Statistical Analysis Plan	
Study Code	D9100C00001
Edition Number	3.0
Date	04 August 2021

A Phase III, Randomized, Multi-Center, Double-Blind, Global Study to Determine the Efficacy and Safety of Durvalumab in Combination With and following Chemoradiotherapy Compared to Chemoradiotherapy Alone for Treatment in Women with Locally Advanced Cervical Cancer (CALLA)

TABLE OF CONTENTS

TABLE OF CONTENTS 2
LIST OF ABBREVIATIONS 6
AMENDMENT HISTORY 10
1 STUDY DETAILS 22
1.1 Study objectives 22
1.1.1 Primary objectives 22
1.1.2 Secondary objectives 23
1.1.3 Exploratory objectives 25
1.2 Study design 26
1.3 Number of subjects 27
2 ANALYSIS SETS 28
2.1 Definition of analysis sets 28
2.1.1 Full analysis set (Intention to treat [ITT]) 29
2.1.2 Safety analysis set 29
2.1.3 Pharmacokinetics analysis set 29
2.1.4 PD-L1 analysis set 29
2.2 Violations and deviations 30
3 PRIMARY AND SECONDARY VARIABLES 33
3.1 Derivation of visit responses 33
3.1.1 RECIST 1.1 target lesions (TLs) 34
3.1.2 RECIST 1.1 non-target lesions and new lesions 39
3.1.3 Overall RECIST 1.1 visit response 41
3.1.4 Independent review 42
3.1.5 Histopathologic confirmation of local progression 43
3.1.6 Overall visit response 44
3.2 Efficacy variables 45
3.2.1 Progression-free survival (PFS) 45
3.2.1.1 Time to first and second subsequent therapy or death 47
3.2.2 Progression-free survival (3 years) 48
3.2.3 Overall survival (OS) 48
3.2.4 Objective response rate (ORR) 49
3.2.5 Complete response (CR) rate 50
3.2.6 Duration of response (DoR) for complete responders 50
3.2.7 Incidence of local progression, distant disease progression, and secondary malignancy 50
3.2.8 Best objective response (BoR) 52
3.3 Patient reported outcome (PRO) variables 52
3.3.1 EORTC-QLQ-C30 53
3.3.2 EORTC QLQ-CX24 54

3.3.3 Patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE) 55
3.3.4 Patient global impression of change (PGIC) and patient global impression of severity (PGIS) 55
3.3.5 Health state utility 56
3.3.6 Compliance 56
3.4 Health care resource use variables 57
3.5 Safety variables 58
3.5.1 General considerations for safety assessments 58
3.5.1.1 Definition of baseline 58
3.5.1.2 Time windows for safety and PRO data 58
3.5.1.3 Handling of missing data 60
3.5.1.4 Safety Follow-up 61
3.5.2 Study treatments 62
3.5.3 Exposure and dose interruptions 63
3.5.3.1 Treatment exposure for durvalumab or placebo 63
3.5.3.2 Treatment exposure for SoCs and Radiotherapy 64
3.5.3.3 Dose intensity 66
3.5.4 Adverse events (AEs) 67
3.5.5 Other significant adverse events (OAE) 68
3.5.6 AEs of special interest (AESI) and AEs of possible interest (AEPI) 68
3.5.7 Laboratory measurements 70
3.5.8 Vital signs 71
3.5.9 Physical examinations 71
3.5.10 Electrocardiograms (ECGs) 71
3.5.11 World Health Organisation (WHO)/Eastern Cooperative Oncology Group (ECOG) performance status (PS) 72
3.6 Pharmacokinetic (PK) variables 72
3.7 Immunogenicity variables 73
3.8 Biomarkers 74
3.9 Prior and concomitant medications and therapies 74
4 ANALYSIS METHODS 74
4.1 General principles 75
4.2 Multiplicity 77
4.3 Analysis methods 78
4.3.1 Efficacy 78
4.3.2 Primary efficacy endpoint 81
4.3.3 Secondary efficacy endpoints 90
4.3.3.1 Progression-free survival at 3 years (PFS (3 years)) 90
4.3.3.2 Progression-free survival in PD-L1 positive high subjects 90
4.3.3.3 Overall survival (OS) 90
4.3.3.4 Overall survival in PD-L1 high subjects 91

4.3.3.5 Objective response rate (ORR) 91
4.3.3.6 Complete response (CR) rate 92
4.3.3.7 Duration of response (DoR) for complete responders 92
4.3.3.8 Incidence of local progression, distant disease progression, secondary malignancy 92
4.3.3.9 Best objective response (BoR) 93
4.3.3.10 Patient reported outcomes 93
4.3.4 Exploratory endpoints 96
4.3.4.1 Health care resource use 96
4.3.5 Safety 96
4.3.5.1 Adverse events 97
4.3.5.2 Adverse events of special interest (AESI) and possible interest (AEPI) 102
4.3.5.3 Summary of long-term tolerability 103
4.3.5.4 Exposure 104
4.3.5.5 Laboratory measurements 104
4.3.5.6 Electrocardiograms 106
4.3.5.7 Vital signs 107
4.3.5.8 Eastern Cooperative Oncology Group performance status 107
4.3.6 Pharmacokinetic data 107
4.3.7 Immunogenicity analysis 107
4.3.8 Demographic, initial diagnosis and baseline characteristics data 108
4.3.9 Concomitant and other treatments 109
4.3.10 Subsequent therapy 109
4.3.11 Biomarker data 109
4.3.12 Genetic data 109
4.3.13 Radiotherapy Quality Control Assessment 110
4.4 COVID-19 110
4.4.1 Censoring Confirmed/Suspected COVID-19 Deaths for PFS Analysis 111
4.4.2 Censoring Confirmed/Suspected COVID-19 Deaths for OS Analysis 111
5 INTERIM ANALYSES 112
5.1 PFS and OS interim analyses 112
5.2 Independent data monitoring committee 112
6 CHANGES OF ANALYSIS FROM PROTOCOL (NOT APPLICABLE) 113
7 REFERENCES 113
8 APPENDIX (NOT APPLICABLE) 114

LIST OF TABLES

Table 1: Primary study objectives and corresponding endpoints/variables 22
Table 2: Secondary study objectives and corresponding endpoints/variables 23

Table 3: Exploratory study objectives and corresponding endpoints/variables 25
Table 4: Summary of outcome variables and analysis populations..... 28
Table 5: TL Visit Responses (RECIST 1.1) 35
Table 6: Example of scaling 38
Table 7: Non-target lesion visit responses..... 40
Table 8: Overall RECIST 1.1 visit responses..... 41
Table 9: Overall visit responses..... 44
Table 10: Visit responses for EORTC QLQ-C30..... 54
Table 11: Visit responses for EORTC QLQ-CX24..... 55
Table 12: Analysis visits and visit windows 59
Table 13: Study treatments..... 62
Table 14: Dose intensity scenarios for durvalumab administered 1500mg per cycle 67
Table 15: Formal statistical analyses to be conducted and pre-planned sensitivity analyses.. 79
Table 16: Discontinuation example..... 94

LIST OF FIGURES

Figure 1: Overview of study design 27
Figure 2: Multiple testing procedures for controlling the type I error rate..... 77

LIST OF ABBREVIATIONS

Abbreviation or special term	**Explanation**
ADA	Antidrug antibody
AE	Adverse event
AEPI	Adverse event of possible interest
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AUC	Area under the curve
AZ	AstraZeneca
BICR	Blinded independent central review
BoR	Best objective response
BP	Blood pressure
Brachy	Brachytherapy
CCRT	Concurrent chemoradiation therapy
CD	Cluster of differentiation
CI	Confidence interval
CMH	Cochran-Mantel Haenszel
CR	Complete response
CrCl	Creatine clearance
CRO	Contract research organization
CRF	Case report form
CSP	Clinical study protocol
CSR	Clinical study report
CT	Computed tomography
CTCAE	Common terminology criteria for adverse event
ctDNA	Circulating tumor deoxyribonucleic acid
CV	Coefficient of variation
DAE	Discontinuation of investigational produce due to adverse event
DBL	Database lock
DCO	Data cut-off
DoR	Duration of response
d.p.	Decimal place
EBRT	External beam radiation therapy

Abbreviation or special term	**Explanation**
ECOG	Eastern Cooperative Oncology Group
ECG	Electrocardiogram
eCRF	Electronic case report form
EORTC	European Organization for Research and Treatment of Cancer
EQ-5D-5L	EuroQol five-dimensional five-level questionnaire
FAS	Full analysis set
FIGO	International Federation of Gynecology and Obstetrics
HR	Hazard Ratio
HRQol	Health-related quality of life
IA	Interim analysis
ICU	Intensive care unit
IDMC	Independent data monitoring committee
IHC	Immunohistochemistry
imAE	Immune-mediated adverse event
IP	Investigational product
IPD	Important protocol deviation
IRMC	Independent review master charter
ITT	Intention to treat
IUO	Investigational use only
KM	Kaplan-Meier
LD	Longest diameter
MedDRA	Medical dictionary for regulatory activities
MMRM	Mixed model repeated measures
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MTp	Multiple testing procedure
nAb	Neutralizing antibody
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NTL	Non-target lesion
OAE	Other significant adverse event
OS	Overall survival
ORR	Objective response rate

Abbreviation or special term	**Explanation**
PD	Progressive disease
PD-L1	Programmed cell death protein 1
PFS	Progression-free survival
PGIC	Patient global impression of change
PGIS	Patient global impression of severity
PK	Pharmacokinetics
PR	Partial response
PRO	Patient reported outcome
PRO-CTCAE	Patient reported outcomes version of the common terminology criteria for adverse events
PS	Performance status
q1w	Every 1 week
q4w	Every 4 weeks
q12w	Every 12 weeks
q24w	Every 24 weeks
QLQ-C30	30 item core quality of life questionnaire
QLQ-CX24	Cervical cancer module
QTcF	Friderica's correction
RDI	Relative dose intensity
RECIST	Response evaluation criteria in solid tumors
REML	Restricted maximum likelihood
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SD	Stable disease
SoC	Standard of care
TEAE	Treatment emergent adverse event
TFST	Time to first subsequent therapy
TL	Target lesion
TMB	Tumor mutation burden
TSH	Thyroid stimulating hormone
TSST	Time to second subsequent anti-cancer therapy
ULN	Upper limit of normal
WHO	World Health Organisation

Abbreviation or special term	**Explanation**
WHO-DD	World Health Organisation drug dictionary

AMENDMENT HISTORY

Category: Change refers to	Date	Description of change	In line with CSP? Y (version) / N / NA	Rationale
The primary or secondary endpoints	09 Jan 2020	Updated the list of secondary and exploratory objectives. Applies to Tables 2 and 3, Sections 4,3,4 and 4.3.4	Y (v2.0)	To align with CSP
	09 Jan 2020	Added text to the primary objective statement. Applies to Table 1	Y (v2.0)	To align with CSP
	04 Aug 2021	Added text for DoR to the SAP and to the list of secondary objectives. Applies to Table 2 and new Sections 3.2.8 and 4.3.3.7		To align with CSP future edition?
	04 Aug 2021	Added OS in PD-L1 positive subjects as a secondary objective and added a new section for the analysis set. Applies to Sections 1.1.2, 2.1, 3.2.5, 4.3.3.4 and Tables 4 and 15	Y (v3.0)	To align with CSP
	04 Aug 2021	Added sections to describe PFS in PD-L1 positive subjects and PFS (3 Years). Applies to new Sections 3.2.2 and 3.2.3	Y (v3.0)	To align with CSP
	04 Aug 2021	Added text to explain what will be done for any OS analysis conducted prior to the final OS analysis where a survival call has not been made. Applies to Section 3.2.4	NA	To improve previous intent
	04 Aug 2021	Added summaries related to subjects who have completed treatment/did not receive treatment when discussion treatment status at progression as well as time from last tumor assessment to DCO in	NA	To improve previous intent

		censored subjects. Applies to Section 4.3.2		
	04 Aug 2021	Removed reference to 20-week assessment for CR rate. Applies to Tables 2 and 15 and Sections 3.2.7 and 4.3.3.6	Y(v4.0)	To align with CSP
	04 Aug 2021	Ensured objectives in Section 3 match the objectives in the CSP	Y(v4.0)	To align with CSP
Derivation of primary or secondary endpoints	09 Jan 2020	Removed reference to 30/90 days after last dose and added text to indicate data will be mapped according to visit windowing rules. Applies to Sections 4.3.5.6, 4.3.5.7, 4.3.5.8	NA	To improve previous intent
	09 Jan 2020	Added text to clarify that subjects will have restaging physical exams at each visit. Applies to Section 3.1	NA	To improve previous intent
	09 Jan 2020	Added text to indicate what happens if a subject with no measurable disease has been enrolled. Applies to Section 3.1.1, Table 8 and Table 9	NA	To improve previous intent
	09 Jan 2020	Added text to indicate the TL visit response if all or more than 1 measurements are missing and the rule when change in assessment method involved clinical examination. Applies to Section 3.1.1	NA	To improve previous intent
	09 Jan 2020	Updated the text on analysis visits and visit windows and definition of baseline updated. Applies to Section 3.5.10.2 and Table 14	NA	To improve previous intent
	09 Jan 2020	Modified text on rules for handling missing data. Applies to Section 3.5.1.3	NA	To align with PHUSE Guidance and improve

			previous intent
09 Jan 2020	Modified text on recording of new lesions and therefore also overall visit response. Applies to Section 3.1.2 and Table 8	NA	To improve previous intent
09 Jan 2020	Modified text on DoR for complete responders. Applies to Section 4.3.3.4	NA	To improve previous intent
09 Jan 2020	Modified the two-missed visits rule. Applies to Section 3.2.1	NA	To align with AZ standard oncology SAP guidance
09 Jan 2020	Updated text on local progression. Applies to Section 3.2.1	Y (v2.0)	To align with CSP
	Modified text to indicate which denominators are to be used for ORR. Applies to Section 3.2.3	NA	To improve previous intent
09 Jan 2020	Added text for the 4-week window that will be applied for the analysis of CR rate. Applies to Section 3.2.4	NA	To improve previous intent
09 Jan 2020	Modified text on exposure and dose interruptions. Applies to Section 3.5.3	NA	To improve previous intent and align with TFLs
04 Aug 2021	Modified text on the derivation of overall visit response and removed Table 9. All references to Table 9 updated and all subsequent table numbers and cross references updated. Applies to Section 3.2.1.1 and throughout	NA	To improve previous intent

	04 Aug 2021	Modified text on DoR in complete responders to remove reference to 20 weeks. Applies to Sections 3.2.9 and 4.3.3.8	NA	To improve previous intent
	04 Aug 2021	Clarified the definition of a non-evaluable assessment. Applies to section 3.2.1	NA	To improve previous intent
	04 Aug 2021	Time from randomization to the date of first documented local progression, distant disease progression/secondary malignancy or death	Y(v4.0)	To align with CSP
Sample size and multiple testing procedure	09 Jan 2020	Updated text on sample size, multiple testing procedure and interim analyses. Applies to Sections 1.3, 4.2 and 5.0	Y (v2.0)	To align with CSP and aims to strengthen the analysis plan for the OS endpoint
	04 Aug 2021	Added details of China cohort including number of subjects required and potential to recruit beyond global LSI. Applies to Section 1.2	Y (v3.0)	To align with CSP
	04 Aug 2021	Updated text on sample size, multiple testing and interim analyses to align with CSP. Applies to Sections 1.3, 4.0, 4.2 and 5.0	Y (v4.0)	To align with CSP
Statistical analyses	09 Jan 2020	Updated the list of statistical analyses and endpoints. Applies to Table 15	Y (v2.0)	To align with CSP
	09 Jan 2020	Added extra sensitivity analyses for (a) suspected progression on physical exam and no confirmatory biopsy and (b) Max-Combo test. Also removed reference to stratified log-	Y (v2.0)	To align with CSP

	rank test for subgroup analysis of PD-L1. Applies to Section 4.3.2		
09 Jan 2020	Modified text on competing risks methodology. Applies to Section 4.3.3.7	NA	To improve previous intent
	Modified text on PRO summaries and removed text on Bonferroni-Holm adjustment text. Applies to Section 4.3.3.9	NA	To improve previous intent
04 Aug 2021	Added text on new subgroup analyses for radiotherapy and lymph nodes. Modified the text to say eCRF or lab data will be used for all subgroup analyses. Applies to Section 4.3.2	NA	To improve previous intent
04 Aug 2021	Modified the statistical analyses for PFS at 3 years to indicate no formal comparisons will be done and only summary information will be provided. Applies to Section 4.3.3	NA	To improve previous intent
04 Aug 2021	Added new sensitivity analysis for PFS where stratification factors are based on eCRF data. Also modified text to indicate imputation for day when month and year are presented for date of birth. Applied so Section 4.3.2	NA	To improve previous intent
04 Aug 2021	Modified the analyses of OS to make it clear that the analyses would follow that used for the primary PFS analyses	NA	To improve previous intent
04 Aug 2021	Added clarification related to COVID-19 censoring. Applies to Section 4.4	NA	To align with COVID-19 related guidance

	04 Aug 2021	Added COVID-19 censoring analysis to Table 15	NA	To improve previous intent
Editorial	09 Jan 2020	Replaced all occurrences of "patient(s)" with "subject(s)". Applies throughout	NA	To align with ICH E6 guidance
	09 Jan 2020	Replaced all occurrences of "IVRS" with "IWRS". Applies throughout	NA	To improve previous intent
	09 Jan 2020	Updated CSP version number from 1.0 to 2.0. Applies to Section	Y (v2.0)	To align with CSP
	09 Jan 2020	Updated text to indicate FIGO staging 2009 will be used. Applies to Section 1.0, 1.1, 2.2, 4.3.2, 4.3.3.7 and Figure 1	Y (v2.0)	To improve previous intent
	09 Jan 2020	Replaced "maintenance therapy" with "adjuvant therapy". Applies to Section 1.2	Y (v2.0)	To align with CSP
	09 Jan 2020	Removed redundant text when discussing PK analysis set. Applies to Section 2.1.3	Y (v2.0)	To improve previous intent
	09 Jan 2020	Reference to Section 3.5.9 corrected to 3.5.10. Applies to Sections 3.5.6 and 3.5.7	NA	To improve previous intent
	09 Jan 2020	Fixed typos and corrected the reference to Table 10 to Table 14. Applies to Section 3.5.10.2	NA	To improve previous intent
	09 Jan 2020	Reference to section 3.5.9.3 corrected to Section 3.5.10.3. Applies to section 3.9.1	NA	To improve previous intent
	09 Jan 2020	Replaced "progression free" with "progression-free". Applies to Sections 1.2, 2.2, 4.3.2 and 4.3.3.5	NA	To improve previous intent

09 Jan 2020	Corrected typos of UK English and changed to US English. Applies to Section 3.1, 4.1, 4.3.2, 4.3.8	NA	To improve previous intent
09 Jan 2020	Added definition of ADA-evaluable set and modified text on ADA assessments. Applies to Section 2.1.2, and 3.7	NA	To improve previous intent
09 Jan 2020	Modified definition of the SAF to include all study treatments. Applies to Section 2.1.2	NA	To improve previous intent
09 Jan 2020	Modified text that indicated what happens if subjects have no evaluable visits or baseline RECIST/physical assessment. Applies to Section 3.2.1	NA	To improve previous intent
09 Jan 2020	Added text to clarify data used for BoR and corrected timeframe for subjects who die with no evaluable assessments. Applies to Section 3.2.7	NA	To improve previous intent
09 Jan 2020	Modified text on EORTC-QLQ-C30 to clarify rules on missing data and correct date of Fayers reference. Applies to Section 3.3.1	NA	To improve previous intent
09 Jan 2020	Modified text on treatment exposure for SoCs by removing repeated text. Applies to Section 3.5.2.2	NA	To improve previous intent
09 Jan 2020	Added text on missed or forgotten doses when dealing with exposure. Applies to Section 3.5.2.2	NA	Updated to align with AZ standard oncology SAP guidance
09 Jan 2020	Moved general considerations for safety variables to the start of safety variables section and re-numbered all subsequent sections/subsections and	NA	To improve previous intent

	cross references. Applies to Section 3.5		
	Modified text on definition of baseline to include dosing with any study treatment. Applies to Section 3.5.1	NA	To improve previous intent
09 Jan 2020	Moved the text on Safety Follow-up to a new section in general considerations. Applies to Section 3.5.1	NA	To improve previous intent
09 Jan 2020	Modified text on references ranges used for laboratory variables to indicate project ranges will be used throughout. Applies to Section 3.5.7	NA	To improve previous intent
09 Jan 2020	Added text to indicate the KM plot and landmark analyses will help understand treatment benefit. Applies to Section 4.3.2	NA	To improve previous intent
09 Jan 2020	Corrected SAP text on Deaths. Applies to Section 4.3.5.1	NA	To improve previous intent
09 Jan 2020	Corrected the definition of prior therapies. Applies to Section 3.9	NA	To improve previous intent
09 Jan 2020	Modified the text on inclusions/exclusions that count as violations and deviations. Applies to Section 2.2	Y (v2.0)	To align with CSP
09 Jan 2020	Added text to indicate summaries of ADA will be presented for each arm and modified the list of summaries. Applies to section 4.3.7	Y (v2.0)	To align with CSP and improve previous intent

	09 Jan 2020	Modified Deviation 4 to include physical exam and replace “assigned to treatment” with “randomized”. Applies to Section 2.2	NA	To improve previous intent
	09 Jan 2020	Replaced “distant disease recurrence” with “distant disease progression”. Applies to Sections 3.2.1, 3.2.6	NA	To improve previous intent
	09 Jan 2020	Changed the ordering of the PFS (3 year), PFS in PD-L1 subjects and OS. Applies to Section 4.3.3	NA	To align with CSR ordering
	04 Aug 2021	Modified the list of IPDs to incorporate two missing IPDs and remove reference to a PD. Applies to section 2.2	NA	To improve previous intent
	04 Aug 2021	Updated the version of the CSP from 2.0 to 3.0. Applies to Section 1	Y (v3.0)	To align with CSP
	04 Aug 2021	Modified text on AESIs to include AESI and AEPI as per AZ guidance. Applies to Section 3,5,6 and 4.3.5.2	NA	To improve previous intent and align with new guidance
	04 Aug 2021	Changed type of chemotherapy to first dose of chemotherapy and appropriate to inappropriate. Applies to Section 4.3.2	NA	To improve previous intent
	04 Aug 2021	Modified definition of the safety analysis set to remove reference to chemotherapy and radiotherapy exposure. Applies to Section 2.1.2	Y (v3.0)	To align with CSP
	04 Aug 2021	Aligned endpoints to match the CSP	Y (v4.0)	To align with CSP
Data presentation	09 Jan 2020	Added text to indicate errors in stratifications will be summarized. Applies to Section 2.2	NA	To improve previous intent

09 Jan 2020	Modified the list of safety data to be summarized by the SAF and ADA-evaluable set. Applies to Table 4	NA	To improve previous intent
09 Jan 2020	Added text on deriving change from baseline for vital signs. Applies to Section 3.5.7	NA	To improve previous intent
09 Jan 2020	Language used for QTcF calculation modified and derivation of RR added, Applies to Section 3.5.10	NA	To improve previous intent
09 Jan 2020	Modified list of summaries for ADA analyses. Applies to Section 3.7	NA	To improve previous intent
09 Jan 2020	Added text to confirm that strata will be checked and condensed prior to unblinding. Applies to Section 4.3.1	NA	To improve previous intent
09 Jan 2020	Added text to indicate the primary source of strata and its comparison with source data. Applies to Section 4.1	NA	To improve previous intent
09 Jan 2020	Removed text stating the KM plots would be produced. Applies to Section 4.3.3.7	NA	To improve previous intent
09 Jan 2020	Modified text to clarify ECG summaries. Applies to Section 4.3.5.6	NA	To improve previous intent
09 Jan 2020	Updated the summaries of demographic, initial diagnosis and baseline characteristics. Applies to Section 4.3.8	NA	To improve previous intent
09 Jan 2020	Updated list of AE summaries. Applies to Section 4.3.5.1, 4.3.5.2	NA	To improve previous intent
09 Jan 2020	Updated list of exposure summaries. Applies to Section 4.3.5.4	NA	To improve previous intent

09 Jan 2020	Updated the ECG summaries. Applies to Section 4.3.5.6	NA	To improve previous intent
09 Jan 2020	Updated the vital signs summaries. Applies to Section 4.3.5.7	NA	To improve previous intent
09 Jan 2020	Added text of heath care resource use. Applies to Section 4.3.4	NA	To improve previous intent
09 Jan 2020	Modified text on concomitant medication and other treatment summaries. Applies to Section 4.3.9	NA	To improve previous intent
09 Jan 2020	Modified text on PFS summaries and subgroup analyses of PFS in PD-L1 subjects and Japan versus rest of the world. Applies to Section 4.3.2	NA	To improve previous intent
09 Jan 2020	Added text on PFS in PD-L1 positive subjects. Applies to Section 4.3.3.4	Y (v2.0)	To improve previous intent
04 Aug 2021	Added text for ECGs to indicate what would be presented when triplicate results are reported for a visit. Applies to Section 4.3.5.6	NA	To improve previous intent
04 Aug 2021	Updated text on AEs and added extra summaries for AEs related to study drug by maximum CTCAE grade. Applies to Section 4.3.5.1	NA	To improve previous intent
04 Aug 2021	Replaced ADA positive with ADA-positive for consistency with other uses and TFLs. Applies throughout	NA	To improve previous intent
04 Aug 2021	Modified the list of summaries for demographics and baseline characteristics to add summaries of agreement between IWRS and eCRF and time to start of subsequent	NA	To improve previous intent

	therapy from treatment discontinuation. Removed previous disease related treatment modalities as this data is not collected. Applies to Section 4.3.7		
04 Aug 2021	Removed references to R from discussions of local incidence and distant disease progression. Applies to Section 4.3.3.9	NA	To improve previous intent
04 Aug 2021	Removed “or death from any cause” from the list of time from randomization summaries that will be produced when looking at incidence of local/distant disease progression. Applies to Section 3.2.10	NA	To improve previous intent
04 Aug 2021	Added text to clarify the time frame of reporting ECGs and Vital signs. Applies to Sections 4.3.5.6 and 4.3.5.7	NA	To improve previous intent
04 Aug 2021	Added text to indicate what summaries may be produced to show the impact of the COVID-19 pandemic. Applies to Section 4.4	NA	To align with CSP
04 Aug 2021	Removed reference to hyper for Glucose summaries as this is not presented. Applies to Sections 3.5.7 and 4.3.5.5	NA	To improve previous intent
04 Aug 2021	Added text on time to second subsequent therapy or death as a supportive analysis of PFS. Applies to Sections 3.2.1.2 and 4.3.2	NA	To improve previous intent
04 Aug 2021	Infection AEs removed from AESI/AEPI section	NA	To improve previous intent

1 STUDY DETAILS

This statistical analysis plan (SAP) contains a more detailed description of the analyses in the clinical study protocol (CSP) and is based on version 4.0 of the CSP.

This SAP will apply to the phase III study to determine the efficacy and safety of durvalumab in combination with and following chemoradiotherapy compared to chemoradiotherapy alone for locally advanced cervical cancer.

The target population for this study is adult female subjects (age ≥18 years) with histologically confirmed cervical adenocarcinoma, cervical squamous carcinoma, or cervical adenosquamous carcinoma. Subjects must have International Federation of Gynaecology and Obstetrics (FIGO (2009)) Stages IB2 to IIB, node positive (N≥1) or FIGO (2009) Stages IIIA to IVA with any N state (N≥0). Nodal staging may be either surgical or by imaging (Response evaluation criteria in solid tumors [RECIST] v.1.1). Subjects should have no evidence of metastatic disease and no prior exposure to surgical, radiation, or systemic therapy for cervical cancer or immunotherapy. Subjects must also have a World Health Organization (WHO)/Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0 or 1 at enrollment and randomization.

1.1 Study objectives

1.1.1 Primary objectives

The primary objectives for this study and the corresponding endpoints/variables are shown in Table 1.

Table 1: Primary study objectives and corresponding endpoints/variables

Objective	Endpoints/variables
To assess the efficacy of durvalumab + SoC CCRT compared with placebo + SoC CCRT in terms of PFS	Endpoints based on investigator assessment according to RECIST 1.1 or histopathologic confirmation of local tumor progression: • PFS: Time from date of randomization until tumor progression or death due to any cause

CCRT Concurrent chemoradiation therapy; PFS Progression-free survival; RECIST Response evaluation criteria in solid tumors; SoC Standard of care

1.1.2 Secondary objectives

The secondary objectives for this study and the corresponding endpoints/variables are shown in Table 2.

Table 2: Secondary study objectives and corresponding endpoints/variables

Objective	Endpoints/variables
To further assess the efficacy of durvalumab + SoC CCRT compared with placebo + SoC CCRT in terms of PFS (3 year), PFS in PD-L1 positive patients, overall survival (OS) (key), OS in PD-L1 positive patients, ORR CR rate, and DoR in patients with a CR	• OS: Time from date of randomization until the date of death by any cause • OS in PD-L1 positive subjects: Time from date of randomization until the date of death by any cause in subjects who are PD-L1 positive (>=1%) based on either tumor or immune cell staining Endpoints based on investigator assessment according to RECIST 1.1 or pathologic assessment of local disease progression: • PFS (3 years): Proportion of patients alive and progression-free at 3 years • PFS in PD-L1 positive subjects: Time from date of randomization until tumor progression or death due to any cause in patients who are PD-L1 positive (≥1%) based on either tumor or immune cell staining • ORR: The percentage of evaluable subjects with investigator-assessed visit response of CR or PR • CR rate: Disappearance of all target and non-target lesions • DoR in subjects with CR: Time from date of first detection of response of CR until the date of disease progression

To assess the effect of durvalumab + SoC CCRT compared with placebo + SoC CCRT on the incidence of local progression, distant disease progression, and secondary malignancy as the first documented progression event	Incidence of local progression, distant disease progression, and secondary malignancy: Number and percentage of subjects who develop local progression, distant disease progression, or secondary malignancy
To assess disease-related symptoms and health-related quality of life (HRQoL) in subjects with cervical cancer treated with durvalumab + SoC CCRT compared with placebo + SoC CCRT using the core quality of life questionnaire (EORTC QLQ-C30) and core quality of life questionnaire cervical cancer module (CX24)	Change from baseline in EORTC QLQ-C30 and EORTC-CX24
To assess the PK of durvalumab when in combination with SoC CCRT	Blood concentration of durvalumab
To investigate the immunogenicity of durvalumab in combination with SoC CCRT	Presence of ADAs for durvalumab
To assess the safety and tolerability profile of durvalumab + SoC CCRT compared to placebo + SoC CCRT	AEs, laboratory findings, vital signs, physical examinations

ADA Antidrug antibody; AE Adverse event; CCRT Concurrent chemoradiation therapy; CR Complete response; DoR Duration of response; EORTC European Organization for Research and Treatment of Cancer; HRQoL Health-related quality of life; ORR Objective response rate; OS Overall survival; PD-L1 Programmed cell death protein 1; PFS Progression-free survival; PK Pharmacokinetics; QLQ-C30 30 item core quality of life questionnaire; QLQ-CX24 Cervical cancer module; RECIST Response evaluation criteria in solid tumors; SoC Standard of care

1.1.3 Exploratory objectives

Table 3: Exploratory study objectives and corresponding endpoints/variables

Exploratory objective:	Endpoint/Variable:
To collect blood and tissue samples for defining biological response to durvalumab +SoC CCRT compared to placebo +SoC CCRT for candidate markers that may correlate with likelihood of clinical benefit **(All countries except China)**	Analysis of blood/tissue samples to assess exploratory biomarkers, which may include, but is not limited to ctDNA, mRNA signatures, CD8 by IHC, and tumor mutational burden (TMB)
To assess treatment-related symptoms using the patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE)	Specific treatment related PRO-CTCAE symptoms
To assess how a subject perceives her overall change in cancer symptoms since the start of study treatment using the PGIC and overall severity of cancer symptoms using the PGIS.	PGIC and PGIS
To explore the impact of treatment and disease state on health status assessed by EuroQol five-dimensional five-level questionnaire (EQ-5D-5L) to support health economic analysis and health technology assessment	EQ-5D-5L
To describe and evaluate resource use associated with durvalumab treatment and underlying disease	Health resource utilization measures including hospitalization, outpatient visits, or emergency department visits

CCRT Concurrent chemoradiation therapy; CD Cluster of differentiation; ctDNA Circulating tumor deoxyribonucleic acid; EQ-5D-5L EuroQol five-dimensional five-level questionnaire; IHC Immunohistochemistry; mRNA Messenger ribonucleic acid; PFS Progression-free survival; PGIC Patient global impression of change; PGIS Patient global impression of severity; PRO-CTCAE Patient-reported outcomes version of the Common Terminology Criteria for Adverse Events; RECIST Response Evaluation Criteria in Solid tumors; SoC Standard of care; TMB tumor mutation burden

1.2 Study design

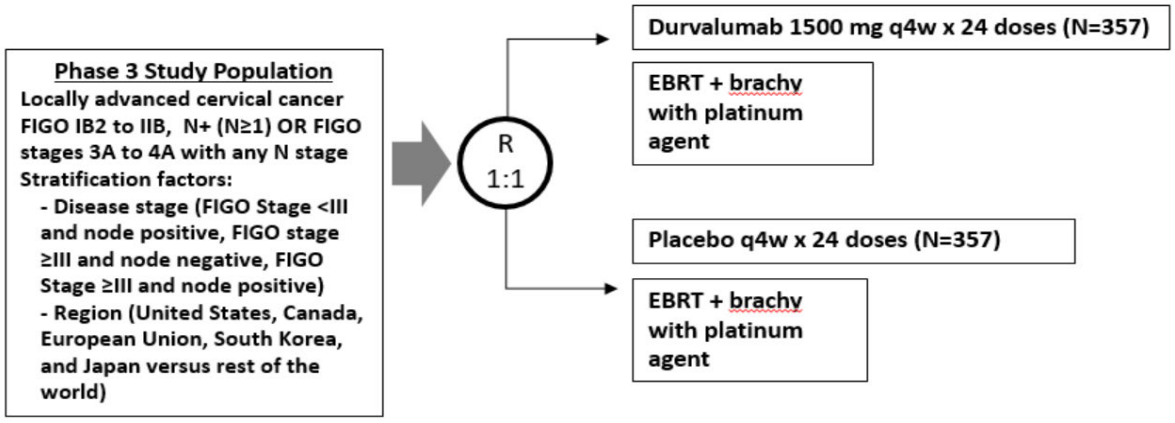
This is a randomized, double-blind, placebo controlled, multi-center, global, Phase III study to determine the efficacy and safety of durvalumab when combined with standard of care (SoC) concurrent chemoradiation therapy (CCRT) and administered as adjuvant therapy following SoC CCRT compared to placebo with SoC CCRT as systemic treatment in subjects with FIGO (2009) Stages IB2 to IVA cervical cancer.

Subjects will be randomized 1:1 to receive either durvalumab 1500 mg or placebo every 4 weeks (q4w) for 24 doses. Subjects in both arms will receive SoC CCRT consisting of external beam radiation therapy (EBRT) plus brachytherapy, and concurrent cisplatin 40 mg/m^2 every 1 week (q1w) for 5 weeks or carboplatin area under the serum drug concentration curve 2 (AUC 2) q1w for 5 weeks (carboplatin only applies for subjects randomized prior to CSP version 3). A 6th week of platinum agent can be given per investigator discretion (CSP Section 6.1.2 for descriptions of the dosing regimens). Randomization will be stratified by disease stage status (FIGO (2009) Stage <III and node positive, FIGO (2009) Stage ≥III and node negative, or FIGO (2009) Stage ≥III and node positive), and region (United States, Canada, European Union, South Korea, and Japan) versus rest of world.

Subjects will receive their assigned treatment until completion of planned therapy, clinical progression or RECIST 1.1-defined radiological progression or histopathologic progression on biopsy unless there is unacceptable toxicity, withdrawal of consent, or another discontinuation criterion is met (CSP Section 7.1 for additional details on discontinuation of study treatment). Following completion of SoC CCRT, subjects will be evaluated by clinical and radiological assessment. During the first 164 weeks after randomization, this assessment will include an interval history for new signs or symptoms, targeted physical examination to evaluate for local progression, and imaging (computed tomography [CT] or magnetic resonance imaging [MRI]) per RECIST 1.1 every 12 weeks (q12w). During years 4, and 5, imaging/histopathologic assessments will be performed every 24 weeks (q24w) and then annually until a progression-free survival (PFS) endpoint has been met or closure of the study (CSP Section 8.1).

An overview of the study design is shown in Figure 1.

Figure 1: Overview of study design



Brachy brachytherapy; EBRT External beam radiation therapy; FIGO (2009) International Federation of Gynecology and Obstetrics; q4w Every 4 weeks; R Randomized

1.3 Number of subjects

Approximately 714 subjects (357 per treatment arm) will be randomized 1:1 to durvalumab plus SoC CCRT or placebo plus SoC CCRT to obtain approximately 227 PFS events in the intention to treat (ITT) population. The randomization will be stratified according to disease stage status (FIGO (2009) Stage <III and node positive, FIGO (2009) Stage ≥III and node negative, or FIGO (2009) Stage ≥III and node positive), and region (United States, Canada, European Union, South Korea, and Japan) versus rest of the world.

In China, recruitment will continue until approximately 105 Chinese subjects have been randomized, irrespective of whether or not the overall study enrollment has been reached. It is anticipated that this target may not be met before the global recruitment of approximately 714 subjects is achieved. Further details on the China cohort will be detailed in a China supplementary SAP

The PFS analysis will occur when approximately 227 PFS events (approximately 32% maturity) have occurred across the durvalumab plus SoC CCRT and placebo plus SoC CCRT treatment arms. If the true hazard ratio (HR) is 0.65 [likely to correspond to a 11% increase in the proportion of subjects progression-free at 3 years from 65% to 76%] (Rose *et al.* 1999), this analysis will have 90% power to demonstrate a statistically significant difference for PFS, assuming a 2-sided 5% significance level. Assuming non-uniform recruitment, a 3-year PFS rate of 65% for placebo + SoC CCRT and an exponential distribution, recruitment of 714 subjects in 18 months would be expected to yield 227 PFS events approximately 53 months following recruitment of the first subject. In addition, the sample size has been derived on the assumption that 5% of subjects will drop out at an exponential rate within 36 months following the randomization of the first subject. If the true OS HR is 0.65, the study is also

sized to provide 89% power for the OS endpoint in the comparison of durvalumab + SoC CCRT versus placebo + SoC CCRT, assuming a 2-sided 5% significance level allowing for 1 interim analysis conducted at the same time as the PFS analysis, with an expected 86% of target OS events. The analysis of OS will only be undertaken if the PFS is statistically significant. The final OS analysis will occur at the earlier of 227 death events or 71 months following recruitment of the first subject. Assuming the same recruitment assumptions for PFS and a 3-year OS rate of 70% for placebo + SoC CCRT (Rose *et al.* 1999), 227 deaths would be expected to occur 61 months following the recruitment of the first subject. The boundaries of the treatment comparison will be derived based upon the actual number of OS events observed at that time.

2 ANALYSIS SETS

2.1 Definition of analysis sets

There are four analysis sets defined for this study. A summary of the analysis sets used for each outcome variable is provided in Table 4.

Table 4: Summary of outcome variables and analysis populations

Outcome variable	Populations
Demographics, baseline characteristics, medical/surgical history, concomitant medications/procedures	Full analysis set
Efficacy data: PFS, ORR, CR rate, DoR in subjects with CR, Incidence of local progression, distant progression and secondary malignancy, PFS (3 year), OS	Full analysis set
PROs	Full analysis set
PFS, OS	PD-L1 analysis set
Safety data: AEs (including SAEs), exposure, laboratory measurements, physical examinations, vital signs, WHO/ECOG PS	Safety analysis set
ADA data	Safety analysis set
PK data	PK analysis set

Outcome variable	Populations

ADA Antidrug antibody; AE Adverse event; CR Complete response; DoR Duration of response; ECOG Eastern Cooperative Oncology Group; ORR Objective response rate; OS Overall survival; PFS Progression-free survival; PRO Patient reported outcomes; PS performance status; WHO World health organization

2.1.1 Full analysis set (Intention to treat [ITT])

The full analysis set (FAS) will include all randomized subjects. The FAS will be used for all efficacy analyses (including patient reported outcomes [PROs]). Treatment arms will be compared on the basis of randomized study treatment, regardless of the treatment actually received. Subjects 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 Safety analysis set

The safety analysis set (SAF) includes all subjects who received at least 1 dose of study treatment (durvalumab or placebo). Safety data will not be formally analyzed but summarized using the SAF, according to the treatment received. Erroneously treated subjects (e.g., those randomized to durvalumab but actually given placebo) will be summarized according to the treatment they actually received. If a subject only receives therapy from the placebo arm, they will be summarized in the placebo treatment group. If a subject receives any amount of durvalumab, they will be summarized in the durvalumab treatment group.

2.1.3 Pharmacokinetics analysis set

The pharmacokinetics (PK) analysis set includes all subjects who receive at least 1 dose of durvalumab per the protocol for whom any post-dose PK concentration data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analyses. The population will be defined by the Study Physician, Pharmacokineticist, and Statistician prior to any analyses being performed. The PK analysis set will be summarized according to the treatment actually received.

2.1.4 PD-L1 analysis set

The programmed cell death protein 1 (PD-L1) analysis set will include a subset of subjects in the FAS who have a baseline PD-L1 expression result as described in Section 3.2.2.

The CSP uses the terms PD-L1 positive and PD-L1 negative. PD-L1 positive has the same meaning as PD-L1 high and PD-L1 negative has the same meaning as PD-L1 low. The SAP and CSR outputs will use PD-L1 high/low.

2.2 Violations and deviations

For this study, the following general categories will be considered important protocol deviations (IPDs) and will be programmatically derived from the electronic case report form (eCRF) data where possible to do so. These will be listed and summarized by randomized treatment group and discussed in the clinical study report (CSR) as appropriate:

- Subjects who deviate from key inclusion criteria per the CSP:
  - Inclusion 5 (IPD 1.1): Histologically confirmed cervical adenocarcinoma, cervical squamous carcinoma, or cervical adenosquamous carcinoma and the following requirements:
    - FIGO (2009) Stages IB2 to IIB, node positive (N≥1) -OR- FIGO (2009) Stages IIIA to IVA with any N stage (N≥0)
    - Nodal staging may be either surgical or by imaging (CT or MRI) with pathological lymph node size defined by a short-axis diameter of ≥10mm (axial plane)
    - No evidence of metastatic disease
  - Inclusion 9 (IPD 1.2): Adequate organ and marrow function as defined below:
    - Hemoglobin ≥9.0 g/dL
    - Absolute neutrophil count $\geq 1.5 \times 10^9$ /L
    - Platelet count $\geq 75 \times 10^9$/L
    - Serum bilirubin ≤1.5 × the upper limit of normal (ULN). This will not apply to subjects with confirmed Gilbert's syndrome, who will be allowed in consultation with their physician and AstraZeneca
    - Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) ≤2.5 × ULN
    - Creatine clearance (CrCl) ≥60 mL/min calculated by Cockcroft-Gault equation (using actual body weight) or by measured 24-hour urine collection

      Females:
      $$\text{Creatinine CL (mL/min)} = \frac{\text{Weight (kg)} \times (140 - \text{Age})}{72 \times \text{serum creatinine (mg/dL)}} \times 0.85$$

  - Subjects who have not signed the informed consent form (IPD 1.3)

- Subjects who deviate from key exclusion criteria per the CSP:

  - Exclusion 3 (IPD 2.1): Evidence of metastatic disease per RECIST 1.1 including lymph nodes ≥15 mm (short axis) above the L1 cephalad body or outside the planned radiation field

  - Exclusion 4 (IPD 2.2): Subjects who have undergone a previous hysterectomy, including a supracervical hysterectomy, or will have a hysterectomy as part of their initial cervical cancer therapy

  - Exclusion 6 (IPD 2.3): Active or prior documented autoimmune or inflammatory disorders (including inflammatory bowel disease [e.g., colitis or Crohn's disease], diverticulitis [with the exception of diverticulosis], systemic lupus erythematosus, Sarcoidosis syndrome, or Wegener syndrome [granulomatosis with polyangiitis, Graves' disease, rheumatoid arthritis, hypophysitis, uveitis, etc]). The following are exceptions to this criterion:

    - Subjects with vitiligo or alopecia

    - Subjects with hypothyroidism (e.g., following Hashimoto syndrome) stable on hormone replacement

    - Subjects with any chronic skin condition that does not require systemic therapy

    - Subjects without active disease in the last 5 years may be included but only after consultation with the Study Physician

    - Subjects with celiac disease controlled by diet alone

  - Exclusion 14 (IPD 2.4): Prior history of vesicovaginal, colovaginal, or rectovaginal fistula

  - Exclusion 17 (IPD 2.5): Any concurrent chemotherapy, investigational product (IP), biologic, or hormonal therapy for cancer treatment. Concurrent use of hormonal therapy for non-cancer-related conditions (e.g., hormone replacement therapy) is acceptable

- Non-compliance with the study protocol that warrants withdrawal from treatment with IP (IPD 3.4)

- Subjects who were randomized who received treatment other than that to which they were randomized (IPD 5.1)

- Subjects who were randomized but who did not receive at least one dose of study treatments

  – Subjects randomized but who did not receive durvalumab or matching placebo (IPD 5.2/5.3)

  – Subjects randomized but who did not receive protocol defined CCRT (IPD 5.4)

- Subjects who have known hypersensitivity to durvalumab or any of the excipients of the product prior to enrollment (IPD 5.6)

- The subject received prohibited concomitant medications (including other anti-cancer agents). Please refer to the CSP Section 6.4 for those medications that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to the database lock (IPD 6.1)

- Baseline assessments incorrectly completed/missed as defined below .

  – Baseline RECIST scan or physical exam > 42 days before date subject was randomized (based upon a 28-day screening period plus 2 weeks allowance, so that only serious violators are identified) (IPD 7.1)

  – No baseline RECIST 1.1 or physical exam assessment on or before date subject was assigned to treatment (IPD 7.2)

Subjects who receive the wrong treatment at any time will be included in the SAF 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 summarized by randomized treatment group. The summary will include a split of important protocol deviations that are and are not COVID-19 related. Deviation 1.1 (subjects randomized but who did not receive durvalumab or matching placebo) will lead to exclusion from the safety analysis set. None of the other deviations will lead to subjects 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 subjects with specific important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis may be performed on the progression-free survival endpoint excluding subjects with deviations that may affect the efficacy of the trial therapy if >10% of subjects in either treatment group:

- Did not have the intended disease or indication or
- Did not receive any randomized therapy

The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of database lock (DBL) 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 case report form (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 the IxRS and eCRF source data) will also be summarized separately to the important deviations.

3 PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of visit responses

For all subjects, the RECIST version 1.1 (Appendix F of the CSP) tumor response data will be combined with physical examination and biopsy data to determine each subject's visit response. It will also be used to determine if and when a subject has progressed and also their best objective response (BoR).

For all subjects, a physical examination including a pelvic examination and (if indicated) biopsy for histopathological determination of tumor progression will be assessed on the same schedule as the tumor assessments. If progression is suspected by the investigator during the pelvic exam, a biopsy must be performed for the purpose of confirming histopathological tumor progression. Some subjects may have a pap smear as part of the physical exam. If progression is suspected from the results of the pap smear, a biopsy must be performed for the purpose of confirming histopathologic tumor progression. The physical examination, pap smear, and biopsy data will be used to determine if and when a subject has histopathologic tumor progression. Throughout this document, the term "physical exam data" includes results from a pap smear where those results are available.

Baseline radiological tumor assessments and the baseline physical exam are to be performed no more than 28 days before the start of randomized treatment and ideally as close as possible to the start of study treatment. Post-baseline tumor assessments and physical exams by the investigator start 20 weeks (± 1 week) after randomization and continue every 12 weeks (±1 week) through 164 weeks (relative to date of randomization). Post 164 weeks, these occur every 24 weeks (±2 weeks) thereafter (relative to date of randomization) until RECIST 1.1 defined- disease progression or histopathologic confirmation of local disease progression plus an additional follow-up assessment consisting only of a RECIST 1.1 tumor assessment that is performed at the next (and no later than the next) scheduled imaging visit following either RECIST 1.1 defined progression or histopathologic confirmation of local disease progression and performed no less than 4 weeks after the prior assessment of PD.

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 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.

From the investigator's review of the imaging scans, the RECIST tumor response data will be used to determine each subject's visit response according to RECIST version 1.1. At each visit, subjects will be programmatically assigned a RECIST 1.1 visit response of complete response (CR), partial response (PR), stable disease (SD) or progressive disease (PD), using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions and depending on the status of their disease compared with baseline and previous assessments. If a subject has had a tumor assessment that cannot be evaluated then the subject will be assigned a RECIST 1.1 visit response of not evaluable (NE), (unless there is evidence of progression in which case the response will be assigned as PD). At each visit, subjects will also have a restaging physical examination including a pelvic examination and (if indicated) biopsy for histopathological determination of tumor progression.

Please refer to Section 3.1.1 for the definitions of CR, PR, SD and PD. The determination of overall visit response using the RECIST 1.1 tumor data, physical exam data, and biopsy data is discussed in Section 3.2.16.

3.1.1 RECIST 1.1 target lesions (TLs)

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥10mm in the longest diameter (LD), (except lymph nodes that must have a short axis ≥15mm) with CT or 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 target lesions (TLs). 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 TL should be identified as non-target lesions (NTLs) at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

Measurable disease (i.e., at least one TL) is one of the entry criteria for the study. However, if a subject who does not have measurable disease is enrolled into the study (i.e., no TLs), evaluation of the 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).

Table 5: 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 in certain situations (i.e. if any of the TLs were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.
Not applicable (NA)	No TLs are recorded at baseline.

Rounding of TL data
For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to one decimal place (d.p.) before assigning a TL response. For example, 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%.

Missing TL data
For a visit to be evaluable then 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 LD recorded. If all TL measurements are missing, then the TL visit response is NE. If there is at least one TL measurement missing and a TL visit response of PD cannot be assigned, the TL visit response is NE.

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.

TL visit responses subsequent to CR
Only CR, PD or NE can follow a CR. 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 (i.e. a pathological lymph node selected as TL has short axis >10mm or the reappearance of previously disappeared lesion) or a new lesion appears, then response will be set to PD

Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR

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

TL too small to measure
If a TL becomes too small to measure, then this will be indicated as such on the case report form 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 (at a subsequent visit) then this will be reviewed by the study team blinded to treatment assignment.

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

Any TL (including lymph nodes), which has had intervention during the study (surgery / embolization), should be handled in the following way. The current study includes radiotherapy as study treatment, and this will not be considered as a TL intervention in the eCRF. 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).

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.

Table 6: Example of scaling

Lesion	Longest diameter (mm) at nadir visit	Longest diameter (mm) at follow-up visit
1	16	18
2	14	16
3	14	16
4	18	18
5	12	Intervention

Sum	**74**	**68**

Lesion 5 has had an intervention at the follow-up visit. It had a baseline measure of 74mm. The sum of lesions 1 – 4 at the follow-up is 68mm. The sum of the corresponding lesions at nadir visit is 62mm. Scale up as follows to give an estimated follow-up visit TL sum of 81mm:

$$\frac{68}{62} \times 74 = 81mm$$

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

Lesions that split in two or more parts
If a TL splits in two or more parts, then the longest diameters of the split lesions should be summed and reported as the longest diameter for the lesion that split.

Lesions that merge
If two or more TLs merge, then the longest diameter of the merged lesion should be recorded for one of the TL sizes and the other TL sizes should be recorded as 0mm.

Change in method of assessment of TLs
CT and MRI are the only methods of assessment that can be used within a trial to assess target lesions. A clinical exam to assess for cervical cancer should be done following the tumor schedule however it cannot be used to assess target lesions. 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 RECIST 1.1 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 7: Non-target lesion 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

NTL Non-target lesion; TL Target lesion

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.

New lesions will be identified via a Yes/No tick box. The 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.

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.

3.1.3 Overall RECIST 1.1 visit response

Table 8 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 8: Overall RECIST 1.1 visit responses

Target	Non-target	New lesions	Overall visit response
CR	CR or NA	Not recorded	CR
CR	Non-CR/Non-PD or NE	Not recorded	PR
PR	Non-PD or NE or NA	Not recorded	PR
SD	Non-PD or NE or NA	Not recorded	SD
PD	Any	Any (i.e., Yes/Not recorded)	PD
Any	PD	Any (i.e., Yes/Not recorded)	PD

Any	Any	Yes	PD
NE	Non-PD or NE or NA	Not recorded	NE
NA	CR	Not recorded	CR
NA	Non-CR/Non-PD	Not recorded	SD
NA	NE	Not recorded	NE

CR Complete response; NA Not Applicable; NE Not evaluable; PD Progressive disease; PR Partial response; SD Stable disease

3.1.4 Independent review

A planned blinded independent central review (BICR) of all radiological imaging data will be carried out using RECIST version 1.1. All radiological scans for all subjects (including those at unscheduled visits, or outside visit windows) will be collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for central analysis. The imaging scans will be reviewed by two independent radiologists using RECIST 1.1 and will be adjudicated, if required (i.e., two reviewers review the scans, and adjudication is performed by a separate reviewer in case of a disagreement). For each subject, the BICR will define the overall visit response (i.e., the response obtained overall at each visit by assessing TLs, NTLs and new lesions) data and no programmatic derivation of visit response is necessary. (For subjects with TLs at baseline: CR, PR, SD, PD, NE; for subjects with NTLs only: CR, SD, PD, NE). If a subject has had a tumor assessment that cannot be evaluated, then the subject will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). Tumor assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision will be used). In the absence of adjudication, the records for all visits for a single reviewer will be used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to report all BICR RECIST information including dates of progression, visit response, censoring and changes in target lesion dimensions. Endpoints (of objective response rate (ORR), PFS and duration of response [DoR]) will be derived programmatically from this information.

Results of this independent review will not be communicated to investigators and the management of subjects will be based solely upon the results of the RECIST 1.1 assessment and physical examination conducted by the investigator.

A BICR of all subjects will be performed for the database lock for PFS, which will cover all scans up to the data cut-off (DCO).

Further details of the BICR will be documented in the independent review master charter (IRMC) Charter.

3.1.5 Histopathologic confirmation of local progression

At response assessment physical exams, subjects will be evaluated for local progression. A physical exam can include a pelvic examination with or without speculum, bimanual examination, rectovaginal examination, abdominal examination and optional pap smear. Physical examinations should be performed following the same schedule as tumor assessments by imaging. Any lesion detected on physical exam that is not evaluable through RECIST 1.1 and is within the radiation field and suspected to be progressive disease will be biopsied. Lesions identified as progressive disease by RECIST 1.1 on CT or MRI do not require a biopsy. At the discretion of the investigator, pap smears are an optional component of the response assessment physical exam. If a subject has a pap smear result indicating potential progression, a biopsy should be performed. Lesions with histopathologic confirmation of disease will be reported as a response of PD.

Note that pap smear is a component of the restaging physical exam. For the restaging physical exam, no suspected progression on the restaging physical exam indicates that there were no findings on the physical exam that were concerning for progressive disease. Suspected progression on the restaging physical exam indicates that either there were findings on the physical exam, the pap smear was performed and the results were indicative of progression, or both. Any positive biopsy result, without a linked pap smear, RECIST or CERCAPE result will be queried.

Radiological assessments should be obtained on all subjects within 28 days of progression if confirmed by histopathology. If the physical examination and biopsy are performed as unscheduled assessments and there is no evidence of active cervical cancer on histopathology, the subject will not be reported as PD, and will continue with all study procedures according to the Schedule of Assessments in the protocol.

3.1.6 Overall visit response

The overall visit responses will be determined using RECIST 1.1 criteria and the findings of response assessment physical exams and required confirmatory biopsies following suspected progression at a physical exam.

A subject's overall visit response will be classified as follows:

- If evidence of progression = yes from the CERCAPE eCRF and confirmation of previous equivocal progression from scan or PE = yes from the PATHROM eCRF then the overall visit response is PD regardless of the other categories. This applies to physical examination and biopsy assessments that are performed as scheduled and unscheduled assessments.

- If the above criteria are not met, then the overall visit response will be determined using RECIST 1.1 as shown in Table 8.

Table 9: Overall visit responses

Overall RECIST response	Physical exam	Biopsy result	Overall
CR	No suspected progression	NA	CR
PR	No suspected progression	NA	PR
SD	No suspected progression	NA	SD
NE	No suspected progression	NA	NE
PD	No suspected progression	NA	PD
Any	Suspected progression	PD	PD
Any	Suspected progression	not PD	overall RECIST response
Missing	No suspected progression	NA	Evaluable
Missing	Suspected progression	not PD	Evaluable
Missing	Suspected progression	PD	PD
Missing	Suspected progression	missing	Non evaluable

Any	Suspected progression	missing	overall RECIST response

CR Complete response; Evaluable Assessment was performed in the absence of a RECIST scan; NA Not Applicable; NE Not evaluable; PD Progressive disease; PR Partial response; SD Stable disease

3.2 Efficacy variables

3.2.1 Progression-free survival (PFS)

The primary endpoint of the trial is PFS. PFS (per RECIST 1.1 as assessed by investigator tumor assessments or histopathologic confirmation of local tumor progression) is defined as the time from randomization until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the subject withdraws from randomized therapy or receives another anti-cancer therapy prior to progression (i.e., date of PFS event or censoring – date of randomization + 1). Objective disease progression is further detailed as one of the following:

- Local progression
- Distant disease progression
- Secondary malignancy

If two progression events are reported but more than 2 months has elapsed between the events than the events will be regarded as two separate progression events and the earlier event and corresponding date of event will be used in the derivation of PFS.

Subjects who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment at which the RECIST 1.1 assessment and/or response evaluation physical exam were evaluable. However, if the subject progresses or dies immediately after two or more consecutive missed visits, the subject will be censored at the time of the latest assessment at which the RECIST 1.1 assessment and/or response assessment physical exam were evaluable prior to the two missed visits (Note: NE visit is not considered as missed visit).

A physical exam will be considered non evaluable if a physical exam is performed and progression is suspected and a biopsy is not reported. Similarly, in the absence of histopathologic confirmation of local tumor progression, an assessment will be considered non-evaluable if the RECIST visit assessment is not performed.

A missed visit will only occur if no scan and no physical exam are performed.

In the absence of histopathologic confirmation of disease progression, given the scheduled visit assessment scheme (i.e., at 20 weeks following randomization then every 12 weeks until week 164 then every 24 weeks thereafter), the definition of two missed RECIST 1.1 assessment visits will be as follows:

- If the previous evaluable assessment is <= day 133 (i.e., week 19) then two missed visits will equate to 33 weeks (i.e., 20 weeks + 12 weeks + 1 week allowing for a late visit at the end of the period)

- If the previous evaluable assessment is > day 133 and <= day 1,057 (i.e., week 151) then two missed visits will equate to 26 weeks (i.e., 2 x 12 weeks + 1 week allowing for an early assessment at the start of the period + 1 week allowing for a late assessment at the end of the period).

- If the two missed visits occur over the period where the visit schedule switches from 12-weekly to 24-weekly this will equate to 39 weeks (i.e., take the average of 12 and 24 to get 18 weeks and apply the same rationale, hence 2 x 18 weeks + 1 week allowing for an early visit at the start of the period + 2 weeks allowing for a late visit at the end of the period). The time period for the previous evaluable exam will be from days 1,058 to 1,141 (i.e., weeks 151 to 163).

- From day 1,142 (week 163) onwards (when the scheduling changes to 24-weekly assessments) two missing visits will equate to 52 weeks (i.e., 2 x 24 weeks + 2 weeks allowing for an early visit at the start of the period + 2 weeks allowing for a late visit at the end of the period).

If the subject has no evaluable visits or does not have a baseline RECIST scan, they will be censored at Day 1 unless they die within two visits of baseline (32 weeks plus 1 week allowing for a late assessment within the visit window); then they will be treated as an event with date of death as the event date.

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

Tumor assessments and physical exams contributing towards a particular visit may be performed on different dates within a 4-week window for the specific timepoint assessment. The following rules will be applied:

- For progression by histopathologic evidence of disease, date of progression will be determined based on the date of the physical exam which showed progression (including pap smear findings for which the date of pap smear would be used for progression).

- For on-study tumor assessments, the date of progression will be determined based on the earliest date of the scan for RECIST 1.1 defined radiological progression or progression seen on physical exam

- For BICR sensitivity analysis, date of progression will be determined based on the earliest of the dates of the component that triggered the progression on the first set of scans that indicates progression for the adjudicated reviewer selecting PD, or of either reviewer where both reviewers select PD as a timepoint response, and there is no adjudication for central review (BICR) data

- For both on-study tumor assessments and the BICR sensitivity analysis, when censoring a subject for PFS the subject will be censored at the latest of the dates contributing to a particular overall visit assessment

Note: for TLs only, the latest scan date is recorded out of all scans performed at that assessment for the TLs and similarly for NTLs only the latest scan date is recorded out of all scans performed at that assessment for the NTLs.

3.2.1.1 Time to first and second subsequent therapy or death

As a supportive summary to PFS, time to first subsequent therapy or death (TFST) is defined as the time from the date of randomization to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomized treatment, or death (i.e. date of first subsequent cancer therapy/death or censoring – date of randomization + 1). Any subject not known to have had a first subsequent anti-cancer therapy will be censored at the last date that the subject was known not to have received a first subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a subject terminated the study for reason other than death before first subsequent therapy, these subjects will be censored at the earliest of their last known to be alive and termination dates. Subjects not receiving randomized treatment would have TFST calculated in the same way (i.e., time from date of randomization to the subsequent therapy or death).

Time to second subsequent therapy or death (TSST) is defined as the time from the date of randomization to the earlier of start date of the second subsequent anti-cancer therapy after discontinuation of first subsequent treatment, or death (i.e. date of second subsequent cancer therapy/death or censoring – date of randomization + 1). Any subject not known to have had a second anti-cancer subsequent therapy will be censored at the last date that the subject was

known not to have received a second subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a subject terminated the study for reason other than death before second subsequent therapy, these subjects will be censored at the earliest of their last known to be alive and termination dates. Subjects not receiving randomized treatment would have TSST calculated in the same way, i.e. time from date of randomization to the subsequent therapy or death.

3.2.2 Progression-free survival (3 years)

Progression-free survival (3 years) will be defined as the Kaplan-Meier estimate of PFS at 3 years.

3.2.3 Overall survival (OS)

Overall survival is defined as the time from the date of randomization until death due to any cause regardless of whether the subject withdraws from randomized therapy or receives another anti-cancer therapy (i.e. date of death or censoring – date of randomization + 1). Any subject not known to have died at the time of analysis will be censored based on the last recorded date on which the subject was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF). Note: Survival calls will be made in the week following the date of DCO for each analysis, and if subjects are confirmed to be alive or if the death date is post the DCO date these subjects will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" subjects at the time of the final OS analysis should be obtained by the site personnel by checking the subject's notes, hospital records, contacting the subject's general practitioner and checking publicly-available death registries. In the event that the subject has actively withdrawn consent to the processing of their personal data, the vital status of the subject 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 subject was known to be alive for those subjects still on treatment (since the SURVIVE module is only completed for subjects off treatment if a survival sweep is not performed). The last date for each individual subject is defined as the latest among the following dates recorded on the case report forms (CRFs):

- AE start 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 anti-cancer treatment
- Date last known alive on survival status CRF
- End of study date

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

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.

3.2.4 Objective response rate (ORR)

To fulfil the analysis of the secondary endpoint of ORR, ORR (per RECIST 1.1 using investigator assessments or pathologic assessment of local disease progression) is defined as the percentage of subjects with at least one investigator-assessed overall visit response of CR or PR (Table 9) and will be based on a subset of all randomized subjects with measurable disease at baseline per the site investigator.

Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Subjects who discontinue randomized treatment without progression, receive a subsequent anti-cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) and then respond will not be included as responders in the ORR.

In addition to ORR, an analysis of confirmed ORR (per RECIST 1.1 using investigator assessments or pathologic assessment of local disease progression) will also be performed. Confirmed ORR is defined as the percentage of subjects with at least one confirmed visit

response of CR or PR. Confirmed ORR is intended as a supplementary analysis to ORR but will not be used for the interpretation of the secondary endpoint of ORR.

ORR will also be analyzed for all randomized subjects, but this will not be used for the interpretation of the secondary endpoint of ORR.

The number of subjects who received surgery at the site of the cervical cancer will be produced.

If applicable, an additional analysis of ORR will be performed in which subjects who have surgery that results in the absence of any tumor after surgery are treated as a CR in the derivation. It is reasonable to assume in this setting that the intervention of therapy has contributed to the ability to perform the surgery and thus the disappearance of the tumor.

3.2.5 Complete response (CR) rate

CR rate (per RECIST 1.1 using investigator assessments or pathologic assessment of local disease progression) is defined as the percentage of subjects with an overall visit response of CR (as defined in Table 9).

3.2.6 Duration of response (DoR) for complete responders

DoR in complete responders (per RECIST 1.1 using Investigator assessments, physical examination, and histopathological confirmation of local disease) will be defined as the time from the date of first documented complete response prior to starting subsequent anti-cancer therapy, until date of documented progression or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of CR.

If a subject does not progress following a response, then their DoR will use the PFS censoring time.

Similarly, the duration of response for subjects with CR or PR will also be defined as per the duration of response for CR responders only.

3.2.7 Incidence of local progression, distant disease progression, and secondary malignancy

The incidence of local progression, distant disease progression, and secondary malignancy will each be considered separately and defined as the number and percentage of subjects with

documented evidence of local progression, distant disease progression, and secondary malignancy.

The time from randomization to the date of first documented local progression, distant disease progression/secondary malignancy or death from any cause will also be summarized.

For the summaries, if two progression events (local progression and distant disease progression) occur within 2 months of each other, then this is referred to as a simultaneous event and will be considered a single event. In this situation the worst case will be taken as the event 'type', but the date of recurrence will be the earliest of the two events. For example, if local progression and distant progression are reported within a 2 month period then for the analysis and reporting of incidence of local progression, distant disease, and secondary malignancy and time to date of first local progression, distant disease progression, secondary malignancy or death from any cause, the event will be counted as distant disease progression, but the date will be the earliest date of the two events.

Distant disease progression, secondary malignancy, and deaths are considered to be competing risks for local progression. Local progression and death are considered to be competing risks for distant disease progression/secondary malignancy. For the time to event endpoint, distant disease progressions and secondary malignancies will be summarized together.

Local progressions will be determined by RECIST 1.1 or histopathologic confirmation of disease. Local progressions will be classified as any enlargement of the target or non-target lesions with no new lesions outside the radiation field. Distant disease progression should be diagnosed by RECIST 1.1 or histopathological confirmation of disease when metastatic lesion is easily accessible for biopsy. Distant disease progression will be classified as any new lesions outside the radiation field.

Secondary malignancies must be confirmed by biopsy. If a biopsy is not able to be performed, then the suspected secondary malignancy will be counted as a distant disease progression for the summary. Secondary malignancies are SAEs coded to the MedDRA system organ class neoplasms benign, malignant and unspecified (including cysts and polyps). All terms recorded in this system organ class at data cut-off are reviewed by AZ clinical and all terms considered secondary malignancies are flagged.

New cancers are those that are not the primary reason for the administration of the study treatment and have been identified after the subjects' inclusion in this study. They do not include metastases of the original cancer.

3.2.8 Best objective response (BoR)

Best objective response (BoR) is calculated based on the overall visit responses of CR, PR, SD, PD and NE as described in section 3.1.6. It is the best response a subject has had following randomization, but prior to starting any subsequent cancer therapy and up to and including RECIST progression or histopathologic confirmation of progression or the last evaluable assessment in the absence of RECIST progression or histopathologic confirmation of progression.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 20 weeks minus 1 week, i.e. at least 134 days (to allow for an early assessment within the assessment window), after randomization. For CR/PR, the initial overall visit assessment that showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

BoR will be determined programmatically based on RECIST, physical exam data, and confirmatory biopsy data from the overall visit response using all site investigator RECIST 1.1, physical exam data, and biopsy data up until the first progression event or start of subsequent anti-cancer therapy. The denominators for each case will be consistent with those used in the ORR analysis.

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

For subjects who die with no evaluable tumor assessments or confirmed histopathological local progression, if the death occurs ≤21 weeks (i.e., 20 weeks + 1 week to allow for a late assessment within the assessment window) after randomization, then BoR will be assigned to the progression (PD) category. For subjects who die with no evaluable tumor assessments or confirmed histopathological local progression, if the death occurs >21 weeks after randomization then BoR will be assigned to the NE category.

A subject will be classified as a responder if the criteria for an overall visit of CR or PR, are satisfied at any time following randomization, prior to progression and prior to starting any subsequent cancer therapy.

3.3 Patient reported outcome (PRO) variables

The following PRO questionnaires will be used to assess the subject experience, including global health status/health-related quality of life (HRQoL), functioning and symptoms: EORTC-QLQ-C30 with the EORTC-QLQ-CX24 cervical cancer module, PGIC, PGIS, PRO-CTCAE and EQ-5D-5L. All items/questionnaires will be scored according to published

guidelines or the developer's guidelines, if published guidelines are not available. All PRO analyses will be based on the FAS, unless stated otherwise.

3.3.1 EORTC-QLQ-C30

The EORTC-QLQ-C30 consists of 30 questions which can be combined to produce 5 functional scales (physical, role, cognitive, emotional, social), 3 multi-item symptom scales (fatigue, pain, nausea/vomiting), 5 individual symptom items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea) and a global measure of health status/QoL. The EORTC-QLQ-C30 will be scored according to the EORTC scoring manual (Fayers *et al.* 2001). An outcome variable consisting of a score of 0 to 100 will be derived for each of the symptom scales/items, the functional scales and the global health status/QoL scale of the EORTC-QLQ-C30 according to the EORTC-QLQ-C30 Scoring Manual. Higher scores on the global health status/QoL and functioning scales indicate better health status/function but higher scores on symptom scales/items represent greater symptom severity.

For each subscale, if ≤50% of the subscale items are missing, then the equation provided in the scoring manual can be applied excluding the missing items. If more than 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.

HRQoL will be assessed using the EORTC-QLQ-C30 global health status/QoL scale which includes 2 items from the QLQ-C30: "How would you rate your overall health during the past week?" (Item 29) and "How would you rate your overall quality of life during the past week?" (Item 30).

At each post-baseline assessment, the change from baseline in symptom and functional scale scores will be calculated.

At each post-baseline assessment, change in symptoms/functioning from baseline will be categorized as improved, stable, or worsened using the clinically meaningful cut-offs.

Definition of clinically meaningful changes

Changes in score compared with baseline will be evaluated. Changes in symptoms, functioning and global health status/QoL from baseline will be categorized as improved, stable or worsened as shown in Table 10. A clinically meaningful change is defined as a change of at least 10 points (Osoba *et al.* 1998). Specifically, a clinically meaningful improvement in a symptom scale/item score is defined as a decrease in the score from baseline of ≥10, whereas a clinically meaningful deterioration is defined as an increase in the score from baseline of ≥10. In contrast, a clinically meaningful improvement in a functional scale or global health status/QoL is defined as an increase in the score from baseline of ≥10, while a

clinically meaningful deterioration in a functional scale or global health status/QoL is defined as a decrease in the score from baseline of ≥10. At each post-baseline assessment, the change in symptoms, functioning, and global health status/QoL from baseline will be categorized as worsened, improved or stable as shown in Table 10.

Table 10: Visit responses for EORTC QLQ-C30

Score	Change from baseline	Visit response
Symptom scales/items	≥+10	Worsened
	≤-10	Improved
	Otherwise	Stable
Functional scales and global health status/QoL	≥+10	Improved
	≤-10	Worsened
	Otherwise	Stable

QoL Quality of life

3.3.2 EORTC QLQ-CX24

The QLQ-CX24 consists of 24 questions that can be combined to produce three multi-item scales (symptom experience, body image, and sexual/vaginal functioning) and six single-item scales (lymphedema, peripheral neuropathy, menopausal symptoms, sexual activity, sexual worry, and sexual enjoyment). The scales and single item measures are also scored from 0 to 100. In sexual activity and sexual enjoyment, higher scores indicate better functioning, however, in the other scales and single-items, higher scores reflect more problematic functioning or status.

At each post-baseline assessment, the change from baseline in symptom and functional scale scores will be calculated.

At each post-baseline assessment, change in symptoms/functioning from baseline will be categorized as improved, stable, or worsened using the clinically meaningful cut-offs.

Definition of clinically meaningful changes

Changes in score compared with baseline will be evaluated. Changes in multi-item scales and single item scales from baseline will be categorized as improved, stable or worsened as shown in Table 11. A minimum clinically meaningful change is defined as a change of 10 points. Specifically, a clinically meaningful improvement in a symptom scale/item score is defined as a decrease in the score from baseline of ≥10, whereas a clinically meaningful deterioration is

defined as an increase in the score from baseline of ≥10. In contrast, a clinically meaningful improvement in a functional item is defined as an increase in the score from baseline of ≥10, while a clinically meaningful deterioration in a functional scale or global health status is defined as a decrease in the score from baseline of ≥10.

Table 11: Visit responses for EORTC QLQ-CX24

Score	Change from baseline	Visit response
Symptom scales/items	≥+10	Worsened
	≤-10	Improved
	Otherwise	Stable
Functional items: Sexual activity/Sexual enjoyment	≥+10	Improved
	≤-10	Worsened
	Otherwise	Stable

3.3.3 Patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE)

The patient reported outcomes version of the common criteria for adverse events (PRO-CTCAE), a PRO version of the CTCAE system developed by the NCI, is included to assess tolerability from the subject's perspective. It was developed in recognition that collecting symptom data directly from subjects can improve the accuracy and efficiency of symptomatic AE data collection. Symptoms have been converted to subject terms (e.g., CTCAE term "myalgia" converted to "aching muscles"). Items capture the presence, frequency, severity and/or interference with usual activities, depending on the AE. Six items that are considered relevant for the trial were selected (CSP Appendix H). For each question, subjects select the value that best describes their experience over the past week.

The PRO-CTCAE items are not currently scored. The data will be descriptively summarized.

3.3.4 Patient global impression of change (PGIC) and patient global impression of severity (PGIS)

The response options of the PGIC are scored as follows: Much Better (+3), Moderately Better (+2), A Little Better (+1), No Change (0), A Little Worse (-1), Moderately Worse (-2) and Much Worse (-3).

The response options of the PGIS are scored using a 6-point scale: 1 = None; 2 = Very Mild; 3 = Mild; 4 = Moderate; 5 = Severe; 6 = Very Severe.

3.3.5 Health state utility

The EQ-5D-5L will be used to explore the impact of treatment and disease state on health state utility.

The EQ-5D-5L, developed by the EuroQol Group, is a generic questionnaire that provides a simple descriptive profile of health and a single index value for health status for economic appraisal. The EQ-5D-5L questionnaire comprises six questions that cover five 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 five options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems). A unique EQ-5D health state, termed the EQ-5D-5L profile, is reported as a five-digit code with a possible 3,125 health states. For example, state 11111 indicates no problems on any of the five dimensions. Respondents also assess their health today using the EQ-VAS, which ranges from 0 (worst imaginable health) to 100 (best imaginable health).

The EQ-5D profile will be converted into a weighted health state utility value, termed the EQ-5D index, by applying a country-specific equation to the EQ-5D-5L profile that represents the comparative value of health states. This equation is based on national valuation sets elicited from the general population and the base case will be the UK perspective. Where a valuation set has not been published, the EQ-5D-5L profile will be converted to the EQ-5D index using a crosswalk algorithm (van Hout *et al.* 2012). The EQ-VAS is reported separately. Further details regarding the evaluation of EQ-5D-5L will be presented in the payer analyses plan.

The evaluable population will comprise a subset of the FAS who have a baseline EQ-5D-5L assessment.

3.3.6 Compliance

Summary measures of overall compliance and compliance over time will be derived for EORTC-QLQ-C30, EORTC QLQ-CX24 and EQ-5D-5L respectively. These will be based upon:

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

- Expected questionnaire: A questionnaire that is expected to be completed at a scheduled assessment time e.g., a questionnaire from a subject who has not withdrawn from the study at the scheduled assessment time but excluding subjects in countries with no available translation. For subjects that have progressed, the latest of progression and last dose of IP follow-up (6 months following progression and 6 months following last dose

of IP) will be used to assess whether the subject 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 forms

- Evaluable questionnaire: A questionnaire with a completion date and at least one subscale that is non-missing

- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all time points, divided by total number of questionnaires expected to be received across all time points multiplied by 100

- Overall subject compliance rate is defined for each randomized treatment group as: Total number of subjects with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of subjects expected to have completed at least a baseline questionnaire multiplied by 100

Compliance over time will be calculated separately for each visit, including baseline, as the number of subjects with an evaluable questionnaire at the time point (as defined above), divided by number of subjects still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires (per definition above), divided by the number of received questionnaires.

3.4 Health care resource use variables

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

- Planned and unplanned hospital attendances beyond trial protocol mandated visits (including physician visits, emergency room visits, day cases and admissions)

- Primary sign or symptom the subject presents with

- Length of hospital stay

- Length of any time spent in an intensive care unit (ICU)

Where admitted overnight, the length of hospital stay will be calculated as the difference between the date of hospital discharge (or death date) and the start date of hospitalization or start of study drug if the start of study drug is after start date of hospitalization (length of hospital stay = end date of hospitalization – start date of hospitalization + 1). Subjects with

missing discharge dates will be calculated as the difference between the last day with available data and the start date of hospitalization. The length of ICU stay will be calculated using the same method.

3.5 Safety variables

Safety is a secondary outcome for this study. Safety and tolerability will be assessed in terms of adverse events (AEs) [including serious adverse events (SAEs)], deaths, physical examinations, laboratory findings, vital signs, electrocardiograms (ECGs) and exposure, which will be collected for all subjects.

Data from all cycles of treatment will be combined in the presentation of safety. The SAF will be used for reporting of safety data.

3.5.1 General considerations for safety assessments

3.5.1.1 Definition of baseline

Baseline will be defined as the last assessment of the variable under consideration prior to the intake of the first dose of study treatment (durvalumab, placebo, chemotherapy, radiotherapy). That is, the latest result prior to the start of study treatment. If two visits are equally eligible to assess subject status at baseline (e.g., screening and baseline assessments both on the same date prior to the first dose with no washout or other intervention in the screening period), the average will be used as the baseline value. For non-numeric laboratory tests where taking the average is not possible, the best value would be taken as baseline as this is most conservative. In the scenario where there are two assessment recorded on the day, one with time recorded and the other without time recorded, the one with the time recorded would be selected as baseline. Where safety data are summarized over time, time on study will be calculated in relation to date of first study treatment.

3.5.1.2 Time windows for safety and PRO data

Time windows will be defined for all presentations of safety and PRO data that summarize values by visit according to the following conventions.

- The time windows should be exhaustive so that data recorded at any time point (scheduled or unscheduled) has the potential to be summarized. Inclusion within the time window should be based on the actual data and not the intended date of the visit

- The window for 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 exist between two consecutive visits, then the upper limit will be taken as the midpoint value minus 1 day. For

demonstration purposes, Table 12 shows the visit windows for vital signs to week 40. Note that assessments will continue until study end.

Table 12: Analysis visits and visit windows

Cycle	Scheduled day	Analysis window (day)
1	Day 1	See Section 3.5.1.1 for baseline definition
2	Day 29	2 to 43
3	Day 57	44 to 71
4	Day 85	72 to 99
5	Day 113	100 to 127
6	Day 141	128 to 155
7	Day 169	156 to 183
8	Day197	184 to 211
9	Day 225	212 to 239
10	Day 253	240 to 267
11	Day 281	268 to 295

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment (as defined in Section 4.3.5.5) will be used (regardless of where it falls in an interval)
- Listings will display all values contributing to a time point for a subject
- For visit-based summaries, if there is more than one value per subject within a time window then the closest value to the scheduled visit date will be summarized. If the values are equidistant from the nominal visit date, then the earlier value will be used. Data listings will highlight the values used in 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 which value is closest to the scheduled visit date

- For summaries at subject level, all values will be included when deriving a subject level statistic such as a maximum regardless of whether they appear in the corresponding visit-based summary

3.5.1.3 Handling of missing data

Missing safety data will generally not be imputed.

Safety assessments 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 will be displayed as "<x" or ">x" in the listings.

For missing diagnostic dates, if day and/or month are missing use 01 and/or Jan. If year is missing, put the complete date to missing.

Adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

For missing start dates for AEs and concomitant medications/procedures, the following will be applied:

- Missing day: Impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date

- Missing day and month: Impute 1st January unless year is the same as first dose date then impute first dose date

- Completely missing date: Impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date

When imputing a start date, ensure that the new imputed date is sensible e.g., prior to the end date of the AE.

For missing stop dates of AEs or concomitant medications/procedures, the following will be applied:

- Missing day: Impute the last day of the month unless month is the same as month of last dose of study drug then impute last dose date

- Missing day and month: Impute 31st December unless year is the same as last dose date then impute last dose date

For completely missing stop dates for a medication, the following will be applied:

- Check whether the medication is still ongoing and when it started in relation to study drug before imputing a date. If the ongoing flag is missing, then assume that the medication is still being taken (i.e. do not impute a date). If the medication has stopped and its start date is prior to first dose date then impute the first dose date; if it started on or after first dose date then impute to the day after the last dose date (first dose date/last dose date of maintenance will take the priority).

Flags will be retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations would not be calculated.

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

- Missing day only: Using the 1st of the month

- Missing day and month: Using the 1st January

Subjects with a partial date of birth (i.e., for those countries where year of birth only is given) will have 1st of the month imputed if the day is missing, and 1st Jan imputed if the day and month is missing.

For partial subsequent anti-cancer therapy dates, the following will be applied:

- Missing day: If the month is the same as treatment end date then impute to the day after treatment, otherwise first day of the month

- Missing day and month: If year is the same as treatment end date then impute to the day after treatment, otherwise 1st January of the same year as anti-cancer therapy date

3.5.1.4 Safety Follow-up

Total Safety Follow-up = min (last dose date +90, date of withdrawal of consent, date of death, date of DCO) – first dose date +1

3.5.2 Study treatments

Study treatment in this study refers to chemotherapy and radiation (CCRT), durvalumab, and placebo. See Table 12 for further details on the IPs. Exposure will be defined for durvalumab or placebo and the SoC CCRTs outlined in Table 13.

Table 13: Study treatments

			Chemoradiotherapy		
	Durvalumab	**Placebo**	**EBRT + brachytherapy**	**Chemotherapy[f]**	
Study treatment name:	Durvalumab (MEDI4736)[a]	Saline solution	External beam radiotherapy and brachytherapy	Cisplatin	Carboplatin
Dosage formulation:	500-mg vial solution for infusion after dilution, 50 mg/mL	Sterile solution of 0.9% (w/v) sodium chloride for injection	As sourced locally	As sourced locally	As sourced locally
Route of administration	IV	IV	Whole pelvic or pelvic and para-aortic radiation and brachytherapy[b]	IV	IV
Dosing instructions:	1500 mg IV q4w	Dosing to match durvalumab	As described in Appendix G of the CSP	40 mg/m^2 q1w × 5 weeks[c,d]	AUC 2 q1w × 5 weeks[d]
Provider	AstraZeneca	Sourced locally by site	Sourced locally by site[e]	Sourced locally by site[e]	Sourced locally by site[e]

[a] Label text prepared for durvalumab (MEDI4736) will show the product name as "MEDI4736" or "durvalumab (MEDI4736)" depending upon the agreed product name used in the approved study master label document. All naming conventions are correct during this transitional period
[b] Pelvic and paraaortic radiotherapy is added based on extent of disease at baseline
[c] Standard antiemetic regimens consist of a steroid and 5HT-antagonist, with or without a substance P antagonist such as aprepitant
[d] An additional 6th week of platinum agent will be administered per investigator discretion
[e] Under certain circumstances when local sourcing is not feasible, an SoC treatment may be supplied centrally through AstraZeneca
[f] For subjects who were treated with carboplatin prior to CSP version 3 see the explanatory paragraph below

EBRT External beam radiotherapy; Gy Grey unit; IV Intravenous; n/a Not applicable; qxw Every x weeks; SoC Standard of care; w/v Weight/volume

In versions of the protocol prior to version 3 carboplatin AUC2 was given for initial treatment or if cisplatin was intolerable. Subjects recruited to the study under this regimen may continue unchanged, as clinically indicated. Carboplatin is to be given AUC 2 q1w × 5 weeks, and an additional 6th week will be administered per investigator discretion. Antiemetics are recommended. Standard antiemetic regimens consist of a steroid and 5HT-antagonist, with or without a substance Pantagonist such as aprepitant.

3.5.3 Exposure and dose interruptions

Exposure variables comprise the measure of treatment durations, number of infusions/cycles received, counts of dose delays and dose interruptions, and dose intakes.

3.5.3.1 Treatment exposure for durvalumab or placebo

Since durvalumab or placebo is dosed every 4 weeks for 24 doses, calculation of exposure (otherwise known as treatment duration) will be defined as follows:

Total (or intended) treatment duration of durvalumab/placebo:

- Total (or intended) treatment duration (days) = min (last dose date where dose > 0mg + 27 days, date of death, date of DCO) – first dose date +1

Actual treatment duration of durvalumab/placebo:

Actual treatment duration (days) = intended treatment duration (days) – total duration of dose delays (days), where intended treatment duration will be calculated as above.

Dose interruptions - infusion

For durvalumab/placebo, a dose interruption is an infusion interruption that occurs during the infusion. The total dose received is >0. The drug can be restarted after the interruption and so it is possible for an infusion interruption to occur and the whole dose to still be administered. If the same infusion was interrupted multiple times, then this would just be captured as one infusion interruption.

Dose delays

A treatment cycle is started when >0 dose of durvalumab/placebo is administered. As such, a dose delay for durvalumab/placebo occurs when the start of a cycle is started at a later date than planned. If durvalumab/placebo is delayed, leading to other drugs that were scheduled to be administered on the same day being administered at a later date (but still the same day that durvalumab/placebo is eventually administered), then in this instance only

durvalumab/placebo is classed as being delayed. This is because the other drugs would have been administered on the correct day relative to durvalumab/placebo.

Dose reductions
Doses that are intentionally permanently reduced are not permitted per CSP for durvalumab/placebo.

Calculation of duration of dose delays (for actual treatment duration)

- Since subjects will receive durvalumab/placebo 1500 mg via IV infusion q4w (for 24 doses), the duration of dose delays will be calculated as follows:
    - For all dosing dates: Total duration of dose delays (days) = Sum of (Date of the dose – Date of previous dose – 28 days)

Thus, if no delays were encountered, the duration would sum up to 0, since infusions were done every 4 weeks (28 days).

Number of infusions
Number of infusions of durvalumab/placebo:

Exposure will also be summarized by the number of infusions received. Cycles of treatment with durvalumab/placebo are of 28 days duration with a single dose on day 1 of each cycle. If a cycle is prolonged due to toxicity, this should be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

Received planned starting dose
A subject is considered to have received their planned starting dose of durvalumab or placebo if the infusion volume before administration is non-zero and the infusion volume after administration = 0.

3.5.3.2 Treatment exposure for SoCs and Radiotherapy

Carboplatin and cisplatin

SoC platinum-based chemotherapy agents, carboplatin and cisplatin, are dosed weekly for 5 weeks, based on Investigator discretion, in addition to radiation therapy. A sixth cycle of chemotherapy is allowed per protocol at investigator discretion. Exposure to carboplatin and cisplatin (otherwise known as treatment duration) will be defined as follows:

Total (or intended) exposure of carboplatin and cisplatin will be summarized separately for subjects who receive 5 or 6 cycles of chemotherapy.

Total (or intended) treatment duration (days) of carboplatin and cisplatin:

- Minimum of (infusion date of the last cycle + 6 days, date of death, date of DCO) – first infusion date of first cycle + 1

Number of doses of carboplatin/cisplatin:

- Exposure will be measured by the number of doses received. A cycle of carboplatin/cisplatin corresponds to a period of one week with a single dose administered. 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

The number of subjects who are initially treated with carboplatin and switch to cisplatin and who are initially treated with cisplatin and switch to carboplatin during the course of the trial will be summarized.

Received planned starting dose
A subject is considered to have received their planned starting dose of carboplatin or cisplatin if the infusion volume before administration is non-zero and the infusion volume after administration = 0.

Dose reductions
Dose reductions, doses that are intentionally permanently reduced, are permitted as per CSP. This term is not used for interruptions or invalidly administered doses. A dose reduction is counted once for each time the dose is reduced.

Dose delays
A dose delay for carboplatin or cisplatin occurs when the first administration of that drug (> 0 dose) in a cycle is administered at a later date relative to the durvalumab/placebo dose. Note that if the drug is completely skipped then this is not classed as a delay (it is classed as a dose interruption).

Dose interruptions - infusion
An infusion interruption of carboplatin or cisplatin is defined the same as for durvalumab/placebo.

Dose interruptions - skipped doses
Since carboplatin and cisplatin are administered multiple times per cycle, a skipped dose is a temporary interruption during a cycle. That is, during the cycle a dose is completely skipped or is taken at a later date than scheduled. Note that this is only applicable to drugs with multiple doses in a cycle (if the first dose is later than planned it would be a delay).

Radiation
Total Dose of Radiotherapy in Grays administered for EBRT will be calculated Fraction Dose multiplied by Number of Fraction doses. This will be for the pelvis only (i.e., of the 45 grays).

The Total Biological Equivalent Dose or EQD2 dose to the target area administered for brachytherapy will be determined from the QS module of the CRF according to whether the subject is receiving Point A directed brachytherapy or volumetric brachytherapy (QSTEST = Point A EQD2 for Point A directed brachytherapy or QSTEST = CTV_HR 90 EQD2 for volumetric brachytherapy).

Total (or intended) exposure

- Minimum of (radiotherapy stop date, date of death, date of DCO) – radiotherapy start date + 1

Actual exposure:

- Actual exposure = intended exposure – total duration of dose delays, where intended exposure will be calculated as above, and a dose interruption is defined as any length of time where the subject has not taken any of the planned dose (taking into account the scheduled off treatment period)

Subjects who permanently discontinue during a dose interruption
If a subject permanently discontinues study treatment in-between cycles or during a dose delay, then the date of last administration of study medication recorded on DOSDISC will be used in the programming.

3.5.3.3 Dose intensity

Dose intensity will be derived for study treatment durvalumab/placebo. Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to last day of dosing. RDI will be defined as follows:

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing for that drug 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 deriving actual dose administered the volume before and after infusion will also be considered.

For durvalumab/placebo, where the last dose is on the cycle 24 visit and if there are no scans post cycle 24, the censoring of progression should occur at cycle 24 for the purposes of RDI.

An example of dose intensity for durvalumab can be found in Table 14. In this example, RDI of durvalumab: d = 5200mg, D= 7500mg. RDI= 100%*(5200/7500) = 69.3%.

Table 14: Dose intensity scenarios for durvalumab administered 1500mg per cycle

Cycle	Date	Duration (days)	Delay (days)	No. cycles, based on duration + delays (A)	No. actual cycles received (B)	Planned total dose D = Max(int(A), B)*1500mg	Actual dose intake d
1	01-Jan-19	28	+ 14				1500
2	12-Feb-19	28	+ 14				1000
3	26-Mar-19	28	+ 0				1500
4	23-Apr-19	28	+ 0				1200
Total		4*28 + 2*14 = 140		140/28 = 5	4	5*1500 = 7500mg	5200mg

3.5.4 Adverse events (AEs)

An AE is the development of any untoward medical occurrence (other than progression of the malignancy under evaluation) in a subject or clinical study subject administered a medicinal product and which does not necessarily have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (e.g., an abnormal laboratory finding), symptom (for example nausea, chest pain), or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product.

The term AE is used to include both serious and non-serious AEs and can include a deterioration of a pre-existing medical occurrence. An AE may occur at any time, including run-in or washout periods, even if no study treatment has been administered.

AEs and SAEs will be collected throughout this study. For this study, on treatment will be defined between date of first dose of study treatment (durvalumab, matching placebo, chemotherapy, radiotherapy) and 90 days following the last dose of study treatment. If an event starts outside of this period and it is considered possible that it is due to late onset toxicity to study treatment, then it should be reported as an AE or SAE.

On treatment AEs (or treatment emergent AEs [TEAEs]) will be defined as any AEs that started after dosing or that started prior to dosing and worsened (by investigator report of a change in intensity) following exposure to treatment. If an AE is not worse than the baseline (pre-dose) severity, then it will not be classified at a TEAE.

The medical dictionary for regulatory activities (MedDRA) [using the latest or current MedDRA version] will be used to code AEs. AEs will be graded according to the National Cancer Institute (NCI) common terminology criteria for adverse event (CTCAE) version 5.0. The CTCAE grade will be assigned by the investigator as follows:

- Grade 1: Mild AE
- Grade 2: Moderate AE
- Grade 3: Severe AE
- Grade 4: Life-threatening or disabling AE
- Grade 5: Death related to AE

Missing start and stop dates for AEs will be handled using the rules described in Section 3.5.1.

3.5.5 Other significant adverse events (OAE)

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

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

3.5.6 AEs 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 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, Hypophysitis, 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.

Immune-mediated adverse events

Durvalumab belongs to a class of drugs called immune checkpoint inhibitors. Because the mechanism of action of this class of drugs is to block the inhibitory signals that prevent T-cell activation, this drug may potentially cause immune-mediated adverse events (imAEs). 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 than 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.

3.5.7 Laboratory measurements

Laboratory data will be collected throughout the study as described in Tables 1 and 2 of the CSP. Blood and urine samples for determination of clinical chemistry, hematology, and urinalysis will be collected as described in Section 8.2.4 of the CSP.

For the derivation of baseline and post baseline visit values, the rules described in Section 3.5.1 of this document considering definition of baseline, visit windows and how to handle multiple records will be used.

Change from baseline in hematology and clinical chemistry variables will be calculated for each post-dose visit on treatment. CTCAE grades will be defined at each visit according to the CTCAE grade criteria using local or project ranges as required, after conversion of lab result to corresponding AZ preferred units. The following parameters have CTCAE grades defined for both high and low values: Potassium, sodium, magnesium, and corrected calcium so high and low CTCAE grades will be calculated. Glucose has CTCAE grades defined only for low values and will be calculated as such.

Corrected calcium product 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)] x 0.02)

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

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

Project reference ranges will be used throughout for reporting purposes. If the project range is unavailable for a test, local ranges will be used. The denominator used in laboratory summaries of CTCAE grades will only include evaluable subjects (i.e., those who had sufficient data to have the possibility of an abnormality). For example,

- If a CTCAE criterion involves a change from baseline, evaluable subjects would have both a pre-dose and at least 1 post-dose value recorded
- If a CTCAE criterion does not consider changes from baseline, to be evaluable the subject needs only to have 1 post dose-value recorded

3.5.8 Vital signs

The following vital signs will be measured as described in Section 8.2.2 of the CSP: Systolic and diastolic blood pressure (BP), pulse rate, temperature, and respiratory rate. Body weight will also be recorded along with vital signs.

Vital signs will be collected at multiple times at same visit for the first infusion (pre-dose, during infusion, and at the end of infusion). At subsequent visits they may be taken at each of these timepoints as per institution and as clinically indicated.

Change from baseline in vital signs variables will be calculated for each post-baseline visit on treatment, for each timepoint as data are available (pre-dose, during infusion, end of infusion). Timepoints are reported by visit for each treatment arm, provided at least one treatment arm has >= 20 subjects with data at a given visit. For derivation of post-baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 3.5 will be used.

3.5.9 Physical examinations

Physical examinations will be performed as described in Section 8.2.1 of the CSP. Abnormalities recorded prior to the first dose of study treatment will be recorded as part of the subject's baseline signs and symptoms. Abnormalities first recorded after first dose of study treatment will be recorded as AEs unless unequivocally related to the disease under study.

3.5.10 Electrocardiograms (ECGs)

Resting 12-lead ECGs will be recoded at screening and as clinically indicated throughout the study as described in Section 8.2.3 of the CSP.

The following ECG variables will be collected: ECG heart rate, PR duration, QRS duration, QT duration, RR duration and overall ECG evaluation.

The overall evaluation of an ECG will either be "normal" or "abnormal" with abnormalities categorized as either "clinically significant" or "not clinically significant". In case of clinically significant ECG abnormalities, 2 additional ECGs will be obtained over a brief period (e.g., 30 minutes) to confirm the finding.

Where QTcF (Friderica) is not reported, it will be calculated programmatically using the ECG values (RR and QT) as follows (where RR are in seconds):

$$QTcF = \frac{QT}{\sqrt[3]{RR}}$$

Alternatively, RR (or QT) can be programmatically derived if not reported but QTcF and QT (or RR, respectively) is reported. RR can be calculated as follows:

$$RR = \left(\frac{QT}{QTcF}\right)^3$$

3.5.11 World Health Organisation (WHO)/Eastern Cooperative Oncology Group (ECOG) performance status (PS)

WHO/ECOG PS will be assessed as described in Section 8.2.5 of the CSP as the following:

0. Fully active; able to carry out all usual activities without restrictions

1. Restricted in strenuous activity, but ambulatory and able to carry out light work or work of a sedentary nature (e.g., light housework or office work)

2. Ambulatory and capable of self-care, but unable to carry out any activities; up and about more than 50% of waking hours

3. Capable of only limited self-care; confined to bed or chair more than 50% of waking hours

4. Completely disabled; unable to carry out any self-care and totally confined to bed or chair

5. Dead

Any significant changes from baseline or screening will be reported as AE.

3.6 Pharmacokinetic (PK) variables

PK concentration data will be collected as described in Section 8.5 of the CSP.

The actual sampling times will be used in the PK calculations. PK parameters, such as peak and trough concentration will be obtained from raw PK data measurements as data allow.

3.7 Immunogenicity variables

Serum samples for ADA assessments will be conducted utilizing a tiered approach (screen, confirm, titer), and ADA data will be collected at scheduled visits shown in the CSP (Section 8.5.1.1). ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titer will be reported as well. In addition, the presence of neutralizing antibody (nAb) may be tested for all ADA-positive samples using a ligand-binding assay. The nAb results will be reported as positive or negative.

The number of subjects in the ADA-evaluable set who fulfil the following the criteria will be determined. The percentage of ADA-positive subjects in each of the category will be calculated, using the number of subjects in the ADA-evaluable set (defined in Section 2.1.2) of the treatment group as the denominator. A subject is defined as being ADA-positive if a positive ADA result is available at any time, including baseline and all post-baseline measurements; otherwise ADA negative.

- ADA-positive at any visit; the percentage of ADA-positive subjects in the ADA-evaluable set is known as ADA prevalence
- Treatment-emergent ADA, which is defined as either treatment-induced or treatment-boosted ADA; the percentage of subjects fulfilling this criterion in the ADA-evaluable set is known as ADA incidence
- ADA-positive post-baseline and positive at baseline
- ADA-positive post-baseline and not detected at baseline (treatment-induced ADA)
- ADA not detected post-baseline and positive at baseline
- Treatment-boosted ADA, defined as a baseline positive ADA titer that was boosted to a 4-fold or higher level (greater than the analytical variance of the assay) following drug administration
- Persistently positive ADA, defined as having at least 2 post-baseline ADA-positive measurements with at least 16 weeks (112days) between the first and last positive measurement. or an ADA-positive result at the last available assessment. The category may include subjects meeting these criteria who are ADA-positive at baseline

- Transiently positive ADA, defined as having at least one post-baseline ADA-positive measurement and not fulfilling the conditions for persistently positive. The category may include subjects meeting these criteria who are ADA-positive at baseline
- nAb positive at any visit

3.8 Biomarkers

Blood samples for exploratory biomarker analyses will be obtained according to the schedules presented in Section 1.1 of the CSP.

Results of the exploratory biomarker data will be reported separately.

3.9 Prior and concomitant medications and therapies

All therapies (drug and non-drug), including herbal preparations, whether prescribed or over-the-counter, that are used within the four weeks prior to initiation of study treatment up until 90 days following last dose of study treatment will be recorded on the eCRF. Details include generic and/or brand names of the medications, World Health Organization Drug Dictionary (WHO-DD) encoding (using the latest or current WHO-DD version), reason for use, route, dose, dosing frequency, and start and stop times.

Prior therapies are defined as those taken prior to study treatment with a stop date prior to the first dose of study medication (durvalumab, matching placebo, chemotherapy or radiotherapy).

Concomitant therapies are defined as those with at least one dose/treatment taken between the date of first dose (inclusive) and the date of last dose (inclusive) of study drug.

Missing start and stop dates for medications will be handled using the rules in Section 3.5.1. Missing coding terms should be listed and summarized as "Not coded".

4 ANALYSIS METHODS

There will be 1 treatment comparison of interest:

- Durvalumab 1500 mg plus SoC CCRT vs placebo plus SoC CCRT

The primary endpoint is PFS using RECIST 1.1 and histopathologic confirmation of tumor progression. The study has been sized to characterize the PFS and OS benefit of durvalumab 1500 mg plus SoC CCRT relative to placebo plus SoC CCRT.

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

The formal statistical analysis will be performed to test the main hypotheses:

- H_0: No difference between durvalumab plus SoC CCRT and placebo plus SoC CCRT
- H_1: Difference between durvalumab plus SoC CCRT and placebo plus SoC CCRT

There will be up to 2 data cut-offs applied for this study:

1. The DCO for the PFS analysis and first OS interim (DCO 1): The DCO for the PFS analysis will be when approximately 227 (32% maturity) PFS events have occurred. The OS interim will occur at the time of the PFS analysis, if PFS is statistically significant. It is anticipated that approximately 86% of the OS events will be available for this OS IA (approximately 195 of 227 OS events).
2. If the criteria for statistical significance is met for PFS but not OS at the time of the analysis for PFS, follow-up will continue until the earlier of 227 OS events or 71 months following randomization of the first subject, at which point OS will be retested.

4.1 General principles

Efficacy and PRO data will be summarized and analyzed on the FAS. Safety and treatment exposure data will be summarized based upon the SAF. Study population and demography data will be summarized based upon the FAS. PK data will be analyzed using the PK analysis set.

Physical exam data will include pap smear data where it is available.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days.

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

- All analyses and reporting will be by treatment arm
- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarized by the number of observations, mean, standard deviation, median, upper and lower quartiles minimum, and maximum. For log-transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum

and maximum. Categorical variables will be summarized by frequency counts and percentages for each category

- Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment group. Overall totals will be calculated for baseline summaries only

- 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 categorical data, percentages will be rounded to 1 decimal place

- The primary source for the stratification factors will be those captured in the IWRS during randomization. However, source data will be used to check concordance against data entered into IWRS.

- The covariates in the statistical modelling will be based on the values entered into the IWRS at randomization, even if it is subsequently discovered that these values were incorrect

- SAS® version 9.1 (or higher) will be used for all analyses

In general, for efficacy 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. 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 the post-treatment value minus the value at baseline. The percentage change from baseline will be calculated as (post-baseline value - baseline value) / baseline value x 100.

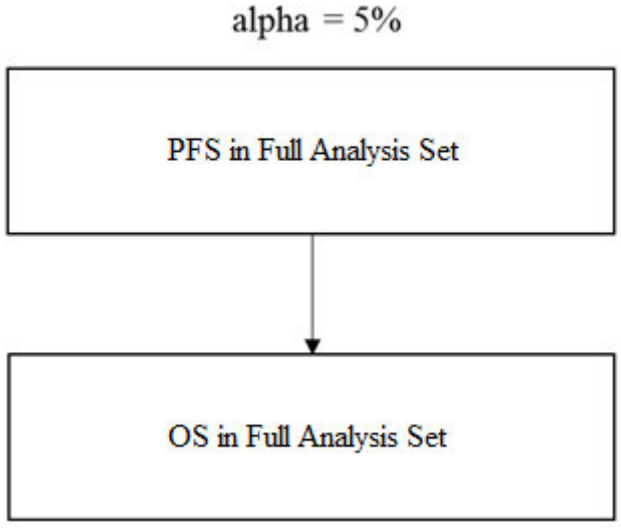
4.2 Multiplicity

The multiple testing procedure (Figure 2) will define which significance levels should be applied to the interpretation of the raw p-values.

In order to strongly control the Type I error at 5% (2-sided), a multiple testing procedure (MTP) will be applied to the primary endpoint of PFS and the key secondary endpoint of OS. The overall 5% type I error rate will first be allocated to test the primary endpoint of PFS for durvalumab plus SoC CCRT versus placebo plus SoC CCRT. If the primary endpoint of PFS is significant, 5% alpha will be recycled to the lower level of the hierarchy, where the 5% alpha will be used for the test of OS for durvalumab plus SoC CCRT versus placebo plus SoC CCRT.

Hypotheses will be tested using a multiple testing procedure with an alpha-exhaustive recycling strategy (Burman *et al.* 2009). With this approach, hypotheses will be tested in a pre-defined order with the hypothesis for PFS tested before the hypothesis for OS. According to alpha (test mass) splitting and alpha recycling, the test mass that becomes available after each rejected hypothesis is recycled to secondary hypotheses not yet rejected.

Figure 2: Multiple testing procedures for controlling the type I error rate



The primary endpoint of PFS will be tested once. The key secondary endpoint of OS will be tested up to 2 timepoints: 1 interim analysis and 1 final analysis. The tests including the interim and final analysis that are for the comparison of durvalumab plus SoC CCRT versus placebo plus SoC CCRT for the analyses of OS will be considered as one test family. As long as 1 test in the family can be rejected, the family is rejected. Thus, the assigned total alpha to the family will be recycled to the next MTP level.

This testing procedure stops when the entire mass is allocated to non-rejected hypotheses. Implementation of this pre-defined ordered testing procedure, including recycling, will strongly control type I error at 5% (2-sided), among the primary hypothesis of PFS and key secondary hypothesis of OS.

For the OS endpoint, there is 1 IA planned, and the alpha level will be controlled at the interim and primary analysis timepoints by using the Lan-DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, with the boundaries for the treatment comparison derived based upon the actual number of OS events observed at the time of analysis. It is anticipated that approximately 86% of the OS events will be available for this OS IA (approximately 195 of 227 OS events). If exactly 86% of the target events are available at the time of the interim analysis, with overall 2-sided alpha level of 5.0%, the 2-sided alpha to be applied at the interim analysis and final analysis would be 3.1% and 4.1%, respectively.

4.3 Analysis methods

4.3.1 Efficacy

Table 15 details which endpoints are to be subjected to formal statistical analysis, together with pre-planned sensitivity analyses, making it clear which analysis is regarded as primary for that endpoint.

Prior to unblinding, the number of subjects across both treatment groups in each level of each strata will be reviewed, and the planned stratification factors may be removed or have levels combined if too few subjects are represented in any cell.

Table 15: Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoints analyzed	Notes
Progression-free survival	Stratified log-rank analysis for: Primary analysis using Investigator RECIST 1.1 assessments or histopathologic confirmation of local tumor progression Sensitivity analyses using BICR assessments (RECIST 1.1) or histopathologic confirmation of local tumor progression Sensitivity analysis to assess possible evaluation-time bias using Investigator RECIST 1.1 assessments or histopathologic confirmation of local tumor progression Sensitivity to assess attrition bias using Investigator RECIST 1.1 assessments or histopathologic confirmation of local tumor progression Sensitivity analysis to assess bias introduced by subjects with suspected progression on physical exam and no confirmatory biopsy Sensitivity analysis to assess impact of non-proportional hazards using the max-combo test Sensitivity analysis to assess impact of COVID-19 deaths (see Section 4.4.1 for further details) Sensitivity analysis to assess bias introduced by mis-stratification
PFS (3 year)	Kaplan Meier estimate of PFS at 3 years

Endpoints analyzed	Notes
PFS in PD-L1 high subjects	Stratified log-rank analysis for: PFS in PD-L1 high subjects
Overall survival	Stratified log-rank analysis for: Overall survival Sensitivity analysis to assess impact of COVID-19 deaths (see Section 4.4.2 for further details)
Overall survival in PD-L1 high subjects	Stratified log-rank analysis for: Overall survival in PD-L1 high subjects
Objective response rate	Secondary analysis for the ITT population Logistic regression for: Secondary analysis for the ITT population using Investigator RECIST 1.1 assessments or histopathologic confirmation of local tumor progression
Complete response rate	Logistic regression for: Secondary analysis for the ITT population using Investigator RECIST 1.1 assessments or histopathologic confirmation of local tumor progression
Duration of response in subjects with a complete response	Summary statistics and KM plot by treatment arm for: Secondary analysis for the ITT population using Investigator RECIST 1.1 assessments or histopathologic confirmation of local tumor progression in subjects who have achieved a best objective response of complete response

Endpoints analyzed	Notes
Incidence of local progression, distant disease progression, and secondary malignancy	Summary statistics and cumulative incidence plot by treatment arm for: Secondary analysis for the ITT population
Change from baseline (EORTC QLQ-C30 and EORTC QLQ-CX24)	Average change from baseline using a Mixed Model Repeated Measurements (MMRM) analysis
EQ-5D-5L (health state utility values and Visual Analog Scale)	Summary statistics for health state utilities and visual analogue scale, including change from baseline.
Healthcare resource use	Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages)

Prior to unblinding, the number of subjects across both treatment groups in each level of each strata will be reviewed, and the planned stratification factors may be removed or have levels combined if too few subjects are represented in any cell.

4.3.2 Primary efficacy endpoint

Progression-free Survival (PFS)

Analysis of the primary endpoint, PFS, will occur when 227 PFS events have occurred (32% maturity). PFS based on the investigator RECIST 1.1 and histopathologic confirmation of progression will be analyzed using a stratified log-rank test adjusting for disease stage status (FIGO (2009) Stage <III and node positive, FIGO (2009) Stage ≥III and node negative, or FIGO (2009) Stage ≥III and node positive), and region ([United States, Canada, European Union, South Korea, and Japan] versus rest of the world) for generation of the p-value and using a method that corresponds to the Breslow approach for handling ties (Breslow, 1974). The HR and its CI will be estimated from a stratified Cox proportional hazards model (Cox, 1972) with ties = Efron and the stratification variables included in the strata statement) and the CI calculated using the profile likelihood approach.

The effect of treatment will be estimated by the hazard ratio (HR) together with its corresponding 95% CI and p-value for the FAS (a HR < 1 will favor durvalumab + SoC CCRT).

KM plots of PFS will be presented by treatment arm and by treatment arm and PD-L1 tumor status. Summaries of the number and percentage of subjects experiencing a PFS event and type of event (RECIST 1.1, histopathologic confirmation of progression, or death) will be provided along with the median PFS and 95% CI for each treatment arm.

Assumptions of proportionality

The assumption of proportionality will be assessed. Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time dependent covariate (adding a treatment-by-time or treatment-by-ln(time) interaction term) to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect can be described by presenting piecewise HR calculated over distinct time-periods for example 0-6m, 6-12m etc. In such circumstances, the HR from the primary analysis 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 a treatment-by-covariate interaction, which will be investigated. In addition, the KM curve along with landmark analyses (e.g., 1-year PFS rate) will also help in understanding the treatment benefit.

PFS Sensitivity Analyses

The following sensitivity analyses will be performed:

1. Evaluation-time bias

A sensitivity analysis will be performed to assess possible evaluation-time bias that may be introduced if scans and physical exams are not performed at the protocol-scheduled time points. The midpoint between the time of progression and the previous evaluable assessment (using the final date of the assessment) will be analyzed using a stratified log-rank test, as described for the primary analysis of PFS. Note that midpoint values resulting in non-integer values should be rounded down. For subjects whose death was treated as a PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). To support this analysis, the mean of subject-level average inter-assessment times will be tabulated for each treatment. This approach will use the site investigator tumor assessments as well as physical examinations, and confirmatory biopsy results.

2. Attrition bias

Attrition bias will be assessed by repeating the PFS analysis except that the actual PFS event times, rather than the censored times, of subjects who progressed or died in the absence of progression immediately following two, or more, non-evaluable tumor assessments will be included. In addition, and within the same sensitivity analysis, subjects who take subsequent therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) prior to their last evaluable assessment or progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier (KM) plot of the time to censoring using the PFS data from the primary analysis and where the censoring indicator of the PFS analysis is reversed.

3. Ascertainment bias

Ascertainment bias will be assessed by analyzing the BICR data. The stratified log-rank test will be repeated on PFS using the BICR data based upon RECIST as well as the physical exam and confirmatory biopsy data. The HR and CI will be presented.

For RECIST progressions, if there is an important discrepancy between the primary analysis using the site investigator data and this sensitivity analysis using BICR data then the proportion of subjects with site but no central confirmation of progression will be summarized; such subjects have the potential to induce bias in the central review due to informative censoring. An approach of imputing an event at the next visit in the central review analysis may help inform the most likely HR value (Fleischer *et al*. 2011), but only if an important discrepancy exists.

Disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group.

4. Randomization bias

Cox proportional hazards modeling will be employed to assess the effect of the pre-specified covariates on the HR estimate. A model will be constructed, containing treatment and the stratification factor(s) alone, to ensure that any output from the Cox modeling is likely to be consistent with the results of the primary analysis using the stratified log-rank test. The result from the initial model and the model containing additional covariates (in the MODEL statement) will be presented.

Additional covariates for this model will be age at randomization, PD-L1 expression, race (Asian vs non-Asian), type of CCRT (IMRT vs other), first dose of chemotherapy (cisplatin vs carboplatin).

The model will include the effect regardless of whether the inclusion of effect significantly improves the fit of the model provided there is enough data to make them meaningful. Missing covariate data will be imputed using the mean (for continuous variables) or the most common category (for categorical factors).

This analysis evaluates the treatment effect adjusting for any potential imbalances in baseline prognostic factors that are not balanced by stratification.

A forest plot illustrating the hazard ratio and 95% confidence interval will be provided to compare the primary and sensitivity analyses of progression-free survival.

5. Suspected progression on physical exam and no confirmatory biopsy

For this analysis, subjects who do not progress by RECIST but have suspected progression by physical exam at an assessment and do not have a confirmatory biopsy will be imputed as progressions with the date of progression based on the date of the physical exam.

The stratified log-rank test will be repeated on PFS using the Investigator data based upon RECIST, physical exam and confirmatory biopsy data, with subjects who have no progression by RECIST and suspected progression by physical exam without confirmatory biopsy data imputed as progressions as described. The HR and CI will be presented.

6. Max-combo test

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

7. Stratification factors based on eCRF data

This sensitivity analysis will repeat the PFS analysis using the stratified log-rank test, as described for the primary analysis of PFS where the stratification factors will be based on the values recorded on the eCRF instead of the IWRS. This analysis will only be performed if more than 5% of subjects were mis-stratified.

Additional supportive summaries/figures

The treatment status at progression of subjects at the time of analysis will be summarized for the FAS. This will include the number (%) of subjects who were on treatment at the time of progression, the number (%) of subjects who have progressed before receiving treatment, the

number (%) of subjects who completed treatment prior to progression, the number (%) of subjects who discontinued study treatment prior to progression, the number (%) of subjects who have not progressed and were on treatment, did not receive treatment, completed treatment or discontinued treatment. This will also provide distribution of number of days prior to progression for the subjects who have discontinued treatment.

In addition, the number of subjects prematurely censored will be summarized by treatment arm together with baseline prognostic factors of the prematurely censored subjects. A subject would be defined as prematurely censored if they had not progressed (or died in the absence of progression) and the latest date of assessment at which the RECIST 1.1 assessment and response evaluation physical exam were evaluable prior to DCO was more than one scheduled assessment interval plus 2 weeks prior to the DCO date.

Additionally, summary statistics will be given for the number of days from censoring to DCO for all censored subjects.

A summary of the number of days between progression assessments (RECIST and physical exam) will be presented for each visit schedule for each treatment arm. A summary of the number of days between RECIST and physical exam assessments over time will also be presented for each treatment arm in the FAS. Furthermore, a summary of the time from last tumor assessment to DCO in censored subjects will be presented for each treatment arm in the FAS.

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

Additionally, summary statistics for the number of weeks between the time of progression and the last evaluable tumor assessment prior to progression will be presented for each treatment group.

Summaries of the number and percentage of subjects who miss two or more consecutive tumor assessments will be presented for each treatment group.

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

Subgroup analysis

Subgroup analyses will be conducted comparing PFS (per RECIST 1.1 using Investigator tumor assessments or histopathologic confirmation of local tumor progression) between

durvalumab plus SoC CCRT and placebo plus SoC CCRT in the following subgroups of the FAS (but not limited to):

- Age at randomization
    - < 65 versus ≥ 65 years of age

    - This will be determined from the date of birth (BIRTHDAT in the DEM module) and date of randomization (RND_DAT in the CRIT1 module) on the eCRF at screening. Subjects 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]). Subjects with a date of birth where day is missing will have an assumed day of 1st of the month. Subjects with a missing age value will be included using the mean age (overall FAS) and categorized accordingly

The primary interpretation of the secondary endpoint of PFS in PD-L1 subjects will be from the stratified log-rank test but a subgroup analysis of PD-L1 expression will also be performed.

- PD-L1 expression
    - ≥1% vs < 1% in either the tumor or immune component and ≥5% vs < 5% in either the tumor or immune component (determined from an investigational use only (IUO) immunohistochemistry (IHC) test result obtained retrospectively after enrolment using a validated VENTANA PD-L1 (SP263) Assay)

    - PD-L1 status expression will be assigned based on the following algorithm:

        - Only sample with sample date on or before first dose will be used

        - Most recent sample “sampled date” should be used, unless this sample is unevaluable, in which case most recent evaluable sample is used

        - For multiple evaluable samples with the same sampled date, the sample with the highest PD-L1 % staining result should be used

- Disease stage status
    - FIGO (2009) <III and node positive, FIGO (2009) Stage ≥ III and node negative, or FIGO (2009) Stage ≥ III and node positive

- PET/CT or PET/MRI at staging vs. no PET/CT at staging

- First dose of cisplatin or carboplatin

- Subjects will be assigned based on whether they have received carboplatin or cisplatin as the first chemotherapy agent

- Radiotherapy
  - Incomplete EBRT and brachytherapy vs. completed EBRT and brachytherapy
  - Did not receive brachytherapy
  - Unacceptable variations vs. no unacceptable variations
  - IMRT/VMAT vs. non-IMRT/VMAT
  - EBRT + brachytherapy administered ≤59 days vs >59 days
  - Midline block administered vs. no midline block administered
  - LN boost administered vs. no LN boost administered
  - Brachytherapy HDR administered vs. other
  - Brachytherapy given vs. no brachytherapy given
- Lymph nodes
  - Paraaortic lymph node vs. no paraaortic lymph node
  - Pelvic lymph node vs. no pelvic lymph node
  - Stage III, IV pelvic or paraaortic lymph node positive vs. stage III, IV pelvic or paraaortic lymph node negative

- Race
  - Asian versus non-Asian
    - This will be determined from the response to “Race” (DEM module) on the eCRF at screening

- Region
  - [United States, Canada, European Union, South Korea, and Japan] versus rest of the world

        - This will be determined from the center number (CENTRE)

- Japan
    - Japan versus rest of the world

- HPV status
    - HPV 16, HVP 18 versus other
        - This will be determined from the response to "Abnormal Result" (HPV module) on the eCRF at screening. For these subgroup analyses any subject with missing values will be excluded from that particular subgroup.

The subgroup analyses for the stratification factors will be based on the values entered into the IVRS, all other factors will be based on values recorded on the eCRF (or laboratory) data.

Other baseline variables may also be assessed if there is clinical or biological justification, or an imbalance is observed between the treatment arms. The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic factors. If a baseline imbalance is observed between treatment arms, ad-hoc subgroup analysis may be used to investigate any potential for impact on the main results.

No adjustment to the significance level for testing will be made since all these subgroup analyses will be considered exploratory and may only be supportive of the primary analysis of PFS. For each subgroup level of a factor, the HR (for the treatment comparisons of interest) and 95% CI will be calculated from a Cox proportional hazards model that only contains a term for treatment. The Cox models will be fit using a SAS PROC PHREG with the Efron method to control for ties, using the by statement to obtain HR and 95% CI for each subgroup level separately. These will be presented on a forest plot including the HR and 95% profile likelihood CI, along with the results of the overall primary analysis.

If there are too few events for a meaningful analysis of a particular subgroup (it is considered inappropriate to present analyses where there are less than 20 events), the relationship between that subgroup and PFS will not be formally analyzed. In this case, only descriptive summaries will be provided.

Unless there is a marked difference between the results of the statistical analyses of the PFS from the site Investigator tumor data and the BICR tumor data, the subgroup analyses will only be performed upon the PFS endpoint using the site investigator data.

Consistency of treatment effect between subgroups
The presence of quantitative interactions between treatment and stratification factors will be assessed by means of an overall global interaction test for plausible subgroups.

This is performed by comparing the fit of a Cox proportional hazards model including treatment, all covariates, and all covariate-by treatment interaction terms, with one that excludes the interaction terms, and will be assessed at the 2-sided 10% significance level. If there are not more than 10 events per stratum for any covariate (i.e., within each stratum of a treatment*covariate interaction [2 treatments * 2 levels of the covariate = 4 stratum]) a pre-defined pooling strategy should be applied to the covariate. If the pooling strategy does not meet the event criteria, then the covariate-by-treatment interaction term should be omitted from the model. Moreover, if the covariate does not have more than 10 events per level of covariate then the main effect of the covariate will also be excluded. If the fit of the model is not significantly improved, then it will be concluded that overall the treatment effect is consistent across the subgroups.

If the global interaction test is found to be statistically significant, an attempt to determine the cause and type of interaction will be made. Stepwise backwards selection will be performed on the saturated model, whereby (using a 10% level throughout) the least significant interaction terms are removed one-by-one and any newly significant interactions re-included until a final model is reached where all included interactions are significant, and all excluded interactions are non-significant. Throughout this process all main effects will be included in the model regardless of whether the corresponding interaction term is still present. This approach will identify the factors that independently alter the treatment effect and prevent identification of multiple correlated interactions.

Any quantitative interactions identified using this procedure will then be tested to rule out any qualitative interaction using the approach of Gail and Simon 1985.

Time to first and second subsequent therapy or death
The time to the start of first and second subsequent therapy will be analyzed using the same methodology and model as that used for the primary analysis of PFS. The HR for the treatment effect together with its 95% CI will be presented. In addition, medians and a Kaplan-Meier plot of the time to the start of subsequent therapy will be presented by treatment arm and the time between progression and starting subsequent therapy will be assessed based upon the date of objective disease progression (per primary definition of progression). This will be summarized per treatment arm, but no formal comparisons will be made. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

In subjects who received a subsequent cancer therapy, a summary table of first subsequent cancer therapies by treatment arm will be provided.

A summary of the number of subjects prematurely censored will also be produced.

4.3.3 Secondary efficacy endpoints

4.3.3.1 Progression-free survival at 3 years (PFS (3 years))

PFS (3 years) will be summarized (using the Kaplan-Meier curve) and presented by treatment arm together with its corresponding 95% CI with no formal comparison or p-value attached.

4.3.3.2 Progression-free survival in PD-L1 positive high subjects

PFS in PD-L1 high subjects will be analyzed using a stratified log-rank test, using the same methodology as for the primary PFS endpoint. The effect of durvalumab + SoC CCRT versus placebo + SoC CCRT will be estimated by the HR together with its corresponding CI and p-value using the same methodology as described for PFS in the subset of subjects with PD-L1 expression of ≥1%. Kaplan-Meier plots will be presented by treatment arm. Summaries of the number and percentage of subjects experiencing a PFS event and type of event (local progression, distant progression, or death) will be provided along with median PFS for each treatment.

PFS in PD-L1 low subjects will be summarized and analyzed in a similar manner to PFS in PD-L1 high subjects.

4.3.3.3 Overall survival (OS)

OS will be analyzed using the same methodology and model as that used for the primary analysis of PFS. The effect of treatment will be estimated by the HR together with its corresponding 95% CI. 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.

Sensitivity analyses

A sensitivity analysis for OS will examine the censoring patterns to rule out attrition bias with regard to the treatment comparisons, achieved by a Kaplan-Meier plot of time to censoring where the censoring indicator of OS is reversed.

The number of subjects prematurely censored will be summarized by treatment arm. A subject would be defined as prematurely censored if their survival status was not defined at the DCO.

In addition, duration of follow-up will be summarized using medians:

- In all subjects: Time from randomization to the date of death (i.e. overall survival) or to the date of censoring (date last known to be alive) for censored subjects regardless of treatment arm

- In censored subjects who are alive at DCO only: Time from randomization to date of censoring (date last known to be alive) for each arm

4.3.3.4 Overall survival in PD-L1 high subjects

Overall survival in PD-L1 high subjects will be analyzed using a stratified log-rank test, using the same methodology as described by the primary PFS endpoint. The effect of durvalumab + SoC CCRT versus placebo + SoC CCRT will be estimated by the HR together with its corresponding CI and p-value using the same methodology as described for the PFS. Kaplan-Meier plots will be presented by treatment arm. Summaries of the number of 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 overall survival for each treatment.

Overall survival in PD-L1 low subjects will be summarized and analyzed in a similar manner to overall survival in PD-L1 high subjects.

These analyses will be performed at the same time as the final analysis of OS.

4.3.3.5 Objective response rate (ORR)

The ORR will be based on the site investigator RECIST 1.1 data as well as physical exam and confirmatory biopsy data and using all scans and physical exams regardless of whether they were scheduled or not. The ORR will be compared between durvalumab plus SoC CCRT versus placebo plus SoC CCRT using logistic regression models, adjusting for the same stratification factors as the primary endpoint as covariates in the model. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favor durvalumab) together with its associated profile likelihood 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 logistic regression analysis will be repeated for confirmed ORR based on investigator RECIST 1.1 data as well as physical exam and confirmatory biopsy data, but this will not be considered the primary interpretation of the secondary endpoint of ORR.

If there are not enough responses for a meaningful analysis using logistic regression, then a Cochran-Mantel Haenszel (CMH) test will be presented. The CMH test will be stratified using

the same stratification factors as the primary endpoint. The results of the analysis will be presented in terms of an odds ratio together with the 95% CI and p-value. The odds ratio, 95% CI and p-value will be obtained using SAS PROC FREQ and the CMH test option. The STRATUM variable used in the TABLE statement will be based on region and disease stage status.

Summaries will be produced that present the number and percentage of subjects with a tumor response (CR/PR) based upon the number of subjects with measurable disease at baseline per the site investigator.

Overall response data will be listed for all subjects (i.e., the FAS).

4.3.3.6 Complete response (CR) rate

The CR Rate will be based on the site investigator RECIST 1.1 data as well as physical exam and confirmatory biopsy data. The CR rate will be compared between durvalumab plus SoC CCRT versus placebo plus SoC CCRT using logistic regression models, using the same methodology as described for the ORR endpoint in Section 4.3.3.4. If there are not enough responses for a meaningful analysis using logistic regression, then a Cochran-Mantel Haenszel (CMH) test will be presented using the same methodology described for the ORR endpoint.

Summaries will be produced that present the number and percentage of subjects with a tumor response of CR.

4.3.3.7 Duration of response (DoR) for complete responders

Descriptive statistics will be provided for the duration of response in subjects with a complete response (CR) observed including the associated Kaplan-Meier curves (without any formal comparison or p-value attached).

The summaries described above will also be produced for duration of response for subjects with a CR or PR.

4.3.3.8 Incidence of local progression, distant disease progression, secondary malignancy

The actual number and percentage of subjects with a progression-free survival event will be presented by treatment arm for the FAS according to whether the first progression-free survival event was a local progression, distant progression, secondary malignancy, or death.

As mentioned in Section 3.2.77, distant disease progressions and secondary malignancies will be summarized together for time to event endpoints.

The cumulative incidence of local progressions and distant disease progressions/secondary malignancies will be compared between treatment groups using competing risk analysis (Gray test, Fine and Gray 1999). Death will be considered a competing risk for local progression and distant disease progression or secondary malignancy. In addition, distant disease progression and secondary malignancy will be considered to be competing risks for local progression and local progression will be considered to be a competing risk for distant disease progression and secondary malignancy. Hazard ratios and 95% confidence intervals from competing risks regression analyses (Fine and Gray, 1999) adjusted for randomized treatment group and the stratification factors will also be presented.

The Gray test and Fine and Gray models can be implemented in SAS version 9.4 by specifying the event code option in PROC PHREG.

- The data derivation requirements are similar to those of a traditional survival analysis, with "survival time" calculated as the time from randomization to the occurrence of an event (or, here one or more possible events) or censored at the time of analysis if the event has not yet occurred. The key difference is a status variable which labels distinct codes for different causes of failure and a distinct code for censored observations

- Cumulative incidence estimates will be plotted by treatment arm, for both the cause of interest and the competing risk(s). Because of the limitations of available computational tools, the effect of treatment arm will be assessed by a Wald Test

- Proportional hazards model for the sub distribution of the competing risk, Fine and Gray models, will include the treatment arm and stratification factors

- As with the proportional hazards model, tests will be performed to assess whether the Fine and Gray model is actually a reasonable fit to the data

4.3.3.9 Best objective response (BoR)

For each treatment arm, best objective response (BoR) will be summarized by n (%) for each category (CR, PR, SD, PD and NE). No formal statistical analyses are planned for BoR.

4.3.3.10 Patient reported outcomes

All patient reported outcomes will be presented using the FAS, unless otherwise stated.

The PRO endpoints identified as primary are symptom experience as assessed by the EORTC -QLQ-CX24 and global health status/QoL, physical functioning, and role functioning as assessed by the EORTC-QLQ-C30.

The analysis will be performed using a linear mixed model for repeated measures (MMRM) analysis of change from baseline in the scores for each assessment timepoint (see next section for further details)

Compliance rates summarizing questionnaire completion at each visit will be tabulated.

Mixed models repeated measures (MMRM analysis)
Change from baseline in the PRO symptom scores of symptom experience as assessed by the EORTC-QLQ-CX24 and global health status/QoL, physical functioning, and role functioning as assessed by the EORTC-QLQ-C30 will be analyzed using a mixed model for repeated measures (MMRM) analysis of all post-baseline scores for each visit . Adjusted mean change from baseline estimates per treatment group and corresponding 95% CIs will be presented along with an overall estimate of the treatment difference, 95% CI, and p-value. The analysis will be to compare the average treatment effect from the point of randomization until 24 months or where 20 or more subjects on each treatment arm have a score. Note that the default method will be used to weight the coefficients in the least square means calculations.

It is acknowledged that subjects will discontinue treatment at different timepoints during the study and that this is an important time with regards to symptoms and HRQoL data collection. To account for this, and to include the discontinuation and follow up visits, a generic visit variable will be derived for each subject in order that the average treatment effect can be analyzed using the above method. Each visit will be assigned a sequential number. The time from randomization to each of these will be derived to select only those visits occurring within the first X months of randomization.

As an example, say a subject X attends the first 4 scheduled visits of a 2-weekly schedule and then discontinues treatment, whilst subject Y discontinues treatment after the first scheduled visit, the first 6 generic visits would be as follows:

Table 16: Discontinuation example

Generic Visit	Study Day	
	Subject X	Subject Y
Baseline	Baseline	Baseline
1	29	28
2	57	50 (discontinuation)

3	85	85
4	113	113
5	130 (discontinuation)	141
6	169	169

The MMRM model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, disease stage status (FIGO (2009) Stage <III and node positive, FIGO (2009) Stage ≥III and node negative, or FIGO (2009) Stage ≥III and node positive) and region ([United States, Canada, European Union, South Korea, and Japan] versus rest of world) as well as the continuous fixed covariate of baseline score and the baseline score-by-visit interaction. Restricted maximum likelihood (REML) estimation will be used. An overall adjusted mean estimate will be derived that will estimate the average treatment effect over visits giving each equal weight. For this overall treatment comparison, adjusted mean estimates per treatment group and corresponding 95% CIs will be presented along with an estimate of the treatment difference, 95% CI (for the 4 key symptoms only) and p-value.

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, autoregressive and compound symmetry.

Multiple imputation techniques for missing values may be considered to explore the robustness of any treatment effect.

Change from baseline
For both the EORTC-QLQ-CX24 and EORTC-QLQ-C30, summaries of absolute and unadjusted change from baseline values for each scale/item will be reported for each treatment arm. Summaries of the number and percentage of subjects in each response category at each assessment time period for each ordinal item (in terms of the proportion of subjects in the categories of improvement, stable, and deterioration as defined in Tables 10 and 11) will also be produced for each treatment arm.

PRO-CTCAE
Data from the PRO-CTCAE will be summarized using the FAS. The number and percentage of subjects with each level of response for each CTCAE item at baseline and over time may be summarized. A bar chart of the incidence by visit will be presented for each CTCAE. Further

summaries to explore the data (i.e., the severity of symptoms) may be produced if needed. Visits will be presented provided at least one treatment arm has >=20 subjects with data at a given visit.

EQ-5D-5L

The evaluable population will comprise a subset of the FAS who have a baseline EQ-5D-5L assessment.

Descriptive statistics will be calculated for each scheduled visit/time point in the study, for each trial arm. These will report the number of subjects, 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.

Summary statistics will be supported by plots of mean and mean change from baseline in EQ-5D index score and VAS score, and associated 95% CIs, by time/scheduled visit and treatment arm. 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.

PGIS and PGIC

PGIC and PGIS summaries will include the number and percentage of subjects with each response at each visit.

4.3.4 Exploratory endpoints

4.3.4.1 Health care resource use

The potential impact the disease and treatment have on health care resource use will be analyzed for the purposes of submissions to payers. Descriptive statistics (as appropriate, including means, median, ranges or frequencies and percentages) will be provided for each arm on the different types of hospital admissions, the length of stay of people admitted in to hospital for at least one overnight stay and length of stay of people admitted to intensive care / high dependency units, as well as the primary sign or symptom the subject presents with. To support submissions to payers, additional analyses may be undertaken, and these will be outlined in a separate Payer Analysis Plan.

4.3.5 Safety

Safety and tolerability data from all cycles will be summarized. No formal statistical analyses are planned for safety data. All safety data will be presented using the SAF.

The following sections describe the planned safety summaries for AEs, laboratory measurements, vital signs, physical examinations, ECG and WHO/ECOG performance status.

Note that additional safety tables (not specified in this SAP) may need to be produced to aid interpretation of safety data.

4.3.5.1 Adverse events

All AEs, both in terms of MedDRA preferred term and CTCAE grade, will be listed and summarized descriptively by count (n) and percentage (%). Any AE occurring before treatment with IP (as defined in Table 13) and which did not worsen during the course of the study will be included in all AE listings but will not be included in the summary tables (unless otherwise stated). These shall be referred to as 'pre-treatment'. However, any AE occurring before the administration of IP that worsens after Study Day 1 will be regarded as treatment emergent and this will be included in the summary tables. Note: If an AE is not worse than baseline (pre-dose) severity then it will not be classified as TEAE.

AEs observed up until 90 days following last dose of study treatment or until the initiation of the first subsequent anti-cancer therapy following discontinuation of study treatment (whichever occurs first) will be used for the reporting of the AE summary tables. This will more accurately depict AEs attributable to study treatment only as opposed to presenting all AEs reported up to 90 days following discontinuation of IP. This is due to the fact that a number of AEs up to 90 days following discontinuation are likely to be attributable to subsequent anti-cancer therapy. However, to assess the longer-term toxicity profile, limited AE summaries will also be produced containing AEs observed up until 90 days following discontinuation of durvalumab plus SoC CCRT or placebo plus SoC CCRT (i.e., without taking subsequent anti-cancer therapy into account). Any events in this period that occur after a subject has received further therapy for cancer (following discontinuation of study treatment) will be flagged in the data listings.

A separate listing of AEs occurring more than 90 days after discontinuation of study treatment will be produced. These events will not be included in AE summaries.

All reported AEs will be listed along with the date of onset (including study day), date of resolution (if AE is resolved), investigator's assessment of CTCAE grade and relationship to study drug. Frequencies and percentages of subjects reporting each preferred term will be presented (i.e., multiple events per subject will not be accounted for apart from any episode level summaries which may be produced).

Summary information (the number and percentage of subjects by treatment arm) by system organ class and preferred term will be tabulated for:

- All AEs

- All AEs possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- All AEs possibly related to SoC CCRT only (as determined by the reporting investigator)
- All AEs possibly related to durvalumab/placebo or SoC CCRT (as determined by the reporting investigator)
- All AEs possibly related to durvalumab/placebo only (as determined by the reporting investigator) by system organ class, preferred term and maximum CTCAE grade
- All AEs possibly related to SoC CCRT only (as determined by the reporting investigator) by system organ class, preferred term and maximum CTCAE grade
- All AEs possibly related to durvalumab/placebo or SoC CCRT (as determined by the reporting investigator) by system organ class, preferred term and maximum CTCAE grade
- Most common AEs (frequency of ≥ 5%)
- AEs with CTCAE grade 3 or 4
- AEs with CTCAE grade 3 or 4, possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- AEs with CTCAE grade 3 or 4, possibly related to SoC CCRT only (as determined by the reporting investigator)
- AEs with CTCAE grade 3 or 4, possibly related to durvalumab/placebo or SoC CCRT (as determined by the reporting investigator)
- Most common AEs with CTCAE grade 3 or 4 (frequency of ≥ 1%)
- AEs with outcome of death
- AEs with outcome of death possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- AEs with outcome of death possibly related to SoC CCRT only (as determined by the reporting investigator)

- AEs with outcome of death possibly related to durvalumab/placebo or SoC CCRT (as determined by the reporting investigator)
- All SAEs
- All SAEs possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- All SAEs possibly related to SoC CCRT only (as determined by the reporting investigator)
- All SAEs possibly related to durvalumab/placebo or SoC CCRT (as determined by the reporting investigator)
- AEs leading to study drug dose delays of durvalumab/placebo only
- AEs leading to study drug dose delays of SoC CCRT only
- AEs leading to study drug dose delays of durvalumab/placebo or SoC CCRT
- AEs leading to discontinuation of durvalumab/placebo only
- AEs leading to discontinuation of SoC CCRT only
- AEs leading to discontinuation of durvalumab/placebo or SoC CCRT
- AEs leading to discontinuation of durvalumab/placebo only, possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- AEs leading to discontinuation of SoC CCRT only, possibly related to SoC CCRT only (as determined by the reporting investigator)
- AEs leading to discontinuation of durvalumab/placebo or SoC CCRT, possibly related to durvalumab/placebo or SoC CCRT (as determined by the reporting investigator)
- AEs leading to dose interruption of durvalumab/placebo only
- AEs leading to dose interruption of SoC CCRT only
- AEs leading to dose interruption of durvalumab/placebo or SoC CCRT
- SAEs leading to discontinuation of durvalumab/placebo only

- SAEs leading to discontinuation of durvalumab/placebo, possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- SAEs leading to discontinuation of SoC CCRT
- SAEs leading to discontinuation of durvalumab/placebo or SoC CCRT
- SAEs leading to dose interruption of durvalumab/placebo only
- SAEs leading to dose interruption of durvalumab/placebo only, possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- SAEs leading to dose interruption of SoC CCRT only
- SAEs leading to dose interruptions of durvalumab/placebo or SoC CCRT
- Other significant AEs
- Other significant AEs, possibly related to durvalumab/placebo only (as determined by the reporting investigator)
- Other significant AEs, possibly related to SoC CCRT only (as determined by the reporting investigator)
- Other significant AEs, possibly related to durvalumab/placebo or SoC CCRT (as determined by the reporting investigator)
- Immune mediated AEs (as determined by the reporting investigator)
- Infusion reaction AEs (as determined by the reporting investigator)

An overall summary of the number and percentage of subjects in each category will be presented. For the truncated AE tables of most common AEs, all events that occur in at least 5% of subjects overall will be summarized by preferred term, by decreasing frequency. This cut-off may be modified after review of the data. When applying a cut-off (e.g., 5%), the raw percentage should be compared to the cut-off, no rounding should be applied first (i.e., an AE with frequency 4.9% will not appear if the cut-off is 5%).

Each AE event rate (per 100 subject years) will also be summarized by preferred term within each system order class for the output summarizing all AEs. For each preferred term, the event rate is defined as the number of subjects with that AE divided by [the total treatment duration (days) of randomized treatment + number of days in the safety follow-up period following

discontinuation of study treatment or up until the start of subsequent therapy, whichever comes first] summed over subjects and then multiplied by 365.25 x 100 to present in terms of per 100 subject years.

AEs will be assigned CTCAE grades and summaries of the number and percentage of subjects will be provided by maximum reported CTCAE grade, system organ class and preferred term.

For each treatment arm, time to first adverse event (days) will be tabulated by system organ class and preferred term. For each AE, time to first onset of the AE from date of first dose will be presented in the listing.

Deaths
Two summaries of all deaths will be provided with number and percentage of subjects by treatment group categorized as:

- Total number of deaths (regardless of date of death)
- Death related to disease under investigation only as determined by investigator
- Death related to disease under investigation and an AE with an outcome of death
    a. AE onset prior to subsequent therapy. Which includes AEs with onset date (or pre-treatment AEs that increase in severity) on or after the date of first dose and up to and including 90 days following the last dose of study medication (durvalumab/placebo) or AE start date ≤ the date of initiation of the first subsequent therapy (whichever occurs first)
    b. AE onset after start of subsequent therapy. Which includes AEs with start date >90 days following the last dose of study medication (durvalumab/placebo) and/or AE start date > the date of initiation of the first subsequent therapy (whichever occurs first)
- AE with outcome of death only
    a. AE onset prior to subsequent therapy. Which includes AEs with an onset date (or pre-treatment AEs that increase in severity) on or after the date of first dose and up to and including 90 days following the last dose of study medication (durvalumab/placebo), or AE start date ≤ the date of initiation of the first subsequent therapy (whichever occurs first)

b. AE onset after the start of subsequent therapy. Which includes AEs with a start date >90 days following the last dose of study medication (durvalumab/placebo) and/or AE start date > date of initiation of the first subsequent therapy (whichever occurs first)

- Death after end of safety follow-up period (last dose of study medication (durvalumab/placebo + 90 days) and not due to disease under investigation
- Unknown reason for death
- Other deaths

This summary will be repeated for all deaths on treatment or within 90 days of the last dose of durvalumab/placebo.

4.3.5.2 Adverse events of special interest (AESI) and possible interest (AEPI)

Preferred terms used to identify AESIs/AEPIs will be listed before DBL and documented in the Trial Master File. Grouped summary tables for certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI/AEPI grouping. For each 'grouped' term, the number (%) of subjects experiencing any of the specified terms will be presented.

Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

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

- Any AESI/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 durvalumab/placebo only
- Any AESI/AEPI possibly related to SoC CCRT only
- Any AESI/AEPI possibly related to durvalumab/placebo or SoC CCRT
- 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 durvalumab/placebo only
- At least one AESI/AEPI leading to discontinuation of SoC CCRT only
- At least one AESI/AEPI leading to discontinuation of durvalumab/placebo or SoC CCRT

An overall AESI/AEPI summary will be presented, including number and percentage of subjects in each of these categories.

Immune-mediated Adverse events (imAEs)

The imAEs (as classified by the Sponsor) will be summarized in the similar manner as for the summaries for AESI/AEPI described above.

4.3.5.3 Summary of long-term tolerability

To assess the long-term tolerability, if there are sufficient subjects with events to warrant it, prevalence plots, life plots and cumulative incidence plots may 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, that is 10 events per treatment group.

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

A life table can be used to describe the time to onset (date of onset – start date of treatment + 1) of the event and specifically when subjects are most at risk of first experiencing the event. The hazard, or in other words the probability of having and AE in a specific time-period (e.g., 0 – 1 months, 1 – 3 months, 3 – 6 months, etc.) given that the subject reaches the time-period

without having an event is plotted for each time-period. These plots may only be produced for AESIs that have ≥10 events, that is 10 events per treatment group.

A cumulative incidence plot is a plot of the raw cumulative incidence and cumulative incidence function over time, these may be presented on separate plots. The raw cumulative incidence is the actual probability that a subject will have experienced their first occurrence of the event at a given time-point. These plots may only be produced for AESIs that have ≥10 events, that is 10 events per treatment group.

4.3.5.4 Exposure

Exposure will be summarized for the SAF. Note all summaries related to carboplatin exposure apply only to subjects initially treated with carboplatin prior to CSP version 3.0. The following summaries will be produced:

- Duration of exposure to durvalumab/placebo (number of infusions, number of infusions group [1, 2 - <6, 6 – <12, 12 - <18, 18 - <24, >=24), total treatment duration (days), actual treatment duration (days), RDI)
- Duration of exposure to radiotherapy (total treatment duration for EBRT (days), actual treatment duration for EBRT (days), total treatment duration for EBRT + brachytherapy (days), actual treatment duration for EBRT + brachytherapy (days), total dose of radiotherapy to the pelvis, biologically equivalent dose, Subjects with boost [lymph node, parametrial and paraaortic], Subjects with any unacceptable variations, Subjects with any unacceptable variations requiring intervention)
- Number of doses received (carboplatin or cisplatin only)
- Summary of dose delays and infusion interruptions for durvalumab/placebo
- Summary of dose delays, reductions and interruptions (infusion interruptions and skipped doses) of chemotherapy (cisplatin or carboplatin)
- Summary of dose interruptions for radiotherapy (EBRT and/or brachytherapy)
- Exposure over time to any protocol mandated study treatment

4.3.5.5 Laboratory measurements

Laboratory data obtained until 90 days after the last dose of study treatment or until the initiation of the first subsequent anti-cancer therapy following discontinuation of study treatment (whichever occurs first) will be used for reporting. This will more accurately depict laboratory toxicities attributable to study treatment only as a number of toxicities up to 90

days following discontinuation of study treatment are likely to be attributable to subsequent anti-cancer therapy.

Data summaries and listings will be provided by AZ preferred units.

All laboratory data will be listed. Flags will be applied to values falling outside – reference ranges (which will be explicitly noted on these listings where applicable), and to values for which CTCAE grading applies.

Scatter plots (shift plots) of baseline to maximum/minimum values (as appropriate) on treatment (i.e., on treatment is defined as data collected between the start of treatment and the relevant follow-up period following the last dose of study treatment) may be produced for certain parameters if warranted after data review.

Boxplots of absolute values by week, and boxplots of change from baseline by week, may be presented for certain parameters if warranted after data review.

Shift tables of laboratory values by worse common toxicity criteria (CTCAE) grade will be produced, and for specific parameters separate shift tables indicating hyper- and hypo-directionality of change will be produced. The laboratory parameters for which CTCAE grade shift outputs will be produced are:

- Hematology: Hemoglobin, Leukocytes, Lymphocytes (absolute count), Neutrophils (absolute count), Platelets
- Clinical Chemistry: ALT, AST, Albumin, Alkaline Phosphatase, Total bilirubin, Magnesium (hypo- and hyper-), Sodium (hypo- and hyper-), Potassium (hypo- and hyper-), Corrected Calcium (hypo- and hyper-), Glucose (hypo-), Gamma-glutamyl transferase, Creatinine

For parameters with no CTCAE grading that are listed in the CSP, shift tables from baseline to worse value on treatment will be provided. Additional summaries will include a shift table of urinalysis (Bilirubin, Blood, Glucose, Ketones, Protein) comparing baseline value to maximum on treatment value.

The denominator used in laboratory summaries of CTCAE grades will only include evaluable subjects. If a CTCAE criterion involved a change from baseline, evaluable subjects are those who have both a pre-dose and at least 1 post-dose value recorded. If a CTCAE criterion does not consider changes from baseline, evaluable subjects are those who have at least 1 post-dose value recorded.

Hy's law
The following summaries will include the number (%) of subjects who have

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

- ALT ≥3x – ≤5x, >5x – ≤8x, >8x – ≤10x, >10x – ≤20x, and >20x ULN during the study
- AST ≥3x – ≤5x, >5x – ≤8x, >8x – ≤10x, >10x – ≤20x, and >20x ULN during the study
- Total bilirubin ≥2x – ≤3x, >3x – ≤5x, >5x ULN during the study
- ALT or AST ≥3x – ≤5x, >5x – ≤8x, >8x – ≤10x, >10x – ≤20x, >20x ULN during the study
- ALT or AST ≥3x ULN and total bilirubin ≥2x ULN during the study (potential Hy's law): the onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation
- Narratives will be provided in the CSR for subjects who have ALT ≥ 3x ULN plus total bilirubin ≥ 2x ULN or AST ≥ 3x ULN plus total bilirubin ≥ 2x ULN at any visit.

Liver biochemistry test results over time for subjects with elevated ALT (i.e. ≥ 3x ULN) or AST (i.e. ≥ 3x ULN), and elevated total bilirubin (i.e. ≥ 2x ULN) (at any time) will be plotted. Individual subject data where ALT or AST plus total bilirubin are elevated at any time will be listed also.

Plots of maximum post-baseline ALT and AST vs. maximum post-baseline total bilirubin, expressed as multiples of ULN, will also be produced with reference lines at 3×ULN for ALT and AST, and 2×ULN for Total bilirubin. In each plot, total bilirubin will be in the vertical axis.

Abnormal Thyroid function
Elevated thyroid stimulating hormone (TSH) will be summarized per treatment group in terms of number (%) of subjects with elevated TSH (higher than the upper normal range), low TSH (lower than lower normal range), elevated TSH post-dose and within normal range at baseline, low TSH post-dose and within normal range at baseline. Shift tables showing baseline to maximum and baseline to minimum will be produced.

4.3.5.6 Electrocardiograms

Summaries of ECG data will include all data obtained up until 30 days after the last dose of study treatment. These data will be mapped in accordance with Section 3.5.1.2. Absolute values and change from baseline for ECG heart rate, PR duration, QRS duration, QT duration, and RR duration will be presented.

The number and percentage of subjects with normal and abnormal (not clinically significant and clinically significant) ECG results will be presented as a shift table from baseline to worst evaluation during the study.

A summary of QTcF intervals at any observation on treatment in each of the treatment arms will be presented.

Where ECGs are recorded in triplicate at a visit, an average of each measurement will be taken in order to obtain a single value to be used in the analysis/summaries. Where an average cannot be taken for the categorical variable of overall evaluation, the worst evaluation will be reported for a visit.

4.3.5.7 Vital signs

Summaries for vital signs data will include all data obtained up until 30 days after the last dose of study treatment. These data will be data mapped in accordance with Section 3.5.1.2. Absolute values and change from baseline for diastolic and systolic BP, pulse, respiratory rate and temperature will be summarized at each scheduled measurement. The denominator in vital sign data should include only those subjects with recorded data.

Boxplots will be produced to show the absolute values and change from baseline at each scheduled measurement.

4.3.5.8 Eastern Cooperative Oncology Group performance status

Summaries of ECOG data will include all data mapped in accordance with Section 3.5.1.2. Absolute values and change from baseline for ECOG PS will be summarized at each visit.

4.3.6 Pharmacokinetic data

PK concentration data for durvalumab will be summarized for all subjects in the PK analysis set.

Serum concentrations of durvalumab will be summarized by nominal sample time using standard summary statistics for PK concentrations (geometric mean, geometric coefficient of variation, geometric mean ± standard deviation, arithmetic mean, standard deviation, minimum, maximum and n). All serum concentrations will be listed.

If the data are suitable, the relationship between PK exposure and efficacy/safety parameters may be investigated graphically or using an appropriate data modelling approach.

4.3.7 Immunogenicity analysis

Immunogenicity results will be listed for all subjects regardless of ADA-evaluable status. The number and percentage of subjects who develop detectable ADA to durvalumab within each ADA response category listed in Section 3.7 will be summarized based on the SAF with a non-missing baseline anti-drug antibody (ADA) result and at least one non-missing post-baseline ADA result. ADA titer and neutralizing ADA data will be listed for samples confirmed positive for the presence of anti-durvalumab antibodies. AEs in ADA-positive

subjects by ADA-positive category will be listed. Details for the presentation and derivation of ADA data is provided in Section 3.7.

The effect of immunogenicity on PK, efficacy and safety will be evaluated if data allow.

4.3.8 Demographic, initial diagnosis and baseline characteristics data

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

- Subject disposition
- Important protocol deviations
- Inclusion in analysis sets
- Demographics (age, age group [<50, ≥ 50 - < 65, ≥ 65 - < 75 and ≥ 75 years], race (White, Black or African American, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska Native, Other) and ethnicity (Hispanic or Latino, Not Hispanic or Latino), Country)
- Subject characteristics at baseline (height, weight, weight group [<70kg, >=70kg - <=90kg, >90kg], body mass index, body mass index group [<18.5, >=18.5 - <25.0, >=25.0 - <30, >=30])
- Subject recruitment by region, country and center
- Stratification factors recorded (at randomization by IWRS, on eCRF and agreement between IWRS versus eCRF)
- Medical history
- Disease characteristics at baseline (WHO/ECOG performance status, primary tumor location, histology type, regional lymph nodes, tumor grade, FIGO (2009) stage and overall disease classification)
- Extent of disease at study entry
- Primary tumor location at initial diagnosis (locally advanced – primary tumor [by FIGO 2009 stage], locally advanced – regional lymph nodes [N0, N1], nodal involvement [pelvic or paraaortic lymph nodes], pelvic lymph node involvement [by FIGO 2009 stage], paraaortic lymph node involvement [by FIGO 2009 stage])

- Medical history (past and current)
- Surgical history
- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer therapy
- Nicotine use, categorized (never, current, former)

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

4.3.9 Concomitant and other treatments

All concomitant and other treatment data will be listed for all subjects in the FAS.

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 FAS by ATC classification codes. Subjects with the same concomitant medication/procedure multiple times will be counted once per medication/procedure. A medication/procedure that can be classified into more than one chemical and/or therapeutic subgroup will be presented in each subgroup.

4.3.10 Subsequent therapy

Subsequent therapies received after discontinuation of study treatment will be summarized by treatment group, together with the number of regimens received.

4.3.11 Biomarker data

Summaries and analyses for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the CSR separately.

4.3.12 Genetic data

CCI [redacted]

4.3.13 Radiotherapy Quality Control Assessment

Radiotherapy quality is monitored through an external vendor with deviations to the protocol specified radiotherapy marked as variations unacceptable if the dose or contour is incorrect or missing or total treatment duration exceeds 59 days. Radiotherapy doses outside of the range set by the protocol are listed as deviation unacceptable. Listings and summary tables of unacceptable variations and deviations unacceptable will be developed. Clinical significance will be assigned following a review from the CALLA radiotherapy steering committee of all applicable cases and will be summarized in the CSR. Clinical significance is defined on whether a dosing deviation results in substantial risk in toxicity and suboptimal efficacy.

4.4 COVID-19

Depending on the extent of any impact, summaries 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 treatment, and other protocol deviations) may be generated, including:

- Disposition (discontinued study treatment due to COVID-19 and withdrew study due to COVID-19)
- Deviations (overall deviations plus if due to COVID-19 and not due to COVID-19)
- Summary of COVID-19 disruptions (visit impact, drug impacted)
- Listing for subjects affected by the COVID-19 pandemic
- Listing for subjects with reported issues in the Clinical Trial Management System due to the COVID-19 pandemic

If there are sufficient number of subjects with confirmed or suspected COVID-19 infections, the following tables will also be produced:

- Number of subjects with adverse events in any category - subject level, patients with confirmed / suspected COVID-19 infection
- Number of subjects with adverse events in any category - subject level, patients without confirmed / suspected COVID-19 infection
- Number of subjects with adverse events by system organ class, preferred term and maximum reported CTCAE grade, patients with confirmed / suspected COVID-19 infection

- Number of subjects with adverse events by system organ class, preferred term and maximum reported CTCAE grade, patients without confirmed / suspected COVID-19 infection

- Number of subjects with adverse events, associated with COVID-19 by system organ class and preferred term

- Number of subjects with adverse events, excluding AEs associated with COVID-19 by system organ class and preferred term

- Number of subjects with confirmed / suspected COVID-19 adverse events, by system organ class and preferred term

- Number of subjects with adverse events, excluding confirmed / suspected COVID-19 AEs by system organ class and preferred term

- Number of subjects with adverse events associated with COVID-19 with outcome of death, by system organ class and preferred term

- Number of subjects with adverse events with outcome of death, excluding AEs associated with COVID-19) by system organ class and preferred term

- Number of subjects with adverse events associated with COVID-19 leading to discontinuation of investigational product, by system organ class and preferred term

- Number of subjects with adverse events leading to discontinuation of investigational product, excluding AEs associated with COVID-19, by system organ class and preferred term

4.4.1 Censoring Confirmed/Suspected COVID-19 Deaths for PFS Analysis

A sensitivity analysis will be conducted to assess for the potential impact of COVID deaths on PFS. This will be assessed by repeating the PFS analysis except that any subject who had a PFS event due to death where primary/secondary cause of death was due to COVID-19 Infection, or a COVID-19 infection reported as a fatal AE, will be censored at their last evaluable assessment prior to their COVID infection death date.

4.4.2 Censoring Confirmed/Suspected COVID-19 Deaths for OS Analysis

A sensitivity analysis will be conducted to assess for the potential impact of COVID deaths on OS. This will be assessed by repeating the OS analysis except that any subject who had a

death with primary/secondary cause as COVID-19 Infection, or a COVID-19 infection reported as a fatal AE will be censored at their COVID infection death date.

5 INTERIM ANALYSES

5.1 PFS and OS interim analyses

No PFS interim analyses are planned in this study.

There will be an interim analysis performed for OS. The first interim analysis for OS will occur at the time of the PFS analysis, if PFS is statistically significant. It is anticipated that approximately 86% of the OS events will be available for this OS IA (~195 of 227 OS events). If exactly 86% of the target events are available at the time of the interim analysis, with overall 2-sided alpha level of 5%, the 2-sided alpha to be applied to the interim and final analysis would be 3.1% and 4.1%, respectively.

If the criteria for statistical significance is met for PFS but not OS at the time of the PFS analysis, follow-up will continue until the earlier of 227 OS events or 71 months following randomization of the first subject, at which point OS will be retested. If the criteria for statistical significance for OS is met at the interim analysis, OS will not be retested, but follow-up will continue until the earlier of 227 OS events or 71 months following randomization of the first subject.

5.2 Independent data monitoring committee

This study will use an external independent data monitoring committee (IDMC) to assess ongoing safety analyses as well as the interim efficacy analysis. The committee will first meet when the first 25 subjects across both treatment arms have completed SoC CCRT and have had at least 28 days of follow-up. The second safety review will take place when the first 60 subjects across both treatment arms have completed SoC CCRT and have had at least 28 days of follow-up. An additional safety review for Japanese subjects will take place when the first 9 subjects in Japan have completed SoC CCRT and had 28 days of follow-up. Safety reviews will be carried out by the IDMC in an unblinded manner. The IDMC will meet at least every 6 months thereafter. Following each meeting, the IDMC will report to the sponsor and may recommend changes in the conduct of the study.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AstraZeneca and are free from conflict of interest.

Following the reviews, the IDMC will recommend whether the study should continue unchanged, be stopped, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca. The report will include the

recommendation and any potential protocol amendments and will not contain any unblinding information.

The final decision to modify or stop the study will sit with the sponsor. The sponsor or IDMC may call additional meetings if at any time there is concern about the safety of the study.

Full details of the IDMC procedures and processes can be found in the IDMC Charter. The safety of all AstraZeneca/MedImmune clinical studies is closely monitored on an ongoing basis by AstraZeneca/MedImmune representatives in consultation with the Patient Safety Department. Issues identified will be addressed; this could involve, for instance, amendments to the Clinical Study Protocol and letters to investigators.

6 CHANGES OF ANALYSIS FROM PROTOCOL (NOT APPLICABLE)

7 REFERENCES

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

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

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

Burman et al 2009
Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Statistics in Medicine 2009; 28:739-61

Fayers et al 1999
Fayers P, Aaronson NK, Bjordal K, Curran D, Groenvold M on behalf of the EORTC Quality of Life Study Group. EORTC QLQ-C30 Scoring Manual: 2nd Edition 1999. Available on request

Fine, J. P., and Gray, R. J. 1999
A Proportional Hazards Model for the Subdistribution of a Competing Risk." Journal of the American Statistical Association 94:496–509.

Lan and DeMets 1983
Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika

1983; 70:659-63

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

Rose et al 1999
Rose PG, Bundy BN, Watkins EB, Thigpen JT, Deppe G, Maiman MA, et al. Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. N Engl J Med 1999;340(15):1144-53

Lin et al 2020
Lin RS, Lin J, Roychoudhury S, Anderson KM, Hu T, Huang B, et al. Alternative Analysis Methods for Time to Event Endpoints Under Nonproportional Hazards: A Comparative Analysis. Statistics in Biopharmaceutical Research 2020;12(2):187-98

8 APPENDIX (NOT APPLICABLE)