors for tumor stage [T2 versus >T2)] and renal function [adequate vs borderline]). The HR with its corresponding 95% CI and the p-value will be estimated using the same approach as specified above for the primary analysis of EFS.

Two subsequent analyses will then be undertaken using this model (1) including TC1 as a categorical covariate in the model (2) including TC25 as a categorical covariate in the model.

Subgroup Analyses

Subgroup analyses will be conducted comparing EFS between the two arms. The subgroup analyses will be done for the FAS using the same subgroups as for pCR.

The purpose of these subgroup analyses is to assess the consistency of treatment effect across potential prognostic factors. For each subgroup, the HR and the corresponding 95% CI will be calculated from an un-stratified Cox PH model with 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.

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 EFS events available for a meaningful analysis of a particular subgroup (it is not considered as appropriate to present analyses where there were less than 20 events across treatment arms in a subgroup), the relationship between that subgroup and EFS will not be formally analyzed. In this case, only descriptive summaries will be provided.

No adjustment to the significance level for testing of the subgroup analysis will be made since all these analyses will be considered as supportive of the analysis of EFS.

For the analysis of EFS in the efficacy interim, see section 5.

Secondary Analyses

An analysis of EFS per investigator or local biopsy review if a biopsy is required for a suspected new lesion will be based on the FAS and the PD-L1 High populations. The stratified log-rank test and stratified Cox PH model on ITT will be adjusted for the three stratification factors (adjusting for tumor stage [T2 versus >T2)], PD-L1 status [high versus low/negative], and renal function [adequate vs borderline]). The stratified log-rank test and stratified Cox PH model for the PD-L1 high population will be adjusted for tumor stage and renal function. The HR, CI of HR, and the p-value will be estimated using the same approach as specified above for the primary analysis of EFS.

4.2.4 Secondary Analysis

4.2.4.1 EFS24

The EFS24 (per BICR or by central pathology review if a biopsy is required for a suspected new lesion) and the corresponding 95% CI from the Kaplan-Meier estimate will be summarized by treatment arm for subjects in the FAS.

4.2.4.2 Proportion of patients who achieve <P2

The primary analysis for proportion of patients who achieve <P2 will be based on assessment per local pathology review. The proportion of patients who achieve <P2 will be compared between Arm 1 and Arm 2 using logistic regression models adjusted for the stratification factors. The results of the analysis will be presented in terms of an odds ratio together with its corresponding 95% CI and p-value.

4.2.4.3 Overall Survival and OS5

The OS analysis will be performed in the FAS using a stratified log-rank test based on the same methodology as described for the EFS endpoint. The effect of Arm 1 versus Arm 2 will be estimated by the HR together with its corresponding (1-adjusted alpha) and 95% CI. The HR and CI will be estimated from the stratified Cox proportional hazards model (Cox 1972). Kaplan-Meier plots of OS will be presented by treatment arm. Summaries of the number and percentage of subjects 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.
A sensitivity analysis for OS will examine the censoring patterns to rule out attrition bias with regard to the primary treatment comparisons, achieved by a Kaplan-Meier plot of time to censoring where the censoring indicator of OS is reversed.

A summary of the duration of follow-up will be summarized for all subjects as well as for censored subjects only, presented by treatment group. Additionally, summary statistics for the number of days from censoring to DCO for all censored subjects will be presented.

OS5 will be summarized using the same methodology as EFS24 (see section 4.2.4.1) for patients in the FAS.

4.2.4.4 MFS

MFS will be summarized using the same methodology as EFS (see section 4.2.3) for patients in the FAS. The effect of treatment will be estimated by the HR together with its corresponding 95% CI.

4.2.4.5 DSS

DSS will be summarized using the same methodology as EFS (see section 4.2.3) for patients in the FAS. The effect of treatment will be estimated by the HR together with its corresponding 95% CI.

4.2.4.6 Proportion of patients who undergo cystectomy

The proportion of patients who undergo radical cystectomy will be compared between Arm 1 and Arm 2 using logistic regression models adjusted for the stratification factors. The results of the analysis will be presented in terms of an odds ratio together with its corresponding 95% CI and p-value.

4.2.4.7 DFS

DFS (per BICR or by central pathology review if a biopsy is required for a suspected new lesion) will be analyzed in the cystectomy analysis set using a stratified log-rank test, using the same methodology as described for the dual primary EFS endpoint (see Section 4.2.3). The effect of treatment will be estimated by the HR together with its corresponding 95% CI. Kaplan-Meier plots of DFS will be presented by treatment arm.

4.2.4.8 PFS2

Time from randomization to second progression or death (PFS2) in the FAS population 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 will be presented to support the analysis. The sensitivity analysis outlined in Section 4.2.3 will not be repeated for PFS2 with the exception of a Kaplan-Meier plot of the time to censoring where the censoring indicator of PFS2 is reversed.

The number and percentage of patients experiencing a PFS2 event and the type of progression will also be summarised by treatment arm, as well as summaries of deaths in the absence of second progression, and categories of PFS2 censoring. Time from randomisation to second progression will be summarised by treatment arm.

4.2.5 Patient Reported Outcomes (PROs)

Primary PRO endpoints include EORTC QLQ-C30: physical functioning, global HRQoL and fatigue, and pain. The FAS will be used for all PRO endpoints.

For all inferential procedures, PRO endpoints will be tested at the 5% significance level and 95% CIs will be produced. Statistical analyses comparing treatment arms will include the following:

- Visit specific adjusted mean change from baseline scores (using MMRM) (EORTC QLQ-C30: global HRQoL, 5 functioning scales, multi-item symptoms).
- Overall (across all visits) adjusted mean change from baseline scores (using MMRM) (same endpoints as first bullet paragraph)
- Time to definitive/sustained deterioration (same endpoints as first bullet paragraph)
- Visit response (improvement, no change, and deterioration) (same endpoints as first bullet paragraph)
- Best overall response (improvement rates) (same endpoints as first bullet paragraph)

Compliance rates summarizing questionnaire completion overall and at each visit and form disposition will be tabulated for EORTC QLQ-C30, PGIC, PGIS, EQ-5D-5L and PRO-CTCAE.

4.2.5.1 EORTC QLQ-C30

The primary assessment of global health status/QoL, functioning, and symptoms will be focused on the adjusted mean change from baseline using a mixed-model for repeated-measures (MMRM) analysis of all the post-baseline scores for each visit. The model will include treatment, visit, treatment by visit interaction, and stratification variables as explanatory variables, and the baseline score as a covariate. Adjusted mean change from baseline estimates per treatment arm and corresponding 95% CIs will be presented along with an overall estimate of the treatment difference, 95% CI, and p-value.

Supportive analysis of global health status/QoL, functioning, and symptoms will be time to definitive or /sustained clinically meaningful deterioration analyzed using a stratified log-rank test as described for the primary EFS endpoint. Separate analysis will be conducted for global health status/QoL, functions, fatigue and pain. The effect between Arm 1 and Arm 2 will be estimated by the HR together with its corresponding CI and p-value. KM plots will be presented by treatment arm. Summaries of the number and percentage of patients who have an event as well as who were censored will be provided along with the medians for each treatment.

Summary tables of visit responses for each EORTC QLQ-C30 scale/item score (global HRQoL, 5 functions and multi-item symptoms) and for each visit (improvement, deterioration and no change) will be presented by treatment arm. In addition, summary tables of the best overall response will be provided for the following domains by treatment arm: multi-item symptoms, global HRQoL, functioning (physical, role, cognitive, social, and emotional) and fatigue.

Multi-item symptoms, global HRQoL and functioning (physical, role, cognitive, social, and emotional) improvement proportions based on best overall response will be compared between each treatment arm using a logistic regression model, controlling for the stratification factors. The odds ratio, p-value, and 95% CI will be presented.

Finally, summaries of absolute and unadjusted change from baseline values of each EORTC QLQ-C30 scale/item will be reported by visit for each treatment arm. Graphical presentations may also be produced as appropriate.

4.2.5.2 PGIC

The PGIC data will be summarized in tables and plotted as appropriate. The number (%) of patients with each level of response over time will be summarized.

4.2.5.3 PGIS

The PGIS data will be summarized in tables and plotted as appropriate. The number (%) of patients with each level of response at baseline and over time will be summarized.

4.2.5.4 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.

4.2.5.5 PRO-CTCAE

The PRO-CTCAE (see Appendix H of the CSP) data will be summarized in tables and plotted as appropriate. The number (%) of patients with each level of response for each PRO-CTCAE item at baseline and over time will be summarized. A bar chart of the incidence by visit may be presented for each PRO-CTCAE item. Further summaries to explore the data (i.e. the severity of symptoms) may be produced.

4.2.6 Safety

Safety and tolerability data will be presented by treatment arm using the safety analysis set. Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) will be reported. No formal statistical analyses will be performed on the safety data.

Data from all cycles of treatment will be combined or grouped by neoadjuvant, post-surgery, and adjuvant periods in the presentation of safety data. AEs (both in terms of Medical Dictionary for Regulatory Activities [MedDRA] preferred terms and CTCAE grade) will be listed individually by patient. The number of patients experiencing each AE will be summarized by treatment group and CTCAE grade. Additionally, data presentations of the rate of AEs per person-years at risk may be produced.

Other safety data will be assessed in terms of physical examination, clinical chemistry, hematology, vital signs, and ECGs. Exposure to durvalumab + G+C combination therapy (neoadjuvant) /durvalumab alone and G+C combination therapy (neoadjuvant)/no adjuvant treatment will be summarized. Dose delay/interruption in Arm 1 and Arm 2 will be summarized.

"On treatment" will be defined as assessments between date of start dose and 90 days following last dose, unless otherwise specified. For the majority of safety summaries, the period of time after the administration of subsequent therapy will not be considered "on treatment".

To support the safety summaries, data from date of informed consent until 90 days after completion of adjuvant phase study visit, and data from date of radical cystectomy to 90 days post radical cystectomy, may also be provided.

The following sections describe the planned safety summaries. However, additional safety tables may be required to aid interpretation of the safety data. By visit summaries will use visits windows described in below.

General Consideration for Safety Assessments

For any safety summaries by period, the following data will be included:

- **Neoadjuvant period**: Date of first dose of neoadjuvant study treatment until the date of surgery, or for subjects without surgery up to min (date of last dose of neoadjuvant treatment + 90 days, date of first dose of subsequent anti-cancer therapy, date of DCO). Note: For assessments recorded on the day of surgery, time will be used to determine if it's pre or post surgery, if time is not available it will be assumed to occur post surgery.
- **Post-surgery period**: Date of the day of surgery until min (date of surgery + 90 days, date of first dose of subsequent anti-cancer therapy, date of DCO). Note: Some subjects may have an overlap between their post-surgery period and their adjuvant period.
- **Adjuvant period:** Date of first dose of adjuvant study treatment (Arm 1) or date of first adjuvant study visit (Arm 2) until min (90 days after the last dose of adjuvant study

treatment (Arm 1) or last adjuvant study visit (Arm 2), date of first dose of subsequent anti-cancer therapy, date of DCO).

- **Overall period:** Date of first dose of study treatment until min (90 days after the last dose of treatment or surgery (Arm 1) (last dose of study treatment or date of surgery, whichever occurs later) or 90 days after the last neoadjuvant treatment, surgery or last adjuvant study visit (neoadjuvant treatment, date of surgery, or adjuvant study visit,, whichever occurs later) (Arm 2), date of first dose of subsequent anti-cancer therapy, date of DCO)

Time windows will need defining for any presentations that summarize 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 halfway 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 Arm 1 and Arm 2 are:

Neoadjuvant treatment period:

Day 1(C1D1), visit window 1 to 4
Day 8(C1D8), visit window 5 to 15
Day 22(C2D1), visit window 16 to 25
Day 29(C2D8), visit window 26 to 36
Day 43(C3D1), visit window 37 to 46
Day 50(C3D8), visit window 47 to 57
Day 64(C4D1), visit window 58 to 67
Day 71(C4D8), visit window 68 to 78

Adjuvant treatment period for Arm1

Day 1(C1D1), visit window 1 to 3
Day 29(C2D1), visit window 4 to 42
Day 57(C3D1), visit window 43 to 70
Day 85(C4D1), visit window 71 to 98
Day 113(C5D1), visit window 99 to 126
Day 141(C6D1), visit window 127 to 154
Day 169(C7D1), visit window 155 to 182
Day 197(C8D1), visit window 183 to 210

- 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 arm visit data should only be summarized if the number of observations is greater than the minimum of 20 and >1/3 of patients dosed. Footnote will be provided if these minimum criteria are not met.

- 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.

- Neoadjuvant 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.

4.2.6.1 Adverse Events

All AEs, both in terms of current MedDRA PT and CTCAE grade, will be summarized descriptively by count (n) and percentage (%) for each treatment arm. The latest MedDRA dictionary version will be used for coding. Excluding AEs after initiation of subsequent therapy will more accurately depict AEs attributable to study treatment only, AEs observed more than 90 days following the date of last dose of study treatment and after an initiation of subsequent therapy are likely to be attributable to subsequent therapy. Any pre-treatment AEs (i.e. AEs starting before the date of first dose of study treatment) that do not increase in severity after the first dose will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated).

AEs in adjuvant phase for patients in Arm 2 will be collected, although those patients will not take any study treatment in adjuvant phase. AE summary tables will also be produced containing AEs starting or increasing in severity up to and including 90 days following the date of last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2) or until the initiation of first subsequent therapy following discontinuation of treatment (whichever occurs first).

To assess the longer-term toxicity profile, an AE summary (by system organ class [SOC], PT and maximum reported CTCAE grade) will also be produced containing AEs starting or increasing in severity up to and including 90 days following the date of last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2), but without taking subsequent systemic anti-cancer therapy into account.

AE summary tables will be produced containing AEs starting from cystectomy up to 90 days after cystectomy.

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).

A separate data listing of AEs occurring more than 90 days after discontinuation study treatment (Arm 1) or last adjuvant phase study visit (Arm 2) will be produced. These events will not be included in AE summaries.

The following summaries present patient incidence (frequencies and percentages), counting each patient only once within each SOC and PT:

- All AEs
- All AEs possibly related to study treatment (as determined by the reporting investigator)
- AEs with CTCAE grade 3 or 4
- AEs with CTCAE grade 3 or 4, possibly related to study treatment (as determined by the reporting investigator)
- AEs with outcome of death

- AEs with outcome of death, possibly related to study treatment (as determined by the reporting investigator)
- Most common AEs
- Most common AEs with CTCAE grade 3 or 4
- All SAEs
- All SAEs possibly related to study treatment (as determined by the reporting investigator)
- AEs leading to discontinuation of study treatment
- AEs leading to discontinuation of study treatment, possibly related to treatment
- AEs leading to dose delay/interruption of study treatment
- Infusion reaction AEs (as determined by the reporting investigator)

Summaries of other significant AEs may be produced.

In addition, truncated AE tables of most common AEs by PT, and of most common AEs of CTCAE grade 3 or 4 by SOC and PT will be produced, including those events that occurred in at least 5% of patients in overall. This cut-off may be modified after review of the data. When applying the cut-off (i.e. x%), the raw percentage, without prior rounding applied, should be compared to the cut-off (i.e. an AE with frequency of 4.9% will not appear if the 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) including cystectomy period summed over patients and then multiplied by 365.25 x 100 to present in terms of per 100 patient years.

Deaths

A summary of all deaths and deaths on treatment or within 90 days of last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2) will be provided.

Adverse Events of Special Interest (AESIs) and Possible Interest (AEPI)

The list of PTs used to identify AESIs/AEPIs will be finalized prior to database lock (DBL) and documented in the Study Master File. Grouped summary tables of certain MedDRA PTs will be produced and may also show the individual PTs which constitute each AESI/AEPI grouping. Groupings will be based on PTs provided by the medical team prior to DBL, and a listing of the PTs in each grouping will be provided.

Summaries of the above-mentioned grouped AE categories will include number (%) of patients who have:

- Any ASEI/AEPI

- Any AESI/AEPI by grouped term, PT by outcome
- Any AESI/AEPI by grouped term, PT and maximum CTCAE grade
- Any AESI/AEPI possibly related to treatment
- Any AESI/AEPI leading to concomitant medication use (steroids)
- Any AESI/AEPI leading to concomitant medication use (high dose steroids)
- Any AESI/AEPI leading to concomitant medication use (endocrine therapy)
- Any AESI/AEPI leading to concomitant medication use (other immunosuppressants)
- At least one AESI/AEPI leading to discontinuation of treatment

Immune Mediated Adverse Events (imAEs)

Programmatically-generated immune mediated adverse events will be presented. Details of the programmatically generated immune mediated adverse event summaries will be confirmed before database lock.

The imAEs (as classified by the Sponsor) will also be summarized in the same manner as for the summaries for AESI/AEPI described above. See further details in the imAE Charter with respect to derivation rules.

Summary of Long-Term Tolerability

To assess the long-term tolerability, if there are a sufficient number of patients with events to warrant it, prevalence plots, life table plots and cumulative incidence plots will be presented for each of the AESI grouped terms 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 categorized by each day after dosing. The prevalence is plotted over time split by treatment arm. 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 can 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 split by treatment. 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, 2006). These plots will only be produced for AESIs that have ≥10 events.

4.2.6.2 Laboratory Assessments

Post baseline data obtained up until the safety follow-up are considered as "On treatment" and will be included in the summary tables. In addition, post baseline data obtained between the start of study treatment and up to and including the earlier of 90 days following last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2) or the date of initiation of the first subsequent anti-cancer therapy, will be used for the reporting of laboratory assessments. Excluding laboratory data after initiation of subsequent therapy will more accurately depict laboratory toxicities attributable to study treatment only, as toxicities observed more than 90 days following the date of last dose of study treatment are likely to be attributable to subsequent therapy. This will more accurately depict laboratory toxicities attributable to study treatment only as a number of toxicities up to 90 days following discontinuation of study medication are likely to be attributable to subsequent therapy.

To assess the longer-term toxicity profile, summaries of laboratory data may also be produced containing data collected up to and including 90 days following last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2), but without taking subsequent anti-cancer therapy into account.

Data summaries will be provided in international system of units (SI).

The following summaries will be provided for laboratory data:

- Absolute value and change from baseline for the minimum and maximum values (for quantitative measurements)

- Shift tables in hematology and clinical chemistry parameters from baseline to maximum CTCAE grade on treatment, indicating hyper- and hypo-directionality of change for electrolytes:

  Hematology: hemoglobin, leukocytes, lymphocytes (absolute count), neutrophils (absolute count), platelets

Clinical chemistry: ALT, AST, alkaline phosphatase (ALP), amylase, bicarbonate, creatinine, gamma-glutamyl transferase (GGT), lipase, total bilirubin, total protein, magnesium (high/low), sodium (high/low), potassium (high/low), calcium (high/low), corrected calcium (high/low), glucose (high/low)

- Incidence of CTCAE grade changes, presenting patients who had a shift of at least two grades from baseline, and patients who changed to grade 3 or 4 since baseline
- Shift tables in thyroid-stimulating hormone (TSH) from baseline to maximum and minimum value on treatment
- Scatter plots (shift plots) of baseline to maximum value / minimum value (as appropriate) on treatment, including or excluding outliers, may be produced for certain parameters if warranted after data review

Liver Enzyme Elevations and Potential Hy's law

To capture all elevated liver enzymes and potential Hy's law cases, the following summaries will be produced:

- Incidence of elevated ALT, AST, and total bilirubin during the study in the following categories:

  ALT, AST, or either ALT or AST: ≥ 3x – ≤ 5x, > 5x – ≤8x, > 8x – ≤ 10x, >10x – ≤ 20x and >20x the upper limit of normal (ULN)

  Total bilirubin: ≥2x – ≤3x, >3x – ≤5x, >5x ULN

  Potential Hy's law: (ALT or AST ≥3x ULN) and total bilirubin ≥2x ULN, where the onset date of the ALT or AST elevation should be prior to or on the date of the total Bilirubin elevation

- Scatter plots of ALT and AST (horizontal axis) versus total bilirubin (vertical axis) by treatment group with reference lines at 3x ULN for ALT and AST and 2x ULN for total bilirubin
- 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 assessment

In addition, liver biochemistry test results over time for patients with ALT or AST ≥ 3x ULN and total bilirubin ≥ 2x ULN together with ALT or AST ≥ 5x ULN (at any time on treatment) will be plotted. Individual patient-level data, presenting all assessments for patients who have ALT or AST ≥ 3x ULN and total bilirubin ≥ 2x ULN at any time on treatment, will be listed.

Assessment of Thyroid Function Test Results

The following summaries will include the number and percentage of patients who have elevated or low TSH.

- TSH > ULN
- TSH > ULN with TSH ≤ ULN at baseline
- TSH > 3 X ULN
- TSH > 3 X ULN with TSH ≤ ULN at baseline
- TSH > 10 X ULN
- TSH > 10 X ULN with TSH ≤ ULN at baseline
- TSH < LLN
- TSH < LLN with TSH ≥ LLN at baseline

A separate summary will present:

- Number of subjects with at least one baseline and post-baseline TSH result
    - On-treatment elevated TSH > ULN and above baseline
    - On-treatment decreased TSH < LLN and below baseline

- Grade change from baseline to on treatment minimum and maximum

Assessment of Renal Function Test Abnormalities

In addition to the analysis for serum creatinine, the number and percentage of patients with creatinine clearance (CrCl) with worst rate during treatment period meeting the following categories will be presented:

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

Creatinine clearance rate will be calculated using serum Creatinine and the Cockcroft-Gault formula (Cockcroft and Gault 1976).

4.2.6.3 ECG

Summaries of ECG will be provided as appropriate. Post baseline ECG data obtained up until the safety follow-up are considered as “on treatment” and will be included in the summary tables. ‘On treatment’ is defined as post baseline data obtained up until 30 days following last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2) or the date of initiation of subsequent anti-cancer therapy (whichever occurs first) will be included in the summary table. A shift table of baseline evaluation to worst evaluation “on treatment” may be produced.

4.2.6.4 Vital Signs

Post baseline vital sign data obtained up until the safety follow-up are considered as "on treatment" and will be included in the summary tables. For vital signs, 'On treatment' is defined as post baseline data obtained up until 30 days following last dose of study treatment (Arm 1) or last adjuvant phase study visit (Arm 2) or the date of initiation of subsequent anti-cancer therapy (whichever occurs first) will be included in the summary table. Summaries of systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate, temperature, respiratory rate and weight will be presented.

4.2.6.5 ECOG Performance Status

All Eastern Cooperative Oncology Group (ECOG) performance status will be summarized over time for patients in the FAS.

4.2.6.6 Clavien-Dindo Assessment

Clavien-Dindo assessment result will be summarized for patients in the cystectomy analysis set. The summary will include the number (%) of subjects in each response category.

4.2.7 Pharmacokinetic Concentration Data

PK concentration data will be listed for each patient and each dosing day, and a summary will be provided for all evaluable patients (PK analysis set).

4.2.7.1 Population Pharmacokinetics and Exposure-Response/Safety Analysis

A population PK model may be developed using a nonlinear mixed-effects modelling approach. If performed, the impact of physiologically-relevant patient characteristics (covariates) and disease on PK will be evaluated. Similarly, 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 outside of the CSR in a separate report, and therefore are not within the remit of this 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.

4.2.7.2 Pharmacokinetic Analysis

The PK analyses will be performed by AstraZeneca Clinical Pharmacology group or designee. PK concentration data and summary statistics will be tabulated by treatment and visit. Individual and mean blood concentration-time profiles will be generated. Samples below the lower limit of quantification (BLQ) will be treated as missing in the descriptive statistics, however if >50% samples are BLQ, the mean, median, geometric mean, and min will be set to BLQ, standard deviation, CV and Geometric CV will be set to NC.

The PK data collected in this study may be utilized with data from other studies for population PK and/or pharmacokinetic/pharmacodynamics analyses.

4.2.8 Immunogenicity Data

Summaries of immunogenicity data will be provided of the number and percentage of patients who develop detectable anti-durvalumab 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-durvalumab antibodies.

The effect of immunogenicity as well as the effect of its neutralizing properties on PK, pharmacodynamics, efficacy, and safety will be evaluated, if the data allow. A detailed plan will be written by the AstraZeneca Clinical Pharmacology group or designee.

4.2.9 Pharmacokinetic/Pharmacodynamic Relationships

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

4.2.10 Biomarker Data

The relationship, if applicable, exploratory biomarkers to clinical outcomes including but not restricted to pCR, EFS, EFS24, and OS5 will be presented for a subset of patients with adequate renal function and in the ITT analysis set who are evaluable for each biomarker.

PD-L1 expression determined by immunohistochemistry and 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.

4.2.11 Demographic and Baseline Characteristics Data

The following will be listed and summarized by randomized treatment group:

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations

- Inclusion in analysis populations
- Demographics:
    1. age
    2. age group (<50, ≥ 50 - < 65, ≥ 65 - < 75 and ≥ 75 years)
    3. sex (male, female)
    4. race
    5. ethnicity
- Patient characteristics at baseline:
    1. height
    2. weight
    3. weight group (<70, ≥70 - <90, and ≥ 90 kg)
    4. Body Mass Index (BMI)
    5. BMI group (18.5, ≥18.5 – 25.0, ≥25.0 - <30.0, ≥30.0 kg/m^2)
- Stratification factors according to the IVRS/IWRS
- Patient recruitment by region, country and centre
- Disease characteristics at study entry (primary tumor location, histology type, primary tumor stage and overall classification)
- Medical history
- Relevant surgical history
- Visual tumor be removed in prior TURBT (yes versus no)
- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer therapy
- Nicotine use, categorized (never, current, former)
- PD-L1 expression (high, low/negative, missing)
- ECOG (0 versus ≥ 1)

The medications will be coded following WHO Drug dictionary (as applicable).

4.2.12 Treatment Exposure

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

- Total exposure of each treatment arm
- Actual exposure of each treatment arm
- Total number of cycles received for each treatment group
- Number of, reasons for, and duration of dose delays/interruptions of durvalumab + G+C combination therapy (neoadjuvant phase) / durvalumab alone(adjuvant) versus G+C combination therapy(neoadjuvant phase) / no adjuvant treatment
- Number of infusions received
- RDI (relative dose intensity) of study medication

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

4.2.13 Subsequent Therapy

Subsequent therapies received after discontinuation of study treatment will have summaries produced by treatment arm, together with number of regimens received.

4.2.14 Concomitant and other treatments

Medications received prior to, concomitantly, or post-treatment will be coded using the AstraZeneca Drug Dictionary Anatomical Therapeutic Chemical (ATC) Classification codes. Concomitant medications will be summarized for the ITT population by ATC classification codes.

For the purpose of inclusion in prior and/or concomitant mediation or therapy summaries (including subsequent anti-cancer therapies), incomplete medication or radiotherapy start and stop dates will be imputed as

- For missing start dates, the following will be applied:
    a. Missing day – Impute the 1st of the month
    b. Missing month – Impute January
    c. Missing day and month – Impute 1st January
    d. Missing year – set as complete missing

- For missing end dates, the following will be applied:
    a. Missing day – Impute the last day of the month
    b. Missing Month – Impute December
    c. Missing day and month – Impute 31st December
    d. Missing Year – set as complete missing

Prior medications, concomitant and post- treatment medications are defined based on imputed start and stop dates as follows:

- Prior medications are those taken prior to the first dose of study treatment.
- Concomitant medications are those with a stop date on or after the first dose date of study treatment (and could have started prior to or during treatment).
- 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.
- Post-treatment medications are those with a start date after the last dose date of study treatment.

In addition, all post-treatment anti-cancer medications and surgical procedures will be summarized for the FAS population.

4.2.15 Coronavirus Disease 2019 (COVID-19)

Depending on the extent of any impact, summaries and listings of data relating to subjects diagnosed with COVID-19 and impact of COVID-19 on study conduct (in particular missed visits, delayed or discontinued study intervention, and other protocol deviations) may be generated. In addition, an EFS sensitivity analysis may be performed by repeating the summaries and analyses such that any subjects who had a death with primary/secondary cause as being COVID-19 related (including infection reported as fatal) will be censored at their COVID related death date. For AE and death, summaries of COVID-19 related events including infections and deaths may be produced.

5 INTERIM ANALYSES

An IDMC will be established to monitor data on an ongoing basis to ensure the continuing safety of patients enrolled in this study, to ensure the integrity of the study, and to oversee the planned EFS futility interim analysis. The IDMC will be composed of five independent experts. The committee will meet approximately 6 months after the first patient randomized or after the first 90 patients have been randomized, whichever occurs first. The subsequent IDMC meetings will occur in approximately 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.

EFS

EFS will be tested at two interim time points and a final time point.

For the dual-primary EFS endpoint in MIBC patients (Arm 1 versus Arm 2), two superiority Interim analyses for EFS will be conducted: the first when the pCR analysis is conducted and the second at a data cut-off when approximately 410 EFS events have occurred (39% maturity, 410/1050) across the 2 arms in FAS, or in April 2024 whichever occurs first. The later interim analysis is expected approximately 31 months after the last participant is randomized. The final analysis will be conducted when approximately 451 EFS events have occurred (43% maturity, 451/1050) across Arm 1 and Arm 2, or in June 2025, whichever occurs first. The final analysis is expected approximately 45 months after the last participant is randomized.

The alpha level allocated to the EFS will be controlled at the interim and final time points using the Lan DeMets 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 first interim analysis has been performed with 301 events and the 2-sided alpha of 0.69% (calculated

assuming 509 total events as stated in the previous protocol) has been spent. Applying this alpha spent at the first interim analysis and considering the revised total of 451 events, if exactly 91% of the target 451 events are available at the time of the second interim analysis, with overall 2-sided alpha level of 4.9%, the 2-sided alpha to be applied at the second interim analysis, and final analysis would be, 3.5%, and 3.9%, respectively. However, the derivation of actual rejection boundary for any interim analysis will use the observed number of events at the interim analysis and the number of events planned for the final analysis. For the planned final analysis, the rejection boundaries will be derived based on the observed number of events and previous rejection boundaries using the generalized Haybittle-Peto method (SAS manual), exhausting any remaining alpha for the analysis.

OS

OS will be tested at 2 interim time points and a final time point in accordance with the hierarchical multiple testing strategy.

The interim analysis of OS will be conducted at the time when the second EFS analysis is conducted (but will only be tested if EFS is positive via the MTP). A second interim analysis will be conducted at the time when the final EFS analysis is conducted. At the time of the first interim analysis, approximately 288 OS events are estimated to occur (27% maturity, 288/1050) across Arm 1 and Arm 2. At the time of the second interim analysis, approximately 318 OS events are estimated to occur (30% maturity, 318/1050) across Arm1 and Arm 2. The final analysis will be conducted when approximately 428 OS events have occurred (41% maturity, 428/1050) across Arm 1 and Arm 2, which is expected to be approximately 60 months after the last participant is randomized.

The alpha level allocated to the OS will be controlled at the interim and final time points using the Lan DeMets spending function that approximates an O'Brien Fleming approach, where the alpha level applied at the interim depends on the proportion of information available. If exactly 67% (288/428) and 74% (318/428) of the target events are available at the time of the first and second interim analysis, with overall 2-sided alpha level of 4.9%, the 2-sided alpha to be applied at the first, the second, and final analysis would be 1.2%, 1.5, and 4.3%, respectively.
The derivation of the actual rejection boundary for any interim analysis will use the observed number of events at the interim analysis and the number of events planned for the final analysis. For the planned final analysis, the rejection boundaries will be derived based on the observed number of events and previous rejection boundaries using the generalized Haybittle-Peto method (SAS manual), exhausting any remaining alpha for the analysis. Full details of the IDMC procedures, processes, and interim analyses can be found in the IDMC Charter.

6 CHANGES OF ANALYSIS FROM PROTOCOL

There is no change to analyses planned in the CSP.

7 REFERENCES

Basch et al 2009

Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationship with clinical outcomes. Journal of National Cancer Institute 2009;101:1624-1632

Breslow 1974

Breslow, NE. Covariance Analysis of Censored Survival Data. Biometrics. 1974; 30:89–99

Cockcroft and Gault 1976

Cockcroft DM and Gault MH. Prediction of creatinine from serum creatinine. Nephron. 1976. 16(1): 31-41

Danna et al 2016

Danna BJ, Metcalf MJ, Wood EL, Shah JB. Assessing symptom burden in bladder cancer: an overview of bladder cancer specific health-related quality of life instruments. BI Cancer 2016;2(3):329-40.

Eble et al 2004

Eble JN, Sauter G, Epstein JI, Sesterhenn IA, editors. Pathology and genetics of tumours of the urinary system and male genital organs. Lyon: IARCPRESS; 2004. (Available upon request).

EuroQol Group 1990

EuroQol-a new facility for the measurement of health-related quality of life. Health Policy 16(3):199-208.

Fayers et al 2001

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A; EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer: 2001.

Finkelstein 1986

Finkelstein DM. A proportional hazards model for interval-censored failure time data. Biometrics. 1986 Dec 1:845-54.

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., and R. Simon. "Testing for qualitative interactions between treatment effects and patient subsets." Biometrics (1985): 361-372.

Grossman et al 2003

Grossman HB, Natale RB, Tangen CM, Speights VO, Vogelzang NJ, Trump DL, et al. Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. N Engl J Med 2003;349:859-66.

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(24):4505-19.

Kim and DeMets 1987

Kim K, Demets DL. Design and analysis of group sequential tests based on the type I error spending rate function. Biometrika. 1987 Mar 1;74(1):149-54.

Lan and DeMets 1983

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983; 70 (3):659-663

Litwin et al 1998

Litwin MS, Lubeck DP, Henning JM, et al. Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: Results of the CaPSURE database. J Urol 1998;159:1988-92.

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.

Osoba et al 1998

Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16(1):139-44.

Pintilie 2006

Competing risks: A practical perspective. Wiley, 2006.

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 structural failure time models. Communications in Statistics-Theory and Methods 1991; 20(8):2609-31.

SAS manual

https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_seqdesign_details31.htm

Sprangers et al 1992

Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. J Clin Epidemiol. 1992;45:743-760.

Sun 1996

Sun J. A non-parametric test for interval-censored failure time data with application to AIDS studies. Statistics in medicine.1996 Jul 15;15(13):1387-95.

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(1):97-108.

Wellner et al 1997

Wellner JA, Zhan Y. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. Journal of the American Statistical Association. 1997 Sep 1;92(439):945-59.

Whitehead and Whitehead 1991

Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 1991;10(11):1665-77.

8 APPENDICES

Table 14 Region

Region	Country
East Europe	Poland, Russia
West Europe	Austria, Belgium, France, Netherlands, Germany, Spain, United Kingdom
North America and Australia	Canada, Australia, United States
East Asia	Japan