Statistical Analysis Plan	
Study Code	D081RC00001; ENGOT-ov46; AGO-OVAR 23; GOG-3025.
Edition Number	2.0
Date	02 September 2021

A Phase III Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab in Combination with Chemotherapy and Bevacizumab, Followed by Maintenance Durvalumab, Bevacizumab and Olaparib in Newly Diagnosed Advanced Ovarian Cancer Patients (DUO-O)

A Phase III Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab in Combination with Chemotherapy and Bevacizumab, Followed by Maintenance Durvalumab, Bevacizumab and Olaparib in Newly Diagnosed Advanced Ovarian Cancer Patients (DUO-O)

Study Lead Statistician PPD

PPD
Date

A Phase III Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab in Combination with Chemotherapy and Bevacizumab, Followed by Maintenance Durvalumab, Bevacizumab and Olaparib in Newly Diagnosed Advanced Ovarian Cancer Patients (DUO-O)

Global Product Statistician	PPD	PPD Date

A Phase III Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab in Combination with Chemotherapy and Bevacizumab, Followed by Maintenance Durvalumab, Bevacizumab and Olaparib in Newly Diagnosed Advanced Ovarian Cancer Patients (DUO-O)

PPD

Trial Steering Committee Chair

PPD

Date

A Phase III Randomised, Double-Blind, Placebo-Controlled, Multicentre Study of Durvalumab in Combination with Chemotherapy and Bevacizumab, Followed by Maintenance Durvalumab, Bevacizumab and Olaparib in Newly Diagnosed Advanced Ovarian Cancer Patients (DUO-O)

Independent Academic Statistician	PPD	PPD **Date**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation/special term	Explanation
ADA	Anti-drug antibodies
AE	Adverse event
AEPI	Adverse events of possible interest
AESI	Adverse events of special interest
AGO	Arbeitsgemeinschaft Gynäkologische Onkologie Study Group
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AML	Acute myeloid leukaemia
ANA	Antinuclear antibody
ANC	Absolute neutrophil count
AST	Aspartate aminotransferase
AUC	Area under the plasma concentration-time curve
AZD2281	Olaparib
AZRand	AZ Global Randomisation System
bd	Twice daily
BICR	Blinded independent central review
BoR	Best overall response
BP	Blood pressure
BRCA	Breast cancer susceptibility gene
BRCAm	Breast Cancer susceptibility gene mutated
CA-125	Cancer Antigen 125
CCI	CCI
CI	Confidence interval
CR	Complete response
CRO	Contract research organisation
CSP	Clinical study protocol
CSR	clinical study report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Event
CCI	CCI
CV	Coefficient of Variation
DAE	Discontinuation of Investigational Product due to Adverse Event
DBP	Diastolic blood pressure
DCO	Data Cut Off
DoR	Duration of response
ECG	Electrocardiogram
E-code	Enrolment code
ECOG	Eastern Cooperative Oncology Group

Abbreviation/special term	Explanation
eCRF/CRF	Electronic case report form
EMA	European Medicines Agency
ENGOT	European Network of Gynaecological Oncological Trial Groups
EORTC-QLQ-OV28	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Ovarian Cancer 28
EORTC-QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
ePRO	Electronic Patient-reported outcomes
EQ-5D-5L	EuroQoL five dimensions, five level health state utility index
FAS	Full analysis set
FDA	Food and Drug Administration (US)
FIGO	Fédération Internationale de Gynécologie et d'Obstétrique
HR	Homologous Recombination
HRD	Homologous Recombination Deficiency
HRR	Homologous Recombination Repair
HRRm	Homologous recombination repair related gene mutation
HRQoL	Health-related quality of life
ICF	Informed consent form
ICR	Independent Central Review
IDMC	Independent Data Monitoring Committee
IDS	Interval debulking surgery
imAE	Immune-mediated adverse event
ir-AE	Immune-related adverse event
IRT	Interactive response technology
ITT	Intention to treat
IV	Intravenous
KM	Kaplan-Meier
LCL	Lower confidence limit
LD	Longest Diameter
LDH	Lactate dehydrogenase
LLOQ	Lower limit of quantification
LPLV	Last Patient Last Visit
MDS	Myelodysplastic syndrome
MEDI4736	Durvalumab
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milli-gram
MMRM	Mixed model for repeated measures
MOA	Mechanism of action
MRI	Magnetic Resonance Imaging
MTP	Multiple Testing Procedure

Abbreviation/special term	Explanation
NA	Not applicable
nAB	Neutralizing antibody
NCI	National Cancer Institute
NE	Non-evaluable
NED	No evidence of disease
NQ	Non-quantifiable
NTL	Non-target lesions
OAE	Other Significant Adverse Event
ORR	Objective response rate
OS	Overall survival
PARP	Polyadenosine 5'diphosphoribose [poly (ADP ribose)] polymerase
PARPi	PARP inhibitor
pCR	Pathological complete response
PD	Progressive disease
PD-1	Programmed cell death protein-1
PD-L1	Programmed death-ligand 1
PFS	Progression-free survival
PFS2	Time to second progression
PGIS	Patient Global Impression of Severity of Cancer Symptoms
PID	Percentage intended dose
PK	Pharmacokinetic
PR	Partial response
PRO-CTCAE	Patient reported outcomes version of the Common Terminology Criteria for Adverse Events
Q3W	Every 3 weeks
Q4W	Every 4 weeks
Q6W	Every 6 weeks
Q12W	Every 12 weeks
Q24W	Every 24 weeks
QAPFS	Quality-adjusted progression-free survival
QTcB	QT interval (corrected for heart rate using Bazett's correction)
QTcF	QT interval (corrected for heart rate using Fridericia's correction)
Q-TWiST	Quality-adjusted time without symptoms of disease or toxicity
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria in Solid tumours. This study will use RECIST version 1.1.
REML	Restricted maximum likelihood
RFS	Relapse free survival
RoW	Rest of the world
SAE	serious adverse event

Abbreviation/special term	Explanation
SAP	statistical analysis plan
SBP	Systolic blood pressure
SD	Stable disease
SoA	Schedule of Activities
SoC	Standard of care
SpO2	Saturation of peripheral oxygen
StD	Standard deviation
tBRCA	Tumour *BRCA1* or *BRCA2* genes
tBRCAm	*tBRCA* mutated
Non-*tBRCAm*	*tBRCA* mutation not detected
TA	Therapeutic Area
TDT	Time to discontinuation or death
TFST	Time to first subsequent therapy
TIL	Tumour-infiltrating lymphocytes
TL	Target lesion
TSST	Time to second subsequent therapy
TWiST	Time without symptoms of disease or toxicity
UCL	Upper confidence limit
ULN	Upper limit of normal
US	United states of America
WHO	World Health Organisation
VUS	Variant of uncertain significance
wt	Wildtype (patients without evidence of BRCA1 or BRCA2 deleterious or suspected deleterious mutations)

AMENDMENT HISTORY

Category: Change refers to	Date	Description of change	In line with CSP? Y (version) / N / NA	Rationale
Derivation of primary or secondary endpoints	02 September 2021	Objectives updated with the primary comparisons now being Arm 3 vs Arm 1 in non-tBRCAm HRD positive population and non-tBRCAm ITT population. The comparison of Arm 2 vs Arm 1 in the non-tBRCAm ITT population is now a key secondary endpoint. (Section 1.1, 4)	Y (v6)	HRD status has been established as a clinically important biomarker in first-line advanced ovarian cancer patients (data from the PAOLA-1 study). Thus, the primary endpoint allows for assessment of effect in the non-tBRCAm HRD positive population and the non-tBRCAm ITT population.
	02 September 2021	Removal of reference to olaparib from the secondary objective to determine the proportion of patients with pCR in patients undergoing IDS. (Section 1.1)	Y (v6)	To clarify this endpoint as IDS occurs prior to receipt of olaparib/placebo within the study.
	02 September 2021	Clarification that secondary endpoints of OS, PFS2, TFST and TSST will be assessed in the populations described for PFS. (Section 1.1)	Y (v6)	To clarify the populations to be assessed in these secondary endpoints.
	02 September 2021	Clarified that safety analysis set summaries in patients who were dosed with investigational product will be summarised by randomised treatment group. The exploratory biomarker analysis set has been removed as not feasible to define one analysis set due to the multiple biomarkers of interest. ADA analysis set defined to support reporting of ADA data. (Section 2.1)	Y (v5)	To clarify reporting of safety data by treatment groups and to clarify analysis set definitions to be applied to support reporting.
	02 September 2021	Updated the definition of the analysis sets to include reference to the non-*tBRCAm* HRD-positive population (Section 2.1, 4)	Y (v6)	Clarified analysis set definitions to be applied to support reporting in the non-*tBRCAm* HRD-positive population.
	02 September 2021	Defined cohort allocation date as first dose of any study treatment at Cycle 2 (Section 2.1)	N/A	The date of allocation is not captured in IRT nor eCRF for the tBRCAm cohort

	02 September 2021	Updated text to be consistent throughout in referring to modified RECIST v1.1. Clarified that baseline scans are prior to Cycle 1, Day 1 to align with SoA and that post-baseline RECIST assessments do not need to be evaluable, due to the planned on study surgery (lesion intervention) in IDS patients. Also aligned text with AstraZeneca TA SAP standard RECIST definitions (Section 3.1, 3.2.1, 3.2.3 and 4.2.2)	Y (v5)	Updated to align statistical text with SoA, reflect planned on study interventions and patient population entered (ie, upfront primary surgery vs IDS). Also to ensure reference to correct RECIST rules, and to align with current standards.
	02 September 2021	Updated the definition of the two-missed visit rule (Section 3.2.1)	N/A	To better reflect the planned RECIST visit schedule as per CSP
	02 September 2021	Clarified that ORR will be assessed in all patients with evaluable disease (Section 3.2.3)	Y (v5)	To align with ORR definition used in other olaparib studies in ovarian cancer.
	02 September 2021	Clarified that second progression can be based on CA125 progression as outlined in study objective and that second progression event is to be subsequent to first subsequent therapy. Also removed the 2 missed visit rule as a reason for censoring (Section 3.2.5)	Y (v5)	To align with study objectives and PFS2 TA standards/regulatory definitions. Two missed visit rule removed as regular visit data not collected after 1st progression with only 2nd progressions collected at this point, and thus this part of the definition cannot be applied.
	02 September 2021	Clarified for derivation of TDT that study treatment refers to investigational treatment (i.e. durvalumab/placebo or olaparib/placebo), and that discontinuation refers to all treatments being discontinued (Section 3.2.9)	N/A	To provide clarity and to be consistent with TDT definition used in other olaparib studies in ovarian cancer.
	02 September 2021	Updated definition of pCR to be 'absence of any residual macroscopic disease together with the absence of residual microscopic disease post-surgery' and clarified that pCR will be assessed in patients with planned IDS at study entry, and in whom surgery is performed (Section 3.2.10)	N/A	To clarify how pCR is derived based on the data collected.

	02 September 2021	Updated the number of ADA patients to be 100 patients as per CSP and clarified ADA definitions/categories (Section 3.5 and 4.4.2)	Y (v5)	To align with CSP and to align ADA definitions with other durvalumab studies.
	02 September 2021	Included definition of treatment emergent adverse events as well as defined durvalumab AEPI/AESIs and imAEs (Section 3.6, 4.4.1.1 and 4.4.1.2)	N/A	To clarify TEAE, AESI/AEPI and imAE definitions.
	02 September 2021	Further clarified olaparib and durvalumab exposure derivations including how patients undergoing IDS have their exposure calculated (separately for cycles before and after IDS and then added together). Removed reference to olaparib PID intensity calculations since olaparib is not dosed until progression. Also included exposure derivation rules for the chemotherapy agents and bevacizumab (Section 3.6.2 and 3.6.3)	N/A	For clarity and given olaparib is administered for a fixed time period.
Multiple Testing Procedure	02 September 2021	Included figure to illustrate the multiple testing procedure and clarified that the interim/final OS analysis boundaries will ultimately be derived based on the actual number of events observed in the study. Also included section 5.1 to clarify the statistical boundaries to be applied at the PFS/OS interim analysis (Section 4.2.1 and Section 5)	N/A	To further clarify alpha spending rules.
Statistical analysis method for the primary or secondary endpoints	02 September 2021	Included a definition of the pooling strategy for strata for the scenario where number of events in individual stratum are too small for meaningful analysis (Section 4)	N/A	To provide guidance as per AZ standards to ensure a meaningful analysis
	02 September 2021	Removed PFS sensitivity analysis for "Randomisation Bias" as analysis is already covered in the section outlining assessment of "Consistency of treatment effect across subgroups" in more detail. In addition, determination of "HR and CI estimation" section removed as provides no useable	N/A	Removed repeated/unnecessary text

		information to help interpret results generated from the primary analysis approach (Section 4.2.2)		
	02 September 2021	Clarified that the subgroup analyses will be produced for Arm 3 vs Arm 1 , Arm 2 vs Arm 1, and Arm 3 vs Arm 2 in the non-*tBRCAm* HRD-positive and ITT populations (Section 4.2.2)	Y (v6)	Updated to align with Health authority feedback.
	02 September 2021	Defined the PD-L1 cut-offs for the PD-L1 subgroup analyses to compare high vs low vs unknown (Section 4.2.2)	N/A	To clarify definition of PD-L1 cut offs and scoring algorithms for PD-L1 high vs low vs unknown status used for subgroup analysis.
	02 September 2021	Included subgroup analysis by surgery status at study entry (upfront primary surgery or planned IDS) and clarified subgroup analysis will be done using a by statement to align with CSP. Also removed example SAS code for analyses as this will be covered within the SAS programs (Section 4.2.2)	Y (v5)	To assess impact of surgery status at study entry and align with CSP/TA standards.
	02 September 2021	Removed the requirement for responses to be confirmed for the analysis of DoR (Section 4.2.8)	N/A	To align the definition of response with ORR
	02 September 2021	Clarified QAPFS and QTWIST analysis approaches and included an analysis of TWIST. Also updated the definition of significant symptoms to align with definitions utilised by other PDL1/PD1 inhibitors (Section 4.3)	N/A	To provide clarity and to be consistent with definition used by other PDL1/PD1 inhibitors. Inclusion of TWIST analysis to enable comparison at time of PFS primary analysis.
	02 September 2021	Removed “Time to Deterioration” from the Patient Reported Outcome analyses (Section 4.3)		Time to deterioration has been removed as this endpoint is unlikely to be meaningful in this setting due to on study planned surgery (including recovery) leading to QoL deterioration and thus potentially confounding any interpretation that is

				associated specifically with study treatment
Data presentations	02 September 2021	Outlined that TEAEs would be summarised overall and by treatment phase (induction and maintenance) where required and clarified the definition of "Maintenance phase" to be used in summaries. Updated the list of AEs to produce summary information for and updated/added further information on AESI/AEPI and imAE reporting to align with durvalumab and olaparib project standards (Section 2.1, 4.4.1.1 and 4.4.1.2)	N/A	To align with project standards and to enable assessment of TEAEs following administration of olaparib maintenance treatment.
	02 September 2021	Included summaries for assessment of Thyroid Test Results and Renal Function Test Abnormalities that are standardly summarised in durvalumab studies (Section 4.4.1.3)	N/A	To follow AZ durvalumab standards
	02 September 2021	Removed reference to summarising "Overall disease classification" and "Extent of disease for baseline RECIST scan" as this information is not collected (Section 4.4.1.7)	N/A	Removed due to not collecting this data
Other	02 September 2021	Updated references of 'IWRS', 'IVRS', 'IxRS' to IRT throughout	Y (v5)	To align with updated AstraZeneca clinical study template
	02 September 2021	Updated study design section to align with CSP updates v6 (Section 1 and 4)	Y (v6)	Align with latest CSP amendments
	02 September 2021	Clarification of event numbers and data maturity required to trigger PFS analysis in the two populations of interest (ie, the non-*tBRCAm* HRD positive population and the non-*tBRCAm* ITT population) (Section 1)	Y (v6)	Event numbers and data maturity required to trigger analyses have been updated to align with the primary endpoint comparing Arm 3 vs Arm 1 in the non-*tBRCAm* HRD positive population and the non-*tBRCAm* ITT population.

	02 September 2021	Exploratory objectives updated to include relapse free survival in patients with NED at baseline, in addition to relapse free survival in patients with NED at the end of chemotherapy (Section 3.7.1)	Y (v5)	To produce summaries of relapse free survival which cover both the overall and maintenance phase of the study
	02 September 2021	Defined analyses and timings specifically for the China cohort (analyses will be conducted for the non-tBRCAm HRD positive population and the non-tBRCAm ITT population) (Section 2.1, 6)	Y (v6)	To support regulatory submissions in China.
	02 September 2021	Updated wording of important protocol deviations to clarify requirements and updated criteria (Section 2.2)	N/A	To align with the study non-compliance handling plan
	02 September 2021	Updated descriptions of visit schedules to instead refer to the CSP for details (Section 3.3)	N/A	To simplify this section in the SAP
	02 September 2021	Included additional measures of PRO compliance and clarified wording (Section 3.3.5)	N/A	To align with PRO compliance reporting utilised in other trials
	02 September 2021	Clarified windowing rules for PROs and safety, including corrections to vitals schedule to align with CSP SoA. Clarified baseline definition for efficacy/PROs. Provided further information on missing data imputation (Section 3.6.8, 4.1 and 4.4.1.1)	N/A	For clarity and to align windowing/baseline rules with SoA in CSP and to clarify missing data imputations.
	02 September 2021	Updated NED at landmark times to be assessed in patients who have no disease at baseline, and clarified that this will be defined by data captured on the baseline RECIST eCRF (Section 3.7.2)	N/A	Since patients tumour response is assessed relative to baseline (prior to randomisation) for this study by investigators and since NED can only be derived in patients who do not have disease
	02 September 2021	Included an exploratory analysis to compare Arm 3 (SoC + durvalumab + olaparib) vs Arm 2 (SoC + durvalumab) as required (Section 4)	N/A	Based on regulatory feedback to perform such comparisons

	02 September 2021	Inclusion of an interim PFS analysis when approximately 50% or more maturity has been observed for the comparisons of Arm 3 vs Arm 1 in the non-tBRCAm HRD-positive population and non-tBRCAm ITT population. (Section 4, 5)	Y (v6)	To include an interim analysis when sufficient maturity and information has occurred in the trial.
	02 September 2021	Outlined additional analyses to be performed as a result of the COVID-19 pandemic. (Section 4.2, 9.4)	N/A	To comply with AstraZeneca requirements for clinical studies ongoing during the COVID-19 pandemic
	02 September 2021	Updated scoring manual for EORTC QLQ-OV28 (Appendix 9.2)	Y (v4)	Update to align with current scoring manual

1 STUDY DETAILS

This statistical analysis plan (SAP) contains a more detailed description of the analyses in the clinical study protocol (CSP) for the DUO-O study (D081RC00001). This SAP is based on version 6.0 of the CSP.

1.1 Study objectives

Table 1 Study objectives

Primary objective:	**Endpoint/variable:**
To determine the efficacy of durvalumab and olaparib assessed by PFS in the first line treatment of non-*tBRCAm* patients with newly diagnosed advanced ovarian cancer.	PFS by investigator assessment using modified RECIST v1.1 – time from randomisation to first progression or death. This will be assessed via: • Determining the efficacy of durvalumab in combination with platinum-based chemotherapy and bevacizumab and continued as maintenance in combination with bevacizumab and olaparib versus SoC platinum-based chemotherapy in combination with bevacizumab in the following populations: o non-*tBRCAm* HRD positive population o non-*tBRCAm* ITT population.
Secondary objectives:	**Endpoint/variable:**
To determine the efficacy of durvalumab assessed by PFS in the first line treatment of non-*tBRCAm* patients with newly diagnosed advanced ovarian cancer.	PFS by investigator assessment using modified RECIST 1.1 – time from randomisation to first progression or death. This will be assessed via: • Determining the efficacy of durvalumab in combination with platinum-based chemotherapy and bevacizumab and continued as maintenance in combination with bevacizumab versus SoC platinum-based chemotherapy in combination with bevacizumab in the non-*tBRCAm* ITT population

To determine the efficacy of durvalumab and olaparib assessed by OS in the first line treatment of non-*tBRCAm* patients with newly diagnosed advanced ovarian cancer.	OS – time from date of randomisation to death This will be assessed via: ◦ Arm 3 vs Arm 1 in non-*tBRCAm* HRD positive population ◦ Arm 3 vs Arm 1 in non-*tBRCAm* ITT population ◦ Arm 2 vs Arm 1 in non-*tBRCAm* ITT population.
To assess the efficacy of durvalumab and olaparib in terms of PFS2, ORR, ORR pre-surgery in IDS group, duration of response, TFST, TSST and TDT in the first line treatment of non-*tBRCAm* patients with newly diagnosed advanced ovarian cancer.	• Time from date of randomisation to second progression by investigator assessment of radiological, clinical or CA125 progression or death (PFS2) • ORR (CR + PR) by investigator assessment by modified RECIST v1.1 ◦ in all patients with evaluable disease at baseline ◦ prior to surgery in those patients planned to have IDS with evaluable disease at baseline. • Duration of response • TFST • TSST • TDT All endpoints will be assessed in the populations described for PFS and OS.
To determine the effects on HRQoL, global health status and ovarian cancer symptoms of the combination of durvalumab and olaparib in the first line treatment of non-*tBRCAm* patients with newly diagnosed advanced ovarian cancer.	• Physical functioning subscale of the EORTC QLQ-C30 and EORTC QLQ-OV28 • Health state utility derived from the HRQoL instrument, the EuroQOL EQ5D-5L • Quality-adjusted Time Without Symptoms of disease or Toxicity (Q-TWiST)[a] • Quality-Adjusted PFS (QAPFS)[a] All endpoints will be assessed in the populations described for PFS and OS.
To determine the effects on pCR for the combination of durvalumab with platinum-based chemotherapy and bevacizumab in the first line treatment of non-*tBRCAm* patients with newly diagnosed advanced ovarian cancer.	• Proportion of patients with pCR in patients undergoing IDS
To characterize the PK and immunogenicity of durvalumab in combination with bevacizumab and olaparib To determine olaparib plasma concentrations via sparse sampling for population PK analyses.	Only in non-*tBRCAm* patients with primary cytoreductive surgery; • Serum concentrations of durvalumab and plasma concentrations of olaparib (samples to be taken in the non-*tBRCAm* cohort only) • ADA to durvalumab

To assess the potential additional clinical benefit of durvalumab added to SoC and olaparib in the first line treatment of *tBRCAm* patients with newly diagnosed advanced ovarian cancer.	PFS, PFS2, ORR, ORR pre-surgery in IDS group, duration of response, TFST, TSST, OS, TDT EORTC QLQ-C30 and EORTC QLQ-OV28, EQ 5D-5L, Q-TWiST [a], TWiST [a], QAPFS [a], proportion of patients with pCR in patients undergoing IDS
Safety objectives:	**Endpoint/variable: All patients**
To evaluate the safety and tolerability of durvalumab in combination with platinum based chemotherapy+/- bevacizumab and continued as maintenance in combination with olaparib +/-bevacizumab in patients with newly diagnosed advanced ovarian cancer To evaluate the safety and tolerability of the combination of durvalumab and bevacizumab given in combination with platinum-based chemotherapy and continued as maintenance in patients with newly diagnosed advanced ovarian cancer	AEs / SAEs, physical examination, vital signs including BP, pulse, ECG and laboratory findings including clinical chemistry / haematology parameters irAEs – given the intended mechanisms of action of durvalumab, particular attention will be given to AEs that may follow enhanced T-cell activation, or other irAE
Exploratory objectives:	**Endpoint/variable: All patients**
To determine the efficacy of olaparib and durvalumab in newly diagnosed advanced ovarian cancer by assessment of relapse free survival in patients who have NED/CR on their end of chemotherapy CT scan.	Relapse free survival for patients who have NED/CR at the end of chemotherapy: time from randomisation to disease progression
To determine the efficacy of olaparib and durvalumab in newly diagnosed advanced ovarian cancer by assessment of the proportion of patients who have NED at 15, 24 and 48 months after initiation of treatment.	Proportion of patients who have NED at 15, 24 and 48 months after randomisation
To explore the impact of treatment and disease state on health state utility and to explore the impact of treatment and disease on resource use	Number, type and reason of hospitalisations and hospital attendances, procedures undertaken and hospital length of stay
To assess patient reported treatment-related side effects of olaparib, durvalumab and bevacizumab	Items selected from the PRO-CTCAE item library (see Section 3.3.2 of SAP)
To assess patients' overall impression of the severity of their cancer symptoms	PGIS
To evaluate tumour HR deficiency status as candidate predictive biomarkers of olaparib and durvalumab in newly diagnosed advanced ovarian cancer patients [b].	May include, but is not limited to the following measurements within the tumour: • Mutation status of HRR genes, • HRD status CCI • CCI [redacted].

To evaluate additional candidate predictive biomarkers of olaparib and durvalumab in newly diagnosed advanced ovarian cancer patients [b].	May include, but is not limited to: • PD-L1, CCI CCI CCI CCI CCI CCI CCI
To further assess the efficacy of treatment through longitudinal analysis of blood samples collected at regular intervals on study [a]	May include but is not limited to CCI response to treatment.
To explore whether resistance mechanisms to treatment can be identified through analysis of tumour and blood samples – archival tumour sample and blood samples at baseline and on progression (tumour sample optional on progression) [a].	Analysis and outcome variables yet to be defined but may include molecular analysis of CCI.
Future exploratory research into factors that may influence development of cancer and/or response to study treatment (where response is defined broadly to include efficacy, tolerability or safety) may be performed on the collected and stored blood or archival tumour samples that were mandatory for entry onto the study or on optional blood or tumour biopsy samples collected during the study [a].	Analysis and outcome variables yet to be defined.
To collect and store DNA according to each country's local and ethical procedures) for future exploratory research into genes/genetic variation that may influence response (i.e., distribution, safety, tolerability and efficacy) to study treatments and or susceptibility to disease (optional) [a].	To identify pharmacogenetic correlates for the response to treatment through the retrospective analysis of DNA extracted from an optional blood sample.

[a] These endpoints may be reported separately to the clinical study report.

[b] Homologous Recombination Repair related gene mutation (HRRm) status, Homologous Recombination Deficiency (HRD) status, and Programmed death ligand 1 (PD-L1) expression status will be reported in the Clinical Study Report (CSR). Other exploratory biomarker will be reported outside the CSR.
CT = Computed tomography; ICF = Informed consent form; MRI = Magnetic resonance imaging.

1.2 Study design

The Phase III DUO-O study will assess the efficacy and safety of durvalumab and olaparib when added to standard of care (SoC) in patients with newly diagnosed advanced ovarian cancer who have either undergone upfront primary surgery or plan to start chemotherapy followed by interval debulking surgery (IDS). The study consists of 2 independent cohorts which are defined by the breast cancer sensitivity gene (*BRCA*) mutated (*BRCAm*) status, based on testing of tumour tissue (ie, *tBRCAm* status) of the patients.

Non-*tBRCAm* cohort

Patients who meet the inclusion criteria and none of the exclusion criteria, have signed the main informed consent and have no deleterious/suspected deleterious mutations in *BRCA1* or *BRCA2* as assessed by central testing (Myriad myChoice HRD plus test, with the exception of patients enrolled from China, where a China-based central assay is used) will be randomised in a 1:1:1 ratio to receive up to a further 5 cycles of platinum-based chemotherapy (in combination with bevacizumab followed by bevacizumab maintenance for up to a total of 22 cycles [15 months of treatment]). The patients will be randomised into three treatment arms:

1 **Arm 1 (SoC):** Patients in Arm 1 will receive saline IV Q3W as a placebo for durvalumab from Day 1 of Cycle 2 up to a total of 35 cycles (24 months) of treatment. At the end of chemotherapy, patients will also receive placebo tablets, matched to olaparib for up to a total of 24 months.
2 **Arm 2:** Patients in Arm 2 will receive durvalumab 1120 mg Q3W from Day 1 of Cycle 2 up to a total of 35 cycles (24 months) of total treatment. At the end of chemotherapy, patients will also receive placebo tablets, matched to olaparib for up to a total of 24 months.
3 **Arm 3:** Patients in Arm 3 will receive durvalumab 1120 mg Q3W from Day 1 of Cycle 2 up to a total of 35 cycles (24 months) of total treatment. At the end of chemotherapy, patients will also receive olaparib tablets, 300 mg twice daily (bd) for up to a total of 24 months.

The randomisation scheme will be stratified according to:

- Timing and outcome of cytoreductive surgery: no macroscopic residual disease after upfront primary surgery vs all others (macroscopic residual disease after upfront primary surgery OR planned IDS)
- Geographic region: North America vs Europe vs RoW.

***tBRCAm* cohort**

Patients who meet the inclusion criteria and none of the exclusion criteria, have signed the main informed consent and have deleterious/suspected deleterious mutations in *BRCA1* or *BRCA2* as identified by central *tBRCA* testing **(Myriad myChoice HRD plus test)** will be allocated into a single arm cohort. Patients in this cohort will receive up to a further 5 cycles of platinum-based chemotherapy (and optional bevacizumab for a total of 15 months of treatment according to standard local practice). Investigational treatments will be given as specified below:

i. Durvalumab 1120 mg Q3W from Day 1 of Cycle 2 up to a total of 35 cycles (24 months) of total treatment.
ii. At the end of chemotherapy, patients will also receive olaparib tablets, 300 mg twice daily (bd) for up to a total of 24 months

1.3 Number of subjects

Approximately 1254 patients with newly diagnosed advanced ovarian cancer patients will be randomised/allocated to study treatment at approximately 225 study sites worldwide. Patients will be assigned to one of two independent cohorts as follows:

- Approximately 1104 patients in the non-*tBRCAm* cohort (with no deleterious/suspected deleterious mutations in *BRCA1* or *BRCA2*) will be randomised to receive 1 of 3 treatment options.
- Approximately 150 patients in the *tBRCAm* cohort (with deleterious/suspected deleterious mutations in *BRCA1* or *BRCA2* as identified by central *tBRCA* testing) will be allocated into an open-label single arm cohort.

The sample size for the non-*tBRCAm* cohort was derived using a validated non-proportional hazards-based AstraZeneca R package for sample size calculations to test the following hypotheses of interest with regards to the efficacy in the non-*tBRCAm* cohort (see Section 4.1) as follows:

- The assumed median PFS of 18 months for Arm 1 based on results seen in the PAOLA-1 study (Ray-Coquard et al 2019) in addition to the expected average time for chemotherapy and the inclusion of the stable disease population in the DUO-O study. The original assumed median PFS for the control arm was 16 months and was based on the data reported in the GOG218, ICON7 and GOG262 clinical trials with bevacizumab in the first line setting, including subgroup analysis from the GOG218 study based on *BRCA*/HRR wild type status (Perren et al 2011; Burger et al 2013; Norquist et al 2018).
- The sample size has been derived on the assumption of a 3 months delay in separation of the PFS curves between Arm 2 vs Arm 1 and between Arm 3 vs Arm 1. The assumed average hazard ratio for PFS for the chemotherapy, bevacizumab, durvalumab + olaparib arm in the non-*tBRCAm* HRD-positive population is 0.49 (approximately 39 months median PFS), in the non-*tBRCAm* ITT population is 0.61 (approximately 30 months median PFS) and for the chemotherapy, bevacizumab and durvalumab arm in the non-*tBRCAm* ITT population it is 0.74 (approximately 24 months median PFS).
- In addition, the sample size has been derived on the assumption that 15% of patients will drop out.

The data cut off (DCO) for the analysis of PFS for the two comparisons of interest (Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population and Arm 3 vs Arm 1 in the non-*tBRCAm* ITT population) will be undertaken at the same calendar time. The primary PFS analysis will be undertaken when approximately 149 PFS events have occurred (58% maturity) for the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population and

approximately 453 PFS events (62% maturity) for the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* ITT population. At the time of the PFS primary analysis approximately 480 PFS events (65% maturity) are expected to have occurred for the comparison of Arm 2 vs Arm 1 in the non-*tBRCAm* ITT population. Assuming a non-linear recruitment period of 26 months, the data cut off for the analysis of PFS will take place at approximately 52 months after the first patient has been randomised.

The PFS comparisons of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population are adequately powered, under the assumed effect sizes using a hierarchical testing procedure with a significance level alpha of 5% (two-sided) for each comparison.

For the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population, if the average true PFS hazard ratio is 0.49, the study will have >90% power to demonstrate a statistically significant difference at a two-sided alpha level of 5% overall . The smallest treatment difference that would be statistically significant is an average hazard ratio of 0.72.

For the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* ITT population, if the average true PFS hazard ratio is 0.61, the study will have >90% power to demonstrate a statistically significant difference at a two-sided alpha level of 5% overall. The smallest treatment difference that would be statistically significant is an average hazard ratio of 0.83.

For the comparison of Arm 2 vs 1 in the non-*tBRCAm* ITT population, if the average true PFS hazard ratio is 0.74, the study will have >80% power to demonstrate a statistically significant difference at a two-sided alpha level of 2.5% overall. The smallest treatment difference that would be statistically significant is an average HR of 0.81.

An interim PFS analysis is planned and is described in detail within Section 5.1.

The power calculations for OS were based on the following assumptions:

- A 3 month delay in the separation of the OS curves between Arm 2 vs Arm 1 and between Arm 3 vs Arm 1
- Median OS of 55 months for Arm 1
- An average hazard ratio of 0.79 for OS for Arm 3 vs Arm 1 comparison in both the non-*tBRCAm* HRD-positive population and the non-*tBRCAm* ITT population (hazard ratio 0.78 from 3 months onwards [approximately 70 months median OS])
- An average hazard ratio of 0.86 for OS for Arm 2 vs Arm 1 comparison in the non-*tBRCAm* ITT population (hazard ratio 0.85 from 3 months onwards [approximately 65 months median OS])

At the time of the primary analysis of PFS, the interim analysis of OS will also be performed. For the comparison of Arm 3 vs Arm 1 it is anticipated that the maturity will be approximately 31% in both the non-*tBRCAm* HRD-positive population and the non-*tBRCAm* ITT population

and for the comparison of Arm 2 vs Arm 1 in the non-*tBRCAm* ITT population, it is anticipated that the maturity will be approximately 32%.

A final analysis of OS may be performed at approximately 50% maturity across the 3 treatment arms in the non-*tBRCAm* ITT population or 5 years following randomisation of the last non-*tBRCAm* patient, whichever occurs sooner. At this time approximately 129 OS events (50% maturity) are expected to have occurred for the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population, 369 OS events (50% maturity) for the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* ITT population, and 379 OS events (51% maturity) for the comparison of Arm 2 vs Arm 1 in the non-*tBRCAm* ITT population.

The power to detect a difference in OS between Arm 3 and Arm 1 will be approximately 17% in the non-*tBRCAm* HRD-positive population and approximately 48% in the non-*tBRCAm* ITT population at the 2.5% level (using a two-sided test). The power to detect OS difference between Arm 2 and Arm 1 in the non-*tBRCAm* ITT population will be approximately 21% at the 2.5% level (using a two-sided test). Note that these estimates assume that no confounding will occur. AstraZeneca anticipates potential confounding of OS data due to availability of PARP inhibitors for ovarian cancer patients.

2 ANALYSIS SETS

Note, Global recruitment to the study will close when approximately 1104 non-t*BRCA*m patients are randomised. If necessary, enrolment in China will continue after global recruitment is closed (ie, last subject randomised from a non-Chinese site) to allow inclusion of a China cohort consisting of approximately 120 non-*tBRCAm* randomised patients. The China cohort will support standalone safety and efficacy analyses of the non-t*BRCA*m patients from sites in China (please see Section 6 for details).

All populations and planned analyses described, relate to the Global population unless otherwise stated. A patient randomised in China prior to global recruitment closure will be included in both the Global ITT population and the China cohort ITT population. A patient randomised in China after the global recruitment closure will be included only in the China cohort ITT population.

2.1 Definition of analysis sets

All patients

The All Patients Set includes all enrolled patients who signed the informed consent form. In addition to the set of All patients, four analysis sets will be defined:

- Full Analysis Set (for efficacy),
- Safety Analysis Set (for safety),
- Pharmacokinetic (PK) Analysis Set (for PK), and
- ADA analysis set

Similar analysis sets will be defined for both the non-*tBRCAm* and *tBRCAm* cohorts except for the PK and ADA analysis sets which will not be defined for the *tBRCAm* cohort as no PK and ADA data will be collected for the open-label arm.

Full analysis set

For the non-*tBRCAm* cohort patients, the Full Analysis Set (FAS) will include all randomised patients with treatment groups assigned in accordance with the randomisation scheme, regardless of whether the treatment was actually received. All efficacy and HRQoL data in the non-*tBRCAm* cohort will be summarised and analysed for this cohort using the FAS on an intention to treat (ITT) basis.

There are two primary efficacy populations in the non-*tBRCAm* cohort: the non-*tBRCAm* HRD-positive population and the non-*tBRCAm* ITT population.

- The non-*tBRCAm* HRD-positive population includes all non-*tBRCAm* patients who are randomised into the study and identified as HRD positive (GIS ≥ 42).
- The non-*tBRCAm* ITT population includes all non-*tBRCAm* patients who are randomised into the study.

For the *tBRCAm* cohort patients, the Full Analysis Set (FAS) will include all cohort allocated patients as defined in the IRT system regardless of whether the treatment corresponding to the cohort was actually received. All efficacy and HRQoL data will be summarised and analysed for this cohort using the FAS on an ITT basis. Note as cohort allocation date was not collected for *tBRCAm* patients in the IRT system, date of allocation to the *tBRCAm* cohort will be defined as the date of first dose of any study treatment (durvalumab, olaparib, carboplatin, paclitaxel or bevacizumab) at Cycle 2.

Safety analysis sets

For the non-*tBRCAm* cohort, the Safety Analysis Set will comprise all patients who received any amount of the investigational treatments (i.e., durvalumab/olaparib), including placebo, regardless of whether that was the randomised therapy intended or indeed whether, in rare cases, they received therapy without being randomised.

Patients who initially received a dose of Durvalumab/Durvalumab placebo will be summarised according to the arm they are randomised to. This is in order to provide a summary of the underlying safety profile that patients should expect when initially prescribed to treatment (i.e. SoC, SoC + Durvalumab, or SoC + Durvalumab + Olaparib).

A select number of safety summaries may also be repeated as required in the non-*tBRCAm* HRD-positive population who received any amount of the investigational treatments (i.e., durvalumab/olaparib), including placebo.

For the *tBRCAm* cohort, the Safety Analysis Set will comprise all patients who received any amount of the investigational treatments (i.e., durvalumab/olaparib), regardless of whether, in rare cases, they received therapy without having the required *BRCA* status.

Maintenance Phase: Number of patients in the Safety Analysis Set who enter the maintenance phase (defined as receiving at least one dose of olaparib/olaparib placebo) will be summarised and additional summaries of safety by phase may also be generated, as required.

Pharmacokinetic analysis set (non-*tBRCAm* patients with primary cytoreductive surgery only)

The PK analysis set will be comprised of all evaluable patients (a subset of those in the full analysis set) dosed with SoC + durvalumab or SoC + durvalumab + olaparib and providing at least one post-dose analysable concentration of durvalumab and/or olaparib. The population will be defined by the Study Physician, Pharmacokineticist, and Statistician prior to any analyses being performed.

ADA Analysis Set (non-*tBRCAm* patients with primary cytoreductive surgery only)

The ADA evaluable subjects are patients in the Safety Analysis Set who received at least 1 dose of durvalumab and have non-missing baseline ADA and at least 1 post-baseline ADA result.

See summary of outcome variables/study endpoints and analysis set associated with them in **Table 2**.

Table 2 Summary of outcome variables and analysis sets

Outcome variable	***Analysis set***[+]
Efficacy data	
PFS	*FAS*
PFS2, OS, TFST, TSST, TDT, ORR*, duration of response*, RFS, NED at landmark time points, symptom/HRQoL endpoints (including EORTC-QLQ-C30 and EORTC-QLQ-OV28, EQ5D-5L, Q-TWiST[a], TWiST[a], QAPFS[a], PGIS, PRO-CTCAE), and proportion of patients with pCR in patients undergoing IDS	*FAS*
Study Population/Demography Data	
Disposition	*All Patients*
Demography characteristics (e.g. age, sex etc.)	*FAS*
Baseline and disease characteristics	*FAS*
Important deviations	*FAS*
Medical/surgical history	*FAS*
Previous anti-cancer therapy	*FAS*

Table 2 Summary of outcome variables and analysis sets

Outcome variable	***Analysis set***[+]
Concomitant medications/procedures	*FAS*
Subsequent anti-cancer therapy	*FAS*
PK Data	
PK data	*PK*
Immunogenicity Data (durvalumab only)	
Immunogenicity (ADA) data	*ADA*
Safety data	
Exposure	*Safety*
Adverse events	*Safety*
Laboratory measurements	*Safety*
Vital signs	*Safety*
ECGs	*Safety*
Biomarker Data	
Biomarkers[a] (including Mutation status of HRR genes, GIS score, PD-L1, CCI CCI CCI CCI CCI CCI CCI CCI CCI).	*FAS*

*Patients who are evaluable for the analysis of ORR are those with evaluable disease at baseline. Patients who are evaluable for the analysis of DoR are those who responded in the ORR analysis.

[a] Outcomes may be reported separately outside the CSR.

[+] The analysis sets are defined separately for patients in the non-*tBRCAm* and *tBRCAm* cohorts. Selected outputs maybe repeated in the non-*tBRCAm* HRD-positive population, as required

2.2 Violations and deviations

The following general categories will be considered important protocol deviations and will be programmatically derived from the eCRF data when applicable. The important protocol deviations (IPDs) will be listed and summarised and discussed in the CSR as appropriate. The following IPDs will be reported:

- Patients randomised/allocated but who did not receive investigational treatment
- Patients who received the investigational treatment at an incorrect dose or received an alternative investigational treatment to that which they were randomised/allocated
- Patients who deviate from key entry criteria (inclusion criteria 1, 3, 4, 5, 7, 10 , and exclusion criteria 1, 2, 4, 7, 8, 24, 25, 27) per the Clinical Study Protocol
- Patient who met discontinuation criteria but continued to be on study treatment
- Patients with no baseline RECIST v1.1 assessment on or before Day 1 of Cycle 1 or a baseline RECIST scan > 42 days before Day 1 of Cycle 1
- Received prohibited concomitant medications (including other anti-cancer agents). Please refer to the CSP section 6.5 for those medications that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to database lock. Concomitant use of hormonal therapy for non-cancer-related conditions (e.g., hormone replacement therapy) or bisphosphonates, if indicated, is acceptable

Further details on the classification can be found in the study non-compliance handling plan. Note any additional important protocol deviations identified beyond those listed above will be captured and documented prior to database lock.

Patients who received the wrong treatment at any time will be included in the safety analysis set as described in section 2.1. During the study, decisions on how to handle errors in treatment dispensing (regarding the continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader and/or statistician.

The important protocol deviations will be listed and summarised using FAS by cohort allocation. For the non-*tBRCAm* cohort patients, IPDs will be listed and summarised by randomised treatment groups. For the *tBRCAm* cohort patients, IPDs will be listed and summarised by the single arm cohort. The deviation listed in the first bullet above will lead to exclusion from the safety analysis set. None of the other deviations will lead to patients being excluded from any of the analysis sets described in Section 2.1 (with the exception of the PK analysis set, a patient will be excluded from the PK analysis if there are any deviations that are considered to impact upon PK analysis).

A per-protocol analysis excluding patients with specific important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis may be performed on the progression free survival endpoint for the non-*tBRCAm* cohort only. This 'deviation bias' sensitivity

analysis is a re-analysis of the PFS excluding patients with deviations that may affect the efficacy of the trial therapy if > 10% of patients in whichever treatment group:

- Did not have the intended disease or indication at screening or
- Did not receive any randomised therapy.

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

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

3 PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST Visit Responses

This section details the implementation of protocol-specific requirements in the assessment of Response Evaluation Criteria In Solid Tumours (RECIST) version 1.1 guidelines (Eisenhauer et al 2009) with regards to the per visit and overall Investigator assessment of tumour burden for this study. RECIST has been modified to allow the assessment of progression due to new lesions in patients with no evidence of disease at baseline.

Patients will undergo regular tumour assessments until documented objective disease progression according to the investigator as defined by modified RECIST v1.1. For all patients, the RECIST tumour response data will be used to determine each patient's visit response. At each visit, patients will be programmatically assigned a modified RECIST v1.1 visit response. Modified RECIST v1.1 criteria will be used to assess patient response to treatment and PFS times and ORR determined. The modified RECIST v1.1 provides guidelines for determining measurable, non-measurable, target lesions (TLs) and non-target lesions (NTLs) and the objective tumour response criteria (CR, PR, SD, PD, non-evaluable [NE], no evidence of disease [NED]) depending on the status of their disease compared with baseline and previous assessments.

If a patient has had a tumour assessment that cannot be evaluated then the patient will be assigned a visit response of non-evaluable (NE), unless there is evidence of progression in which case the response will be assigned as PD.

Baseline assessments should encompass all areas of known predilection for metastases in the disease under evaluation and should additionally investigate areas that may be involved based on signs and symptoms of individual patients. Baseline radiological tumour assessments are to be performed no more than 28 days prior to Day 1 of Cycle 1 of chemotherapy. Patients with upfront cytoreductive surgery should have the baseline scan conducted post-surgery.

Following the baseline assessment, the first follow-up tumour assessment according to modified RECIST v1.1 (the end of chemotherapy assessment scan) will be performed within 3 weeks ±1 week of the last dose of chemotherapy and before the start of maintenance. Thereafter, assessments will be performed every 12 weeks ±2 weeks for the first 3 years (156 weeks) and then every 24 weeks ± 2 weeks, relative to the date of end of chemotherapy assessment scan, according to the planned study schedule until disease progression by the modified RECIST v1.1. Patients requiring IDS will have an additional scan pre-surgery (see Table 12 Section 8.1 of CSP for schedules for both types of surgery).

If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits. This schedule by type of debulking surgery received is to be followed to minimise any unintentional bias caused by some patients being assessed at a different frequency than other patients.

RECIST outcomes (i.e. PFS, ORR etc.) will be calculated programmatically for the site investigator assessments (see section 3.2) from the overall visit responses.

Please refer to Table 2 and Table 3 below for the definitions of CR, PR, SD, PD and NE.

Patients with measurable or non-measurable disease or no evidence of disease assessed at baseline by computed tomography (CT) / magnetic resonance imaging (MRI) will be entered in this study.

3.1.1 Target lesions (TLs)

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (LD), (except lymph nodes which must have short axis ≥ 15 mm) with computed tomography (CT) or magnetic resonance imaging (MRI) and which is suitable for accurate repeated measurements. For discussion on measurable and unmeasurable lesions, see Appendix H2 of CSP.

A patient can have a maximum of 5 measurable lesions recorded at baseline, with a maximum of 2 lesions per organ (representative of all lesions and suitable for accurate repeated measurement) and these are referred to as target lesions (TLs).

The site and location of each TL should be documented as well as the longest diameter for non-nodal lesions (short axis for lymph nodes). All measurements should be recorded in millimetres. At baseline the sum of the diameters for all TL will be calculated and reported as the baseline sum of diameters. At follow-up visits the sum of diameters for all TL will be calculated and reported as the follow-up sum of diameters.

If more than one baseline scan is recorded, then measurements from the one that is closest and prior to Cycle 1 Day 1 will be used to define the baseline sum of TLs. It may be the case that,

on occasion, the largest target 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.

Note: For patients who do not have measurable disease at entry (i.e. no TLs) but have non-measurable disease, evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions (see section 3.1.3 for further details). If a patient does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

For patients with no disease at baseline (i.e. no TLs and no NTLs), evaluation of overall visit responses will be based on absence/presence of new lesions. If no TLs and no NTLs are recorded at a visit, both the TL and NTL visit response will be recorded as NA and the overall visit response will be no evidence of disease (NED). If a new lesion is observed, then the overall visit response will be PD.

Table 3 TL Visit Responses

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

The percentage increase/decrease taking baseline as a reference for post-baseline visit will be calculated as:

[Post-baseline sum of diameters of TL – Baseline sum of diameters of TL] *100/ (Baseline sum of diameters of TL)
And similarly, for the percentage change at a post-baseline visit using previous study minimum (post-baseline nadir) as a reference.

Rounding of TL data

Before assigning target lesion response of PD and PR, the TLs percentage changes from baseline and previous minimum should be rounded to 1 decimal place. For example, 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%.

Missing TL data

For a visit to be evaluable, all target lesion 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.

Lymph nodes

For lymph nodes, if the size reduces to < 10 mm 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 < 10 mm and all other TLs are 0 mm then although the sum may be >0 mm the calculation of TL response should be over-written as a CR.

TL visit responses subsequent to CR

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

- **Step 1:** If all TLs meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL are also met i.e. if a lymph node longest diameter (LD) increases by 20% but remains < 10 mm.
- **Step 2:** If some TL measurements are missing but all other lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to NE

irrespective of whether when referencing the sum of TL diameters, the criteria for PD is also met.

- **Step 3:** If not all TLs 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 target lesion 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 5 mm 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 (for example, irradiation/surgery/embolisation), should be handled in the following way and once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours:

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

Special Case: Lesions after Interval Debulking Cytoreductive Surgery

For patients who undergo IDS, lesions removed would be treated similarly as lesions that have had interventions.

Example of scaling

Lesion	Longest diameter at nadir visit	Longest diameter at follow-up visit
1	72	71
2	67	64
3	43	40
4	86	85
5	25	Intervention
Sum	293	260

Lesion 5 has had an intervention at the follow-up visit. The sum of lesions 1-4 at the follow-up is 260 mm. The sum of the corresponding lesions at baseline visit is 268 mm.

Scale up as follows to give an estimated TL sum of 284mm:

$$\frac{260}{268} \times 293 = 284\text{mm}$$

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

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

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of target lesions

CT (preferred) and MRI are the only methods of assessment that can be used within the trial for TL measurement. 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 Non-Target Lesions (NTLs) and new lesions.

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

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

NTL response will be derived based on the Investigator's overall assessment of NTLs as follows in **Table 4**:

Table 4 NTL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
Progression (PD)	Unequivocal progression of existing non-target lesions. 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 non-target lesions were not assessed or had a lesion intervention and, in the Investigator's opinion, they are not able to provide an evaluable overall non-target lesion assessment at this visit. Note: For patients without target lesions at baseline, this is relevant if any of the non-target lesions were not assessed at this visit and the progression criteria have not been met.
Not Applicable (NA)	Only relevant if there are no NTLs at baseline

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

New lesions

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 (PD) irrespective of the TL and NTL response.

The finding of a new lesion should be unequivocal: i.e. not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumour.

New lesions will be identified via a Yes/No tick box. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses.

If the question ‘Any new lesions since baseline’ has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present and should be treated as NE in the derivation of overall visit response. If the new lesion is equivocal, for

example because of its small size, continued therapy and follow-up evaluation will clarify if it represents truly new disease. If repeat scans confirm there is definitely a new lesion, then progression should be declared using the date of the initial scan.

NOTE: The scenario whereby new lesion response is NE, should only occur in exceptional cases, as missing data for the new lesion field should always be queried.

Symptomatic deterioration

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

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

3.1.3 Overall visit response

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

Table 5 Overall Visit Response

Target Lesions	Non-target lesions	New Lesions	Overall Visit Response
CR	CR (or NA)	No (or NE)	**CR**
CR	Non-CR/Non-PD or NE	No (or NE)	**PR**
PR	Non-PD or NE or NA	No (or NE)	**PR**
SD	Non-PD or NE or NA	No (or NE)	**SD**
PD	Any	Any	**PD**
Any	PD	Any	**PD**
Any	Any	Yes	**PD**
NE	Non-PD or NE or NA	No (or NE)	**NE**
NA	CR	No (or NE)	**CR**
NA	Non-CR/Non-PD	No (or NE)	**SD**
NA	NE	No (or NE)	**NE**
NA	NA	No (or NE)	**NED**

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, NE = not evaluable, NED = no evidence of disease, NA = not applicable (only relevant if there were no TL/NTL at baseline).

3.1.4 Independent review

A blinded independent central review (BICR) of scans will be performed as sensitivity analysis to identify and quantify any bias in the local investigator-led determination of progression for patients in the non-*tBRCAm* cohort. Retrospective central reading of the scans from *tBRCAm* cohort patients is not planned, but may be conducted, if needed.

The planned BICR of all radiological imaging data will be carried out using modified RECIST v1.1. Anonymised copies of all radiological scans from all patients (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 modified RECIST v1.1 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 patient, 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 example:

- For patients with TLs at baseline: CR, PR, SD, PD, NE;
- For patients with NTLs only: CR, SD, PD, NE;
- For patients with no disease identified at baseline: PD, no evidence of disease [NED], NE).
- If a patient has had a tumour assessment that cannot be evaluated, then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD).

RECIST assessments/scans contributing towards a visit may be performed on different dates and for the BICR, 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 assigned 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 ICR RECIST information including dates of progression, visit response, censoring and changes in target lesion dimensions. Only PFS will be derived programmatically from this information.

Results of this independent review will not be communicated to investigators and the management of patients will be based solely upon the results of the modified RECIST v1.1 assessment conducted by the investigator.

A BICR of all patients will be performed for the final database lock for PFS, which will cover all the scans up to the DCO. After the primary PFS analysis, BICR review of scans will no longer be required.

Further details of the BICR will be documented in the BICR Charter. The BICR charter will detail the BICR conducted by the AstraZeneca appointed Contract Research Organisation (CRO) and will be developed in advance at the start of the study.

3.2 Efficacy Variables

3.2.1 Progression free survival (PFS)

For the non-*tBRCAm* cohort, PFS is defined as the time from randomisation to the date of objective disease progression according to modified RECIST v1.1 or death (by any cause in the absence of progression) regardless of whether the patient discontinues randomised therapy or receives another anti-cancer therapy prior to progression.

That is;

PFS = [Date of RECIST progression/death or censoring – date of randomisation + 1].

For the *tBRCAm* cohort, PFS is defined as the time from allocation to the *tBRCAm* cohort (see section 2.1 for definition) to the date of objective disease progression according to RECIST v1.1 or death (by any cause in the absence of progression) regardless of whether the patient discontinues study therapy or receives another anti-cancer therapy prior to progression.

That is;

PFS = [Date of RECIST progression/death or censoring – date of allocation to cohort + 1].

Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last RECIST assessment. However, if the patient progresses or dies immediately after two or more consecutive missed RECIST visits, the patient will be censored at the time of the latest RECIST assessment prior to the two missed visits (Note: a NE visit is not considered as a missed visit).

If the patient has no RECIST visits or does not have baseline data, they will be censored at Day 1 (date of randomisation/cohort allocation) unless they die within two RECIST visits of randomisation/cohort allocation: 27 weeks plus 2 weeks allowing for a late assessment within the visit window (Note: the pre-surgery scan for IDS patients will be treated the same as an unscheduled scan for this derivation).

Given the scheduled visit assessment scheme (i.e. end of chemotherapy (week 15 from randomisation/allocation at Cycle 2 in case of PDS patients, or as soon as possible after week 15 in the case of IDS patients), twelve-weekly through to week 171 from randomisation/allocation (i.e. 156 weeks from end of chemotherapy scan per CSP) and then every 24 weeks thereafter), the definition of 2 missed visits will change. The definition of two missed visits is given below (Note: the pre-surgery scan for IDS patients will be treated the same as an unscheduled scan for the purpose of this derivation):

- If the previous RECIST assessment is prior to week 15 then two missed visits will equate to 29 weeks (i.e. 15 weeks + 12 weeks plus 2 weeks allowing for a late assessment within the visit window = 29 weeks).

- If the previous RECIST assessment is between week 15 and less than week 157 then two missed visits will equate to 28 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 12 weeks + 2 week for an early assessment + 2 week for a late assessment = 28 weeks).

- If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from twelve-weekly to every 24 weeks this will equate to 40 weeks (i.e. take the average of 12 and 24 weeks which gives 18 weeks and then apply same rationale, hence 2 x 18 weeks + 2 week for an early assessment + 2 week for a late assessment = 40 weeks). The time period for the previous RECIST assessment will be from week 157 to week 169.

- From week 169 onwards (when the scheduling changes to every 24 weeks assessments), two missing visits will equate to 52 weeks (i.e. 2 x 24 weeks + 2 week for an early assessment + 2 week for a late assessment = 52 weeks).

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

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

- For BICR assessments, date of progression will be determined based on the **earliest** of the scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of the reviewer who read baseline first if there is no adjudication for the BICR data.
- For investigational site assessments, date of progression will be determined based on the **earliest** of the RECIST assessment/scan dates of the component that triggered the progression.
- For both BICR and investigational site assessments, when censoring a patient for PFS the patient will be censored at the **latest** of the RECIST assessment/scan 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.

The primary analysis will be derived based on investigator assessment of the radiological scans.

A sensitivity analysis based on the blinded independent central review (BICR) of the radiological scans from the non-*tBRCAm* cohort patients will be carried out.

3.2.2 Overall survival (OS)

For the non-*tBRCA* cohort, overall survival (OS) is defined as the time from randomisation to death due to any cause regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy. That is;

OS = (date of death or censoring – date of randomisation + 1).

For the *tBRCA* cohort, OS is defined as the time from allocation to the *tBRCAm* cohort (see section 2.1 for definition) until death due to any cause regardless of whether the patient withdraws from study therapy or receives another anti-cancer therapy. That is;

OS = (date of death or censoring – date of cohort allocation + 1).

Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Note: Survival calls will be made in the week following the date of each DCO, and if patients are confirmed to be alive or if the death date is post the DCO date, these patients will be censored at the date of DCO.

The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the OS analyses should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly-available death registries. In the event that the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note that for any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment. The last date for each individual patient is defined as the latest among the following dates recorded on the case report forms (CRFs):

- AE start and stop dates
- Admission and discharge dates of hospitalization
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST CRF
- Start and stop dates of subsequent anticancer treatment

- Date last known alive on survival status CRF
- End of study date

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

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

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

3.2.3 Objective response rate (ORR)

Objective response rate (ORR) is defined similarly for the non-*tBRCAm* and *tBRCAm* cohorts as the number (percentage) of patients with at least one investigator-assessed visit response of CR or PR and will be based on a subset of all randomised/allocated patients who have evaluable disease at baseline per the site investigator.

Data obtained up until progression, or the last RECIST assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue randomised/allocated 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.

Categorisation of objective tumour response assessment will be based on the modified RECIST v1.1 criteria of response (see section 3.1). Target lesion (TL) progression will be calculated in comparison to when the tumour burden was at a minimum (i.e., smallest sum of diameters previously recorded on study). For patients with non-measurable disease only at baseline, categorisation of objective tumour response assessment will be based on the modified RECIST v1.1 criteria of response: CR, PD and Non-CR/Non PD. Patients with no disease at baseline will be assessed according to modified RECIST v1.1 criteria with responses of NED or PD.

ORR will be assessed;

- In all patients with evaluable disease at baseline per the site investigator across the entire period of the study (ie, chemotherapy phase and maintenance phase) and may be produced for the two phases separately
- Prior to surgery in those patients planned to have IDS with evaluable disease at baseline. For example, calculate ORR where only responses prior to curative surgery are included in the numerator.

3.2.4 Best objective response (BoR)

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in section 3.1. It is the best response a patient has had following randomisation, but prior to starting any subsequent cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorisation of BoR will be based on RECIST using the following response categories: CR, PR, SD, NED (applies only to those patients entering the study with no disease at baseline), PD and NE.

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 15 weeks minus 1 week, i.e. at least 98 days (to allow for an early assessment within the assessment window), after randomisation. 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 from the overall visit response using all investigator data up until the first progression event. It will also be determined programmatically based on RECIST using all site investigator data up until the first progression event. The denominators for each case will be consistent with those used in the ORR analysis.

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

For patients who die with no evaluable RECIST assessments, if the death occurs ≤29 weeks (i.e. 27 weeks + 2 weeks to allow for a late assessment within the assessment window) after randomisation, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs > 29 weeks after randomisation then BoR will be assigned to the NE category.

A patient will be classified as a responder if the RECIST criteria for a CR or PR are satisfied at any time following randomisation, prior to RECIST progression and prior to starting any subsequent cancer therapy.

3.2.5 Time from cohort allocation / randomisation to second progression (PFS2)

For the non-*tBRCAm* cohort, time from randomisation to second progression or death (PFS2) is defined as the time from the date of randomisation to the earliest of the progression event subsequent to first subsequent therapy or death. The date of second progression will be recorded by the investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: investigator assessment of radiological progression, cancer antigen 125 (CA125) progression or symptomatic progression or death.

For the *tBRCAm* cohort, PFS2 is defined as the time from allocation to the *tBRCAm* cohort to the earliest of the progression event subsequent to first subsequent therapy or death. The date

of second progression will be recorded by the investigator in the eCRF and defined according to local standard clinical practice and may involve any of the following: investigator assessment of radiological progression, cancer antigen 125 (CA125) progression or symptomatic progression or death.

The PFS2 event after the primary endpoint PFS event will be reviewed every 12 weeks and the event status recorded till the determination of PFS2 or death. Patients alive and for whom a second disease progression has not been observed should be censored at date last known alive and without a second disease progression (i.e., censored at the latest of the PFS or PFS2 assessment date if the patient has not had a second progression or death).

3.2.6 Duration of response (DoR)

Duration of response (DoR) is defined similarly for the non-*tBRCAm* and *tBRCAm* cohorts using the corresponding FAS among patients with a response (CR or PR), as the time from the date of first documented response (i.e., the first time at which the visit response is PR or CR) according to modified RECIST v1.1 as assessed by the investigator until date of documented progression or death in the absence of disease progression. That is;

DoR = (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 PR or CR.

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

3.2.7 Time to first subsequent therapy or death (TFST)

For the non-*tBRCAm* cohort, time to start of first subsequent therapy or death (TFST) is defined as the time from randomisation to the earlier of first subsequent therapy start date following study treatment discontinuation, or death. That is;

TFST = (date of first subsequent cancer therapy/death/censoring – date of randomisation + 1).

For the *tBRCAm* cohort, TFST is defined as the time from the allocation to the *tBRCAm* cohort to the earlier of first subsequent therapy start date following study treatment discontinuation, or death. That is;

TFST = (date of first subsequent cancer therapy/death/censoring – date of cohort allocation + 1).

Subsequent therapies will be reviewed to assess which represent clinically important treatments intended to control ovarian cancer. Any patient not known to have had a first subsequent anti-cancer therapy or died at the time of the analysis will be censored at the last date that the patient was known to have not received subsequent anti-cancer therapy, i.e. the last of the follow-up visits where this was confirmed (obtained from the TTSCAPRX form).

If a patient terminated the study for reason other than death before first subsequent therapy, these patients will be censored for TFST evaluation at the earliest of their last known to be alive and termination dates.

Patients who did not receive any of the study treatments, the first alternative cancer therapy they receive will be the initial therapy. In this situation, TFST will be calculated as time from the allocation to the *tBRCAm* cohort, or randomisation to the non-*tBRCAm* cohort to the start of the initial alternative therapy or death.

For the non-*tBRCAm* cohort,

TFST = (date of start of first alternative cancer therapy/death/censoring – date of randomisation + 1).

For the *tBRCAm* cohort,

TFST = (date of start of first alternative cancer therapy/death/censoring – date of cohort allocation + 1).

3.2.8 Time to second subsequent therapy or death (TSST)

For the non-*tBRCAm* cohort, time to start of second subsequent therapy or death (TSST) is defined as the time from randomisation to the earlier of the second subsequent anti-cancer therapy start date following study treatment discontinuation, or death. That is;

TSST = (date of start of second subsequent cancer therapy/death/censoring – date of randomisation + 1).

For the *tBRCAm* cohort, time to start of second subsequent therapy or death (TSST) is defined as the date of cohort allocation to the earlier of the second subsequent anti-cancer therapy start date following study treatment discontinuation, or death. That is;

TSST = (date of start of second subsequent cancer therapy/death/censoring – date of cohort allocation + 1).

Any patient not known to have had a second subsequent anti-cancer therapy or died at the time of the analysis will be censored at the last date that the patient was known to have not received a second subsequent anti-cancer therapy, i.e. the last of the follow-up visits where this was confirmed (obtained from the TTSCAPRX form).

If a patient terminated the study for reason other than death before a second subsequent therapy, these patients will be censored for the TSST evaluation at the earliest of their last known to be alive and termination dates.

Patients who did not receive any of the study treatments, the first alternative cancer therapy they receive will be the initial therapy such that the second anti-cancer therapy would be the second alternative anti-cancer therapy. TSST will then be calculated as time from date of allocation to the *tBRCAm* cohort, or randomisation to the non-*tBRCAm* cohort to the start of the second alternative anti-cancer therapy or death.

That is;

For the non-*tBRCAm* cohort,

TSST = (date of start of second alternative cancer therapy/death/censoring – date of randomisation + 1).

For the *tBRCAm* cohort,

TSST = (date of start of second alternative cancer therapy/death/censoring – date of cohort allocation + 1).

3.2.9 Time to study treatment discontinuation or death (TDT)

For the non-*tBRCAm* cohort, time to permanent study treatment discontinuation or death (TDT) is defined as the time from randomisation to the earlier of the date of permanent study treatment discontinuation or death. That is;

TDT = (date of study treatment discontinuation/death/censoring – date of randomisation + 1)

For the *tBRCAm* cohort, time to permanent study treatment discontinuation or death (TDT) is defined as the time from allocation to the *tBRCAm* cohort to the earlier of the date of permanent study treatment discontinuation or death. That is;

TDT = (date of study treatment discontinuation/death/censoring – date of cohort allocation + 1)

Any patient not known to have died at the time of analysis and not known to have discontinued study treatment will be censored based on the last recorded date on which the patient was known to be alive.

Note, for the purposes of the analysis of TDT, study treatment is defined as investigational treatment: durvalumab/placebo or olaparib/placebo. Treatment discontinuation is defined as all investigational products being discontinued. For example:

- If a patient receives both durvalumab/placebo and olaparib/placebo the date of treatment discontinuation will be the later of the date of discontinuation of each drug respectively.
- If a patient only receives one of the two investigational products then treatment discontinuation will be the date of discontinuation for the investigational product administered.
- If a patient does not receive any investigational product then they will be censored at day 1

3.2.10 Pathological complete response (pCR) in IDS patients only

Pathological complete response (pCR) is defined as the absence of any residual macroscopic disease together with the absence of residual microscopic disease post-surgery. The pathological complete response (pCR) rate is defined similarly for the non-*tBRCAm* and

tBRCAm cohorts as the number (percentage) of patients with pCR and will be based on a subset of all randomised/allocated patients who surgery status at study entry is planned IDS and the surgery is performed.

3.3 Patient reported outcome variables

Patient-reported outcomes will be assessed using;

- European Organisation for Research and Treatment of Cancer quality of life core questionnaire (EORTC QLQ-C30 version 3.0) and the ovarian cancer disease-specific module (EORTC QLQ-OV28).
- Patient reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE).
- Patient global impression of severity of cancer symptoms (PGIS)
- EuroQoL five dimensions, five level (EQ-5D-5L) health state utility index

3.3.1 EORTC QLQ-C30 and EORTC QLQ-OV28

The EORTC QLQ-C30 and EORTC QLQ-OV28 will be assessed throughout the study at timepoints defined in the CSP.

3.3.1.1 EORTC QLQ-C30

The EORTC QLQ-C30 (Aaronson et al 1993) is a questionnaire developed to assess cancer patients' measurement of global aspects of health-related quality of life. The QLQ-C30 incorporates eight multi-item scales: five functional scales (physical, role, cognitive, emotional, and social); three symptom scales (fatigue, pain, nausea/ vomiting); and a 2-item global health related quality-of-life scale (global health status/QoL). Five single-item symptom measures are also included providing measurements on additional symptoms commonly reported by cancer patients (dyspnoea, loss of appetite, insomnia, constipation, diarrhoea) and 1 item on the financial impact of the disease. See scoring algorithm for the EORTC-QLQ-C30 in Appendix 9.1. A clinically meaningful change on EORTC QLQ-C30 subscales/scores is defined as an absolute change in score of ≥10 points (Osoba et al 1998).

3.3.1.2 EORTC QLQ-OV28

The EORTC QLQ-OV28 supplements the EORTC quality of life questionnaire-core 30 (QLQ-C30) to assess the health-related quality of life of women with ovarian cancer treated in clinical trials. It is meant for use among ovarian cancer patients varying in disease stage and treatment modality (i.e. surgery, chemotherapy, radiotherapy, etc.). The 28-item module, OV28, assesses three functional scales (body image, sexuality, attitude to disease/treatment) and five symptom scales (abdominal/GI symptoms, peripheral neuropathy, hormonal/menopausal symptoms, other chemotherapy side-effects, hair loss). See scoring algorithm for the EORTC-QLQ-OV28 in Appendix 9.2. A clinically meaningful change on EORTC QLQ-OV28 subscales is defined as an absolute change in score from baseline of ≥10 points (Osoba et al 1998).

3.3.2 PRO-CTCAE

The PRO version of the CTCAE is an item library of symptomatic AEs experienced by patients while undergoing treatment of their cancer. It was developed to measure the relevant attributes of a symptom directly from patients for the purposes of understanding symptomatic toxicity from the patients' perspective.

For this study, 9 items are considered relevant for assessment, (i.e., nausea, vomiting, diarrhoea, decreased appetite, abdominal pain, itching, muscle pain, fatigue and chills).

The PRO-CTCAE will only be administered in those countries where a linguistically validated version exists. PRO-CTCAE will be assessed throughout the study at timepoints defined in the CSP. See scoring algorithm for the PRO-CTCAE in Appendix 9.3.

3.3.3 Patient global impression of severity of cancer symptoms (PGIS)

The PGIS is included to assess how a patient perceives her overall current severity of cancer symptoms. This is a single item with five response options: “no symptoms”, “very mild”, “mild”, “severe” and “very severe”.

PGIS will be assessed throughout the study at timepoints defined in the CSP.

3.3.4 EQ 5D-5L

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 consisting of the EQ 5D-5L descriptive system and the EQ Visual Analogue scale (EQ VAS).

EQ 5D-5L will be assessed throughout the study at timepoints defined in the CSP.

EQ-5D-5L descriptive system

The EQ-5D-5L descriptive system questionnaire comprises of 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). For each dimension, respondents select which statement best describes their health on that day from a possible five options of increasing levels of severity (1 = “No problems”; 2 = “Slight problems”; 3 = “Moderate problems”; 4 = “Severe problems”; 5 = “Unable to/ extreme problems”). This results in a 1-digit number expressing the level selected for that dimension. A unique health state, termed the EQ-5D-5L profile, is reported as a five-digit code with a possible 3,125 health states from responses to the 5 dimensions. For example, state 11111 indicates no problems on any of the five dimensions, while state 12345 indicates no problems with mobility, slight problems with washing or dressing, moderate problems with doing usual activities, severe pain or discomfort and extreme anxiety or depression.

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. The index values facilitate the calculation of quality-adjusted life years (QALYs) that are used to inform economic evaluations of health care

interventions. The country specific weighting applied in the derivation of the index, 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.

EQ VAS

Respondents also assess their health today using the EQ VAS. The EQ VAS records the respondent's self-rated health status on a visual analogue scale with endpoints labelled 'the best health you can imagine' corresponding to higher values on the VAS score and 'the worst health you can imagine' corresponding to lower values on the VAS score. This information can be used as a quantitative measure of health as judged by the individual respondents.

3.3.5 PRO compliance

Summary measures of overall compliance and compliance over time will be derived for the following PRO instruments used in this study: EORTC-QLQ-C30, EORTC-QLQ-OC-28, and EQ-5D-5L.

These will be based upon the following definitions:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.
- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For all patients, up to a maximum of 36 months post Visit 2 (Cycle 1 Day 1), will be used to assess whether the patient is still under PRO follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.
- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that can be determined.
- Compliance rate (visit): Compliance rate for each specific visit (including baseline) is defined as: Total number of patients with an evaluable questionnaire for the visit, divided by total number of patients with questionnaires expected for the visit multiplied by 100, i.e. (Evaluable ÷ Expected) * 100.
- Completion rate (visit): Completion rate for each specific visit (including baseline) is defined as: Total number of patients with an evaluable questionnaire for the visit divided by total number of randomised/allocated patients multiplied by 100, i.e. (Evaluable ÷ Number of randomised/allocated patients) * 100.
- Evaluable rate (visit): Evaluable rate for each specific visit (including baseline) is defined as: Total number of patients with an evaluable questionnaire for the visit divided by total number of patients with an received questionnaire for the visit multiplied by 100, i.e. (Evaluable ÷ Received) * 100.
- Overall patient compliance rate is defined as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined

above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by 100.

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 patient 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 hospitalisation. That is;

Length of hospital stay = (end date of hospitalisation – start date of hospitalisation + 1).

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

3.5 Pharmacokinetic/ADA variables (in non-*tBRCAm* patients only)

Blood samples for determination of durvalumab concentrations in serum and olaparib concentration in plasma will be taken in a subset of approximately 100 non-*tBRCAm* patients (approximately 33 evaluable patients per arm) who have already had upfront primary cytoreductive surgery at those sites that are able to take PK assessment samples (see Section 8.5 of CSP for details on sampling dates and times). Anti-drug antibodies (ADA) samples for durvalumab will only be taken in this subset of non-t*BRCA*m patients.

PK variables

Samples for determination of drug concentrations of durvalumab in serum will be analysed by contract laboratories on behalf of AstraZeneca, using a validated bioanalytical method. Full details of this analytical method will be provided in a separate bioanalytical report. Placebo samples will not be analysed.

Also, samples for determination of drug concentrations of olaparib in plasma will be analysed by central laboratories on behalf of AstraZeneca, using an appropriate bioanalytical method. Full details of this analytical method will be described in a separate bioanalytical report. Placebo samples will not be analysed.

Plasma and serum concentrations below the lower limit of quantification (LLOQ) will be handled as follows:

- At a time point where 50% or less of the values are below the LLOQ (BLQ), all BLQ values will be set to LLOQ, and all descriptive statistics will be calculated.
- At a time point where more than half of the values are BLQ, the mean, SD, geometric mean and CV% will be reported as 'Not Determined'. The maximum value will be reported from the individual data, and the minimum and median will be reported as BLQ.
- The number of BLQ values (n below LLOQ) will be reported for each time point.

The PK concentrations will be reported to the same precision as the source data. For descriptive statistics of concentrations, non-quantifiable (NQ) values of plasma concentrations will be handled as follows:

- If, at a given time point, 50% or less of the plasma concentrations are NQ, the mean, SD, geometric mean, %CV, and %GCV will be calculated by substituting the limit of quantification (LOQ) for values which are NQ.
- If more than 50%, but not all, of the concentrations are NQ, the mean, geometric mean, SD, %CV, and %GCV will be reported as not calculable (NC).
- If all the concentrations are NQ, the geometric mean and mean will be reported as NQ and the SD, %CV, and %GCV as NC.

Immunogenicity

Immunogenicity results will be analysed descriptively by summarizing the number and percentage of patients who develop detectable anti-drug antibodies (ADA) and/or anti-drug neutralising antibody (nAB) for durvalumab.

Immunogenicity Variables

Serum samples for ADA (anti-drug antibody) assessments will be conducted utilizing a tiered approach (screen, confirm, titer), and ADA data will be collected at scheduled visits shown in the CSP. ADA results from each sample will be reported as either positive or negative. Positive negative cut points that were previously statistically determined from drug-naïve samples will be employed to determine ADA positive or negative status. If the sample is positive, the ADA titer will be reported as well. In addition, the presence of neutralizing antibody (nAb) will be tested for all ADA-positive samples using a ligand binding assay. The nAb results will be reported as positive or negative.

The baseline ADA result is defined as the reported result of the pre-dose sample. ADA evaluable subjects are patients in the safety analysis set who received at least 1 dose of durvalumab and have non-missing baseline ADA result and at least 1 post-baseline ADA. The number of patients in the ADA analysis set (defined in Section 2.1) who fulfil the following the criteria will be determined. The percentage of ADA-positive patients in each of the category will be calculated, using the number of patients in the ADA analysis set of the treatment group as the denominator.

- ADA positive at any timepoint including baseline and all post-baseline. The percentage of ADA positive patients in the ADA evaluable population is known as ADA prevalence.

- ADA incidence (treatment-emergent ADA), defined as the percentage of ADA evaluable population with either treatment-induced or treatment-boosted ADA.
- 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 within at least 16 weeks (112 days) between the first and last positive measurement or an ADA positive result at the last available assessment. The category may include patients 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 patients meeting these criteria who are ADA positive at baseline.
- nAb positive at any visit, defined as the proportion having neutralizing ADA at any point in time.

3.6 Safety

Safety and tolerability will be assessed in terms of AEs (including SAEs), deaths, laboratory data, vital signs, ECG and exposure. These will be collected for all patients.

3.6.1 Adverse events

Serious adverse events (SAEs) from time of signature of the main screening ICF throughout the treatment period and including the safety follow-up periods (See Section 1.1.3.2 of CSP) corresponding to;

- 30 days from the last dose of olaparib/placebo treatment (the follow-up period for olaparib)
- 90 days from the last dose of durvalumab/placebo (the follow-up period for durvalumab)

All other AEs will be collected from Day 1 of Cycle 1 throughout the treatment period and including the follow-up periods.

An adverse events will be defined as treatment emergent if it starts, or worsens (by investigator report of a change in intensity) at or after the start of the first dose of any of the investigational study treatments, including durvalumab/placebo or olaparib/placebo, throughout the treatment period and including the safety follow-up periods until the later date of 30 days after the last dose of olaparib/placebo and 90 days after last dose of durvalumab/placebo. The latest version of Medical Dictionary for Regulatory Activities (MedDRA) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE version 5 or later version if available).

The Investigator will assess causal relationship between study treatments and each AE, and answer 'yes' or 'no' to the question '*Do you consider that there is a reasonable possibility that the event may have been caused by the study treatments?*'

For SAEs, causal relationship will also be assessed for other medications, including all study medications, as well as any study procedures. Note that for SAEs that could be associated with any study procedure the causal relationship is implied as 'yes'.

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 OAEs and reported as such in the Clinical Study Report.

A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs. Examples of these are marked haematological and other laboratory abnormalities, and certain events that lead to intervention (other than those already classified as serious), dose reduction or significant additional treatment.

Olaparib AEs of special interest

Olaparib adverse events of special interest (AESI) are events of scientific and medical interest specific to the further understanding of the safety profile of olaparib and may require close monitoring and rapid communication by the investigators to AstraZeneca. An AESI may be serious or non-serious.

These AESIs are identified as a list of categories provided by the patient safety team. Other categories may be added as necessary or existing terms may be merged.

An AstraZeneca medically qualified expert, after consultation with the Global Patient Safety Physician, has reviewed the AEs of interest and identified which preferred terms contribute to each AESI. A further review will take place prior to DBL to ensure any further terms not already included are captured within the categories.

Durvalumab AEs of special interest and AEs of possible interest

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.

AESI are defined as AEs that include, but are not limited to, events with a potential inflammatory or immune mediated mechanism that may require 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).

The AEPIs reported in the AstraZeneca-sponsored durvalumab studies 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.

The AESIs and AEPIs have been categorized into the following AESI/AEPI categories: Pneumonitis, Hepatic events, Diarrhoea/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 preferred terms contribute to each AESI/AEPI. A further review will take place prior to Database lock (DBL) to ensure any further terms not already included are captured within the categories.

Immune-mediated Adverse Events (imAE)

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

Further details are provided in a separate imAE charter. In addition, the Sponsor may perform medical review of those AESIs and classify them as imAEs or not imAEs via an independent manual adjudication process.

3.6.2 Exposure and dose interruptions

Exposure (days) will be defined as follows:

For olaparib/placebo:
Total (or intended) exposure (days) of olaparib/placebo

- Total (or intended) exposure = earliest (last dose date of olaparib/placebo > 0 [mg], date of death, date of DCO) – (first dose date of olaparib/placebo) +1)

Actual exposure of olaparib/placebo

- Actual exposure = intended exposure – total duration of dose interruptions
- Where intended exposure will be calculated as above, and a dose interruption is defined as any length of time where the patient has not taken any of the planned dose.

To calculate actual exposure, dose interruptions will include those where a patient forgot to take a dose. No adjustment however will be made for any dose reductions that may have occurred.

Patients undergoing IDS in the period will have their total/actual exposure calculated separately for cycles before and after IDS and added together.

For durvalumab/placebo:

Total (or intended) exposure (days) of durvalumab/placebo

- Total (or intended) exposure = earliest (last dose date of durvalumab/placebo where dose > 0 [mg] + 20 days, date of death, date of DCO) – (first dose date of durvalumab/placebo) +1)

Actual exposure of durvalumab/placebo

- Actual exposure = intended exposure – total duration of dose delays
- Where intended exposure will be calculated as above. Calculation of duration of dose delays (for actual exposure) is:
  For all dosing dates: Total duration of dose delays (= Sum of (Date of the dose – Date of previous dose – 20 days)). Thus, if no delays were encountered, the duration would sum up to 0, since infusions were done every 21 days.

Dose reductions of durvalumab are not permitted and the calculation of actual exposure makes no adjustment for any dose reductions that may have occurred.

Patients undergoing IDS in the period will have their total/actual exposure calculated separately for cycles before and after IDS and added together.

For chemotherapy agents (paclitaxel and carboplatin) and bevacizumab:

Total (or intended) and actual exposure for paclitaxel (or paclitaxel substitute alternative treatment), carboplatin (or carboplatin substitute alternative treatment), and bevacizumab (or biosimilar) will be derived using similar calculations as for durvalumab/placebo and derived from first dose of each respective agent.

Number of treatment cycles received

Exposure of chemotherapy agents, bevacizumab, and durvalumab/placebo will also be measured by the number of cycles received. A cycle corresponds to a period of 21 days. If a

cycle is prolonged due to toxicity, this should still be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

Missed or forgotten doses

Missed and forgotten doses should be reported for patients who receive olaparib/placebo. Missing or forgotten doses should be recorded on the Exposure (Olaparib) module as a dose interruption with the reason recorded as "Subject forgot to take dose". These missed or forgotten doses will not be included as dose interruptions in the summary tables, but the information will appear in the listing for dosing. However, these missed and forgotten doses will be considered in the derivation of actual exposure.

Patients who permanently discontinue during a dose interruption or delay

If a patient permanently discontinues study treatment during a dose interruption or delay, then the date of last administration of study medication recorded on DOSDISC will be used in the programming as the last dose date. The dose interruption that happens immediately before patient permanently discontinue the study treatment will not be included in the summary table as interruption (i.e. an interruption will only be included if the drug was restarted).

Compliance

Patients who receive olaparib/placebo will be assessed for treatment compliance. Percentage compliance will be defined as:

> {(No. tablets dispensed in period – no. tablets returned from period)/ (no. days of study drug exposure in period × expected tablets per day)} × 100%

Where expected tablets per day will take into account once daily or bid dosing. Overall compliance may be calculated over various periods if the dose has been modified, to consider the differing expected tablets per day or the protocol-specified dose interruptions. As in the determination of exposure, missed doses will not be adjusted for.

Before calculation of summary measures, individual compliance rates will be capped at 100%, however uncapped compliance rates will be listed by subject.

Percentage compliance will not be assessed for durvalumab/placebo.

3.6.3 Dose intensity

Olaparib/placebo

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

- RDI = 100% × d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing.

d is determined by adding up the received dose for each day from the date of first olaparib/placebo dose to the date of last olaparib/placebo dose. D = 300 mg x 2 x total (intended) exposure

Durvalumab/placebo

In deriving dose intensity parameters for durvalumab/placebo, these additional considerations should be made;

- Dosing with durvalumab/placebo will have these apply;
  - Dose reductions are not permitted.
  - Dose interruptions and dose delays may occur in which case a dose cycle may be prolonged, but this cycle should be still counted as one cycle. A cycle is counted if treatment is started even if the full dose is not delivered.
  - If a decision is made to permanently discontinue study treatment in-between cycles or during a cycle delay, then the date of the last administration of durvalumab/placebo recorded will be used in the determination of dose intensity
- When deriving actual dose administered the volume before and after infusion will also be considered.

Relative dose intensity (RDI) then 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 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.

3.6.4 Laboratory data

Parameters related to laboratory data including whole blood haematology/haemostatis, serum and plasma clinical chemistry, urinalysis (see CSP Table 13) will be assessed based on samples that will be collected throughout the study, from screening to follow-up visit. Additionally, assessments on coagulation, bone marrow analysis, CA-125 tumour marker, and pregnancy test by serum or urine sample will be determined.

For derivation of post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 3.6.8 below will be used.

As applicable, values will be converted to standard units and will be graded using CTCAE version 5.0. Maximum post-baseline CTCAE grade will also be calculated, as required.

3.6.5 ECG

Resting 12-lead ECGs will be performed within 28 days prior to starting study treatment and

when clinically indicated. Measurements should be taken after the subject has been rested in a supine position for at least 5 minutes. All ECGs will be assessed locally to determine whether they are clinically significantly abnormal / not clinically significantly abnormal. If there is a clinically significant abnormal finding, it will be recorded as an AE by the Investigator if applicable (see CSP Sections 8.3.6 and 8.3.7)

ECG data obtained up until the 30 days from date of last dose of olaparib/placebo treatment or 90 days from last infusion of durvalumab/placebo will be used for reporting. ECG changes from baseline will be determined as a safety evaluation.

3.6.6 Vital signs

Vital signs (BP, pulse and temperature) data obtained up until the end of safety follow-up (30 days post last dose for subjects receiving olaparib/placebo or 90 days for subjects receiving durvalumab/placebo whichever is later). For derivation of post baseline visit values considering visit window and to handle multiple records, derivation rules as described in Section 3.6.8 below will be used. Any clinically significant changes in vital signs should be recorded as an AE, if applicable (see CSP Sections 8.3.6 and 8.3.7).

3.6.7 Physical examination

Physical examination assessments including assessments of the following: general appearance, respiratory, cardiovascular, skin, head and neck (including ears, eyes, nose and throat), lymph nodes, thyroid, abdomen, musculo-skeletal (including spine and extremities) and neurological systems will be performed at screening and as specified according to the study schedule (see CSP Table 2, Table 3 and Table 4). Any clinically significant changes in physical examination should be recorded as an AE, if applicable (see CSP Sections 8.3.6 and 8.3.7).

3.6.8 General considerations for safety and PRO assessments

Time windows will need to be defined for any presentations that summarise values by visit. The following conventions should apply:

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the time window should be based on the actual date and not the intended date of the visit.
- All unscheduled visit data should have the potential to be included in the summaries.
- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day. For example, the visit windows for vital signs data (with 3 weeks between scheduled assessments up to week 108 from first randomisation/allocation (Cycle 37)) are:

  Day 22, visit window 2 –32
  Day 43, visit window 33 – 53
  Day 64, visit window 54 – 74
  Day 85, visit window 75 – 95

Day 106, visit window 96 – 116
Day 127, visit window 117 – 140
…

- In addition, for safety assessments only, the safety follow-up visit is defined as the latest of either 30 days following discontinuation of olaparib/placebo or 90 days following discontinuation of durvalumab/placebo, or until the initiation of the first subsequent therapy (including non-palliative radiotherapy), whichever occurs first.
- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).
- Listings should display all values contributing to a time point for a patient.
- For visit based summaries:
    - If there is more than one value per patient within a time window then the closest value should be summarised, or the earlier in the event the values are equidistant from the nominal visit date. The listings should highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all post baseline values collected are used including those collected at unscheduled visits regardless of whether or not the value is closest to the scheduled visit date
    - To prevent very large tables or plots being produced that contain many cells with minimal data, for each treatment group visit data should only be summarised if the number of observations is greater than the minimum of 20 and > 1/3 of patients dosed.
- For summaries at a patient level, all values should be included, regardless of whether they appear in a corresponding visit-based summary, when deriving a patient level statistic such as a maximum.
- Baseline for safety assessments will generally be defined as the last non-missing measurement prior to first dose with randomised/allocated study treatment (olaparib/durvalumab/placebo). For laboratory data and vital signs data, any assessments made on Cycle 2 Day 1 will be considered pre-dose. Where safety data are summarised over time, study day will be calculated in relation to date of first dose of randomised/allocated treatment (olaparib/durvalumab/placebo).

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

3.7 Exploratory Variables

3.7.1 Relapse free survival

Relapse-free survival (RFS) also known as disease-free survival (DFS) is defined as the time from randomization/cohort allocation until first recurrence disease (progression) according to modified RECIST v1.1 (per investigator assessment) or death due to any cause, whichever was observed first.

The population for this analysis will be patients who are in the FAS and were assessed as having no disease at baseline (as captured in the baseline RECIST eCRF) and also will be repeated in those who had NED according to modified RECIST v 1.1 (per investigator assessment) according to the last visit at end of the chemotherapy phase of the study.

Patients who have no evidence of recurrent disease/not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last RECIST v1.1 assessment. However, if the patient has recurrent disease/progresses or dies after two or more consecutive missed RECIST visits, the patient will be censored at the time of the latest RECIST v1.1 assessment prior to the two missed visits. If the patient has no RECIST visits or does not have baseline data, they will be censored at day 1 unless they die within 2 visits of randomisation/cohort allocation.

The derivation will utilise the same rules as those adopted for the primary PFS endpoint. In addition, further analysis may be performed in patients who were assessed as CR according to modified RECIST v1.1 (per investigated assessment) according to the last visit at end of the chemotherapy phase of the study.

3.7.2 Proportion of patients with NED at landmark time points

The proportion of patients with NED at landmark time points will be defined as the number of patients assessed as NED according to the RECIST v1.1 assessment (per investigator assessment) at the visit corresponding to the time point from randomization/allocation, as a proportion of all patients in the FAS in each treatment arm who were assessed as having no disease at baseline (as captured in the baseline RECIST eCRF).

The landmark times to be assessed for are 15, 24, and 48 months and will be summarised (using estimates from the relapse free survival KM curve) and presented by treatment group.

3.7.3 Biomarker endpoints

In addition to the pre-defined biomarker subgroups, the presence and status of other exploratory biomarkers may be assessed and summarized;

- Mutation status of HRR genes in the tumour (BRCA1, BRCA2, ATM, BRIP1, PALB2, RAD51C, BARD1, CDK12, CHEK1, CHEK2, FANCL, RAD51B, RAD51D, RAD54L)
  - Myriad Genomic Instability Score (GIS)
  - Myriad tumour mutational score
- CCI [redacted]
- CCI [redacted].
- CCI [redacted] levels and response to treatment

These endpoints may be reported separately to the clinical study report.

4 ANALYSIS METHODS

The primary objective of the study is to compare PFS (per RECIST v1.1 as assessed by investigator) for Arm 3 vs Arm 1 in non-*tBRCAm* HRD-positive population and in the non-*tBRCAm* ITT population. A key secondary comparison is to compare PFS (per RECIST 1.1 as assessed by investigator) for Arm 2 vs Arm 1 in the non-*tBRCAm* ITT population.

Formal statistical analyses will be performed to test the following corresponding hypotheses of interest for this study in the non-*tBRCAm* cohort for each population of interest as follows;

H_{03}: Arm 3 = Arm 1 vs H_{13}: Arm 3 ^= Arm 1
and
H_{02}: Arm 2 = Arm 1 vs H_{12}: Arm 2 ^= Arm 1

Where H0 = the null hypothesis; H1 = the alternate hypothesis and:

- Arm 1 = bevacizumab in combination with platinum-based chemotherapy
- Arm 2 = bevacizumab in combination with platinum-based chemotherapy + durvalumab
- Arm 3 = bevacizumab in combination with platinum-based chemotherapy + durvalumab + olaparib.

In the non-*tBRCAm* HRD-positive population an exploratory analysis comparing Arm 2 vs Arm 1 will also be undertaken, as required. In addition, an exploratory analysis comparing Arm 3 vs Arm 2 will also be undertaken in the non-*tBRCAm* HRD-positive population and in the non-*tBRCAm* ITT population, as required.

Pooling strategy

For the analysis of primary endpoint of PFS, the following pooling strategy will be applied across all three arms by population (ITT population and HRD-positive population). If the number of events in the individual stratum are too small for a meaningful analysis (less than 5 events per stratum; a stratum is defined as strata1*strata2*...strataX*treatment; so with 2 stratification factors with 2 and 3 levels respectively and two treatments per comparison we have 2*3*2=12 stratum) stratification factors which will be removed in the following order until there are at least 5 events in each stratum: Geographic region (North America/Europe/RoW), Timing and outcome of cytoreductive surgery (no macroscopic residual disease after upfront primary surgery vs all others). All analyses will then be conducted with the same stratification factors as the primary analysis of PFS. If there are secondary endpoints or sensitivity analyses that still do not conform to the 5 event rule per stratum, an unstratified analysis will be conducted. This will be supported by an unstratified analysis of the primary endpoint.

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

Timing of efficacy analysis

The timing of efficacy analyses is summarised in Table 6 and is described below.

Table 6 Timing of efficacy analyses

Analysis	Timepoint
PFS interim analysis*	Approx. 128 PFS events across Arm 3 and Arm 1 in the non-*tBRCAm* HRD-positive population and approx. 390 PFS events across Arm 3 and Arm 1 in the non-*tBRCAm* ITT population.
Primary PFS/First interim OS analysis**	Approx. 149 PFS events across Arm 3 and Arm 1 in the non-*tBRCAm* HRD-positive population and approx. 453 PFS events across Arm 3 and Arm 1 in the non-*tBRCAm* ITT population.
Final OS analysis	Approx 50% maturity have occurred across the 3 non-*tBRCAm* treatment arms or 5 years after the last non-*tBRCAm* patient is randomised, whichever occurs sooner.

* At this time approximately 414 PFS events are anticipated across Arm 2 and Arm 1 in the non-tBRCAm ITT population.

** At this time approximately 480 PFS events are anticipated across Arm 2 and Arm 1 in the non-tBRCAm ITT population.

No further analyses of PFS are planned beyond the primary PFS analysis unless requested by Health Authorities. Other analyses planned at the same time include PFS2, TFST, TSST and TDT analysis which will use the same methodology and model. At the time of the primary analysis of PFS, an interim analysis of OS will also be performed.

At the time of the DCO and unblinding for the PFS analysis in the non-*tBRCAm* cohort, the data collected from the *tBRCAm* cohort will be appropriately summarised.

4.1 General principles

Efficacy and PRO data will be summarised and analysed on the FAS analysis set. Safety and treatment exposure data will be summarised based upon the safety population. Study population and demography data will be summarised based upon the FAS analysis set.

Data will be presented in data listings by treatment group and subject number. All summaries will be presented by treatment group, unless otherwise specified. A month is operationally defined to be 30.4375 days.

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

- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarised 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.

- 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.
- Categorical variables will be summarised by frequency counts and percentages for each category. Unless otherwise stated, percentages will be calculated out of the population total and for each treatment group. Percentages will be reported by values rounded to 1 decimal place.
- SAS® version 9.4 (as a minimum) will be used for all analyses.
- Where analysis models are stratified by the randomisation stratification factors; timing and outcome of cytoreductive surgery: no residual macroscopic disease after primary surgery vs all others (macroscopic residual disease after upfront primary surgery OR planned IDS), and Geographic region: North America vs Europe vs RoW (or the stratification factors are included in the model as covariates), the strata obtained from the randomisation code will be used, not the values recorded in the electronic case report form (eCRF).
- In general, for efficacy and PRO endpoints the last observed measurement prior to Cycle 1 Day 1 will be considered the baseline measurement. However, if an evaluable assessment is only available after Cycle 1 Day 1 but before the first dose of randomized/allocated treatment (durvalumab/olaparib/placebo) then this assessment will be used as baseline.
- For safety endpoints the last observation before the first dose of randomised/allocated investigational treatment will be considered the baseline measurement unless otherwise specified.
- Assessments on the day of the first dose of randomised/allocated treatment 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

$$((post\ baseline\ value\ -\ baseline\ value))/(baseline\ value)\ \times 100.$$

Additional summaries of efficacy and other variables may be produced as a separate report(s) for specific regions, as required by local health authorities.

Depending on the extent of any impact, summaries of data relating to patients diagnosed with COVID-19, and impact of COVID-19 on study conduct (in particular missed visits, delayed or discontinued study treatments, and other protocol deviations) may be generated, see Section 9.4 for further information.

An independent statistical analysis of the primary endpoint will be done by the Arbeitsgemeinschaft Gynäkologische Onkologie (AGO) Study Group and will be described and reported separately.

4.2 Analysis methods

Table 7 provides the summary of all formal statistical analyses planned for this study together with pre-planned sensitivity analyses.

Table 7 Formal statistical analyses to be conducted and pre-planned sensitivity analyses[+]

Endpoint Analysed	Notes
Progression free survival (PFS)	**Analysis Set: FAS (non-*tBRCAm* cohort)** Stratified log-rank test using RECIST data from investigator assessment Hazard ratio with 95% Profile likelihood CI using stratified Cox proportional hazards model with the stratification variables as strata. Plots and summaries of number (%) subjects with progression or death events at landmark time points using Kaplan-Meier. ***Sensitivity analyses:*** • Evaluation time bias; stratified log-rank test using RECIST data from investigator assessment • Attrition bias (using alternative censoring rules); stratified log-rank test using RECIST data from investigator assessment • Ascertainment bias; stratified log-rank test using BICR data • Deviation bias (if meaningful to do); stratified log-rank test using RECIST data from investigator assessment • Unstratified analysis of the primary endpoint • Proportional hazards assumption ***Subgroup analysis:*** For each subgroup factor of interest, the hazard ratio and 95% profile likelihood CI will be calculated from an unstratified Cox proportional hazards model with treatment as a factor, there will be a separate Cox PH model for each subgroup level.
Overall survival (OS)	**Analysis Set: FAS (non-*tBRCAm* cohort)** Stratified log-rank test Hazard ratio using a stratified Cox proportional hazards model Plots and summaries of number (%) subjects with death events at landmark timepoints using Kaplan-Meier. Subgroup analysis using unstratified Cox proportional hazards model at time of the final OS analysis
Progression free survival (PFS)	**Analysis Set: FAS (*tBRCAm* cohort)** Kaplan-Meier estimation of survival using RECIST data from investigator assessment
Overall survival (OS)	**Analysis Set: FAS (*tBRCAm* cohort)** Kaplan-Meier estimation of overall survival
Time from randomisation to second progression (PFS2)	**Analysis Set: FAS (non-*tBRCAm* and *tBRCAm* cohorts)** Stratified log-rank test (non-*tBRCAm* only) Hazard ratio using a stratified Cox proportional hazards model (non-*tBRCAm* only) Plots and summaries of number (%) subjects with second progression or death events at landmark times using Kaplan-Meier.
Time to first subsequent therapy or death (TFST)	**Analysis Set: FAS (non-*tBRCAm* and *tBRCAm* cohorts)** Stratified log rank test (non-*tBRCAm* only) Hazard ratio using a stratified Cox proportional hazards model (non-*tBRCAm* only)

	Plots and summaries of number (%) subjects with first subsequent therapy or death events at landmark timepoints using Kaplan-Meier.
Time to second subsequent therapy or death (TSST)	**Analysis Set: FAS (non-*tBRCAm* and *tBRCAm* cohorts)**
	Stratified logrank test (non-*tBRCAm* only)
	Hazard ratio using a stratified Cox proportional hazards model (non-*tBRCAm* only) Plots and summaries of number (%) subjects with second subsequent therapy or death events at landmark times using Kaplan-Meier.
Time to study treatment discontinuation or death (TDT)	**Analysis Set: FAS (non-*tBRCAm* and *tBRCAm* cohorts)**
	Stratified logrank test (non-*tBRCAm* only)
	Hazard ratio using a stratified Cox proportional hazards model (non-*tBRCAm* only) Plots and summaries of number (%) subjects with discontinuation or death events at landmark timepoints using Kaplan-Meier.
Objective response rate	**Analysis Set: FAS (non-*tBRCAm* and *tBRCAm* cohorts)**
	Summary statistics on objective response rates
	ORR pre-surgery in IDS group
	Summaries will be produced that present the number and percentage of patients with a tumour response (CR/PR) based upon the number of patients with evaluable disease at baseline
Duration of response	**Analysis Set: FAS (non-*tBRCAm* and *tBRCAm* cohorts)**
	Descriptive summary of the Duration of response according to the investigator assessment (In responding patients with evaluable disease at baseline)
	Kaplan-Meier plot and estimate of summary of time to response
EORTC-QLQ-C30 and EORTC-QLQ-OV28	**Analysis Set: FAS (non-*tBRCAm* and *tBRCAm* cohorts)**
	Piecewise linear modelling of change from baseline in functioning subscale (non-*tBRCAm* only).
EQ-5D-5L	**Analysis Population: FAS (non-*tBRCAm* and *tBRCAm* cohorts)**
	Summary statistics on Health state utility
	Plots of summary information
	QAPFS using mixed effects models
	Duration of "good quality of life" using Q-TWiST analysis
pCR	**Analysis Population: FAS (non-*tBRCAm* and *tBRCAm* cohorts)**
	pCR rate in IDS group who have had surgery
PK analysis	**Analysis Population: PK Analysis Set (non-*tBRCAm* cohort)**
	Descriptive statistics will be summarized for all individual concentration data per drug (olaparib/durvalumab) and per time points.
ADA analysis	**Analysis Population: ADA Analysis Set (non-*tBRCAm* cohort)**
	Descriptive statistics will be summarized for all ADA data

[+] Comparisons between treatment groups will only performed in the non-*tBRCAm* cohort as well as non-*tBRCAm* HRD-positive population (as required), whereas for the *tBRCAm* cohort, only descriptive summaries will be shown.

4.2.1 Multiplicity

- To facilitate a strong control of the type I error at the 5% two-sided level, a multiple testing procedure will be employed (see Figure 1). The overall 5% type I error rate will be allocated to the primary PFS comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population. If the PFS analysis for this comparison is statistically significant at

the time of either the interim analysis or the final analysis, 5% alpha (two-sided) will be allocated to the next level in a pre-defined order:

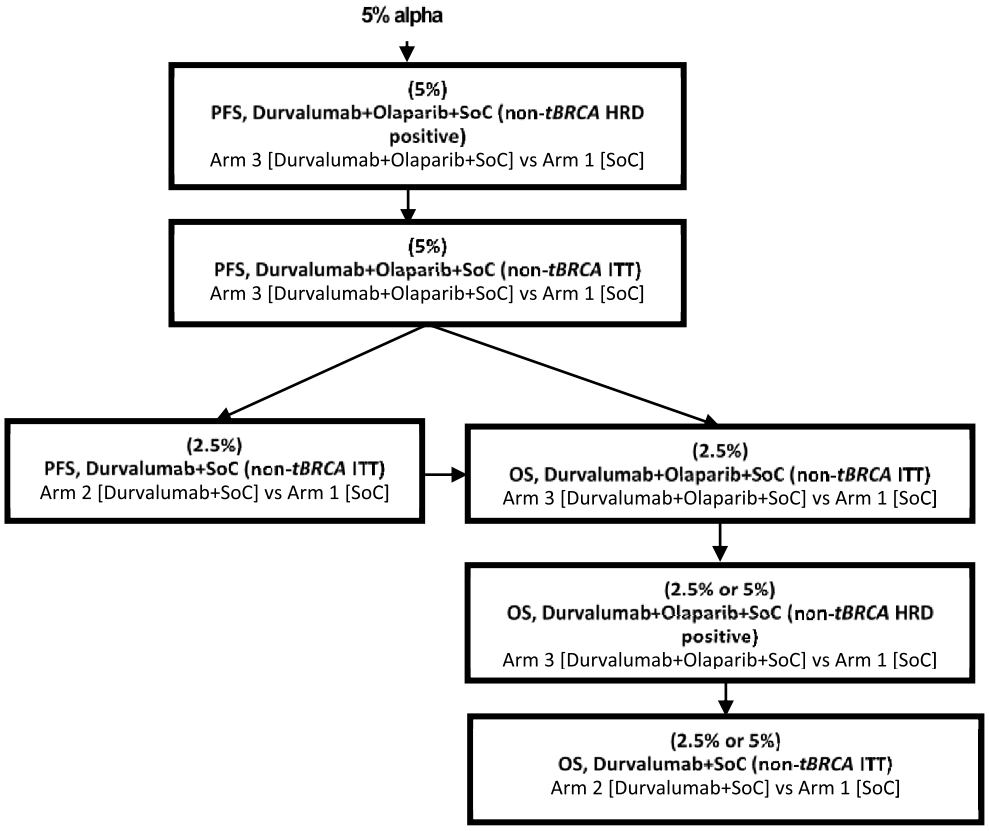
- 5% alpha will be assigned to PFS comparison of Arm 3 vs Arm 1 non-*tBRCAm* ITT population. If statistical significance for PFS comparison of Arm 3 vs Arm 1 non-*tBRCAm* ITT population is met, the 2.5% (two-sided) test mass is recycled to test the PFS comparison of Arm 2 vs Arm 1 non-*tBRCAm* ITT population and the other 2.5% alpha (two-sided) assigned to the OS comparison of Arm 3 vs Arm 1 non-*tBRCAm* ITT population.
- If statistical significance for PFS comparison of Arm 2 vs Arm 1 non-*tBRCAm* ITT population is met, the 2.5% (two-sided) test mass is recycled to test the OS comparison of Arm 3 vs Arm 1 non-*tBRCAm* ITT population and this will be tested at 5% alpha.
- If statistical significance for OS comparison of Arm 3 vs Arm 1 non-*tBRCAm* ITT population is met, the test mass (2.5% or 5% [two-sided]) is recycled to test the OS comparison of Arm 3 vs Arm 1 non-*tBRCAm* HRD-positive population.
- If statistical significance for OS comparison of Arm 3 vs Arm 1 non-*tBRCAm* HRD-positive population is met, the test mass (2.5% or 5% [two-sided]) is recycled to test the OS comparison of Arm 2 vs Arm 1 non-*tBRCAm* ITT population.

A PFS interim analysis and an OS interim analysis is planned (see Section 5 for details). Note: If any interim analysis or primary analysis is statistically significant, the overall 2.5% or 5% (two-sided) alpha will be allocated to the next level[1]. If the interim results do not meet the criterion of stopping for superiority for a given hypothesis, then follow-up will continue until the final target number of PFS/OS events for that comparison has been observed, following which the hypothesis will be retested. If the hypothesis is then rejected, subsequent testing will continue hierarchically. The testing procedure will ensure strong control of the family-wise error rate (Figure 1).

To describe the nature of the benefits of durvalumab and durvalumab + olaparib compared with the comparator arm, the following endpoints will be tested at a 2-sided significance level of 5%; ORR, PFS2, TFST, TSST, TDT and change from baseline score in the physical functioning subscale of the EORTC-QLQ-C30.

[1] The 5% alpha (two-sided) allocation for the primary and secondary PFS endpoints will be controlled at the interim and the final analysis timepoints separately for each PFS comparison by using a bespoke spending function, where a fixed significance level will be assigned at the interim and the remaining significance level assigned to the final analysis, taking account of correlation (Stone 2010). For OS, the 2.5%/5% alpha (two-sided) would be controlled at the interim and the final analysis timepoints separately for each comparison by using Lan De Mets O'Brien-Fleming spending function, where the significance level applied at the interim analysis depends upon the proportion of information (ie, information fraction) available.

Figure 1: Illustration of Multiplicity strategy in the Randomised non-*tBRCAm* Cohort



4.2.2 Primary variable - Progression free survival (PFS)

For the non-*tBRCAm* cohort, the primary PFS analysis of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population will be analysed using a log rank test stratified by timing and outcome of cytoreductive surgery and geographic region in accordance with the pre-defined pooling strategy (Section 4) for the generation of the p-value. Stratification variables will be defined according to data from the IRT system. If there are any patients who were mis-stratified, a sensitivity analysis will be carried out using the (correct) baseline data collected in the eCRF.

The HR and its confidence interval will be estimated from a stratified Cox Proportional Hazards model (with ties = Efron and the stratification factors as covariates) and the CIs calculated using a profile likelihood approach. The hazard ratio together with its 95% CI and p-value will be presented (a hazard ratio less than 1 will favour the comparator arm).

The primary analysis of Arm 3 vs Arm 1 in the non-*tBRCAm* ITT population as well as the key secondary PFS comparison of Arm 2 vs Arm 1 in the non-*tBRCAm* ITT population and exploratory PFS comparisons (for example, PFS comparison of Arm 3 vs Arm 2 in the non-*tBRCAm* ITT population) will be analysed using the same methodology. Each comparison (Durvalumab+SoC vs SoC, Durvalumab+Olaparib+SoC vs SoC and Durvalumab+Olaparib+SoC vs Durvalumab+SoC) will be made from separate models.

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

The primary analysis will be based on the programmatically derived PFS investigator-recorded assessment of disease progression by RECIST v1.1 and using all scans regardless of whether they were scheduled or not. The proportion of patients alive and progression free at 6 monthly intervals from randomisation (6, 12, 18 and 24) will be summarised (using the KM curve) and presented by treatment group.

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

For the *tBRCAm* cohort, summaries of the number and percentage of subjects experiencing a PFS event, and the type of event (RECIST or death) will be provided along with median PFS for each treatment. Also, the proportion of patients alive and progression free at 6 monthly intervals will be summarised (using the KM curve) and presented by treatment group.

Additional supportive summaries/graphs (non-*tBRCAm* cohort only)

The number of patients prematurely censored will be summarised by treatment arm together with baseline prognostic factors of the prematurely censored patients. A patient is defined as prematurely censored if they had not progressed and the latest scan prior to DCO was more than one scheduled tumour assessment interval (+ 2 weeks for protocol allowed window) prior to the DCO date. A KM plot, with tick marks to identify censored observations, of PFS will be presented by treatment group.

In addition, duration of follow-up will be summarised using median time from randomisation to date of censoring (date last known to be non-progressor) in censored (not progressed) patients only, presented by treatment group.

Proportionality assumption (non-*tBRCAm* cohort only)

The assumption of proportionality will be assessed. Proportionality will be tested firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, Cox model with a time dependent covariate would be fitted 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. 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. The treatment effect will also be described by presenting piecewise HRs calculated over distinct time-periods of 0-6, 6-12, 12-18, 18-24, >24 months. The piecewise model will

be implemented by the addition of a time varying covariate/factor (based on the periods in the previous sentence) as per Collet 2003.

PFS sensitivity analyses (non-*tBRCAm* cohort only)

(a) Evaluation-time bias

Sensitivity analyses will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled time points. The midpoint between the time of progression and the previous RECIST assessment will be analysed using a stratified log-rank test, as described for the primary analysis of PFS. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). Note that if the event contributing to the analysis is death in the absence of progression, then the time to event remains unchanged (i.e. the time to death is not replaced by the midpoint between death and the previous RECIST assessment). To support this analysis, the mean of subject-level average inter-assessment times will be tabulated for each treatment. This approach will use the investigator RECIST assessments.

(b) Attrition bias

Attrition bias will be assessed by repeating the primary 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, missed RECIST assessments will be included. In addition, subjects who take subsequent therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) prior to their last RECIST assessment or progression or death will be censored at their last RECIST assessment prior to taking the subsequent therapy.

Additionally, a Kaplan-Meier plot of the time to censoring where the censoring indicator of the primary PFS analysis is reversed (what was originally a censored event in the primary PFS analysis becomes an actual event and what originally was a PFS event becomes a censored event) will be presented.

(c) Ascertainment bias

Ascertainment bias will be assessed from carrying out a blinded independent central review (BICR). The objective of the BICR is to detect potential evaluation bias in the investigator assessment of PFS. A stratified log-rank test will be repeated using BICR RECIST data using the same ties and stratification factors as described for the primary analysis of PFS. The HR and 95% CI will be presented using stratified Cox Proportional Hazard model.

If there is an important discrepancy between the primary analysis using investigator assessments and this sensitivity analysis using BICR assessments, then the proportion of subjects with site but no central confirmation of progression will be summarised; such patients have the potential to induce bias in the central review due to informative censoring. The 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 is determined to exist by the study team.

Disagreements between investigator and BICR assessment of RECIST progression will be presented for each treatment group. The summary will include the early discrepancy rate which is the frequency of BICR progressions declared before the investigator review progressions (≥2 weeks earlier and including progressions declared by BICR but not investigator) as a proportion of all BICR review progressions, and the late discrepancy rate which is the frequency of BICR review progressions declared after the investigator review progressions (≥2 weeks later and including progressions declared by investigator review but not BICR) as a proportion of all discrepancies (including early and late discrepancies) (Amit et al 2010).

In the case where the distribution of discrepancy in progression assessment between BICR and local investigator across treatment groups is not similar, the PFS analysis may be biased due to informative censoring. The potential impact of informative censoring on parameter estimate will be assessed through sensitivity analysis, using either the methods of Jackson et al or Hsu and Taylor (Jackson et al 2014, Hsu and Taylor 2009) when considering time dependent covariates. This work will be presented separately and will not form part of the CSR.

(d) Deviation bias (if meaningful to do)

As a sensitivity analysis to the primary PFS endpoint, an analysis excluding patients with deviations that may affect the efficacy of the trial therapy will be performed if > 10% of patients:

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

A stratified log-rank test will be repeated using the investigator RECIST data, using the same ties and stratification factors as described for the primary analysis of PFS. The HR and 95% CI will be presented using stratified Cox Proportional Hazard model.

(e) Other sensitivity analyses

.

Discrepancy between IRT reported and eCRF reported stratification factors

A sensitivity analysis to the primary PFS endpoint will be conducted to evaluate an assessment of the discrepancy between IRT reported stratification factors and the eCRF reported factors and their potential impact on the primary analysis.

(f) Additional supportive summaries

In addition, the number of patients prematurely censored will be summarised by treatment arm together with baseline prognostic factors of the prematurely censored patients. A patient would be defined as prematurely censored if they had not progressed (or died in the absence

of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval plus 2 weeks prior to the DCO date.

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

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

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

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

Subgroup Analysis (non-*tBRCAm* cohort only)

Subgroup analyses will be conducted comparing PFS (per RECIST 1.1 using investigator assessments) between the following treatments, Arm 2 vs Arm 1 and Arm 3 vs Arm 1 (i.e. Durvalumab+SoC arm versus SoC arm and Durvalumab+Olaparib+SoC arm versus SoC arm), in the non-*tBRCAm* ITT population and HRD-positive population as well as for the exploratory comparison of Arm 3 vs Arm 2 (Durvalumab+Olaparib+SoC arm versus Durvalumab+SoC arm) in the non-*tBRCAm* ITT population and HRD-positive population. The purpose of the subgroup analyses is to assess the consistency of treatment effect across potential or expected prognostic factors.
The following subgroups of the full analysis set will be analysed for PFS:

Stratification factors:

- Timing and outcome of cytoreductive surgery (no residual macroscopic disease after upfront primary surgery vs all others [residual macroscopic disease after upfront primary surgery OR planned IDS]),
- Geographic region (North America vs Europe vs RoW);

Additional factors:

- Age at randomization (<65 years vs ≥65 years). Age will be derived from “age” (DM module) on the eCRF at screening
- ECOG performance status (PS0 vs PS1); This will be determined from “Performance status” (ECOG Performance Status module) on the eCRF at screening
- Stage of disease at diagnosis (Stage III vs Stage IV);
- Surgery status at study entry (upfront primary surgery or planned IDS)
- Myriad HRD status (HRD positive vs HRD negative vs HRD unknown)*, as required; Positive: sample with a genomic instability score above or equal to 42 using the Myriad myChoice HRD Plus; negative: sample with genomic instability score below

42 using the Myriad myChoice HRD Plus; unknown: sample that failed genomic instability score testing.

- Homologous recombination repair related gene mutation (HRRm) status (HRRm positive vs HRRm negative vs HRRm unknown). As defined by the presence of a deleterious or suspected deleterious mutation in one of the following HRR genes (ATM, BRIP1, PALB2, RAD51C, BARD1, CDK12, CHEK1, CHEK2, FANCL, RAD51B, RAD51D, RAD54L)
- PD-L1 expression (high vs low vs unknown). High: sample with PD-L1 expression at any intensity above or equal to predefined cut-offs; Low: sample with PD-L1 expression at any intensity below predefined cut-offs. Unknown: sample where PD-L1 expression was not available either due to a test fail (unevaluable sample or assay failure) or sample slide out of cut-slide stability. The following cut-offs will be assessed:
    - an immune cell only positivity score and ≥25% cut-off (IC25),
    - a tumour and immune cell positivity score and ≥5% cut-off (TIP5).

  PD-L1 expression will be evaluated applied to sections stained using Ventana SP263 immunohistochemistry assay. Other exploratory cut-offs may also be assessed as required.

* This subgroup will only be analysed in the ITT population

Other biomarker subgroups can be added prior to database lock based on emerging clinical trial evidence.

In addition, the following subgroup based on post-baseline status will be further described within an exploratory analysis plan and may be reported outside the main CSR.

- Response to chemotherapy (CR/PR/NED vs non CR/PR/NED) at the end of chemotherapy.

An analysis will not be performed if there are too few events available for a meaningful analysis of a particular subgroup (i.e., if there are less than 20 events across both treatment groups in a subgroup).

The subgroup analyses for the stratification factors will be based on the values entered into the IRT system, all other factors will be based on values recorded on the eCRF as indicated above, or third party vendor data.

Other baseline variables may also be assessed if there is clinical or biological justification or an imbalance is observed between the treatment arms.

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 to the primary analysis of PFS.

For each subgroup level of a factor, the HR and 95% CI will be calculated from an unstratified Cox proportional hazards model that only contains treatment as a term. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties and using a BY statement for the subgroup factor. These HRs and associated two-sided 95% profile likelihood CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis. An unstratified analysis of PFS using the FAS population will be presented alongside the subgroup analyses.

Consistency of treatment effect between subgroups (non-*tBRCAm* cohort only)

The presence of quantitative interactions will be assessed by means of an overall global interaction test for plausible subgroups:

The plausible subgroups with biological rationale for an interaction with treatment consist of all baseline covariates (including stratification factors and subgroups mentioned above). Note, only a single PD-L1 cut-off (TIP5) will be included when testing for interactions.

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, 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.

4.2.3 Overall survival (OS)

OS data will be analysed at the time of the primary analysis of PFS and will use the same methodology and model in the non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT

population (provided there are sufficient events available for a meaningful analysis, see the pooling strategy in Section 4, if not descriptive summaries will be provided).

The sensitivity analysis outlined for PFS in Section 4.2.2 will not be repeated for OS except for a Kaplan-Meier plot of the time to censoring where the censoring indicator of the OS event is reversed (what was originally a censored event in the primary OS analysis becomes an actual event and what originally was a OS event becomes a censored event).

A summary of survival status at the time of analysis will be produced. This will summarise the number of patients who have died, who are still in survival follow-up, who are lost to follow-up or who have withdrawn consent.

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

- In censored (not died) patients only: Time from randomisation/allocation to date of censoring (date last known to be alive)
- In all patients: Time from randomisation/allocation to the date of death or to the date of censoring for censored patients.

The percentage OS at 6-month intervals will be summarised for the analysis of OS at the primary PFS analysis, and the percentage OS at 12-month intervals for the analysis of OS at the final survival follow up.

A descriptive analysis of OS will occur at the time of the interim PFS analysis, and an interim OS analysis at the time of the primary PFS analysis. A final analysis of OS may be performed at approximately 50% maturity or 5 years after the last non-*tBRCAm* patient is randomised to treatment, whichever occurs sooner.

The subgroup analyses that will be done for PFS will be done for OS at the time of the final OS analysis.

4.2.4 Progression free survival 2 (PFS2)

Time from randomisation to second progression or death (PFS2) will be analysed in the non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population using the same methods as outlined for PFS and adjusting for the same set of covariates – stratification factors, but no subgroup analysis will be performed. The HR for the treatment effect together with its 95% CI will be presented together with the median PFS2 time by treatment arms. Also, Kaplan-Meier plots will be presented to support the analysis.

The number and percentage of patients experiencing a PFS2 event and the type of progression (progression by investigator assessment of radiological progression, CA-125 progression, symptomatic progression, or death) will also be summarised by treatment arm. Time from randomisation to second progression will be summarised by treatment arm.

The sensitivity analysis outlined for PFS in Section 4.2.2 will not be repeated for PFS2 except for a Kaplan-Meier plot of the time to the censor event where the censoring indicator of the

PFS2 event is reversed (what was originally a censored event in the PFS2 analysis becomes an actual event and what originally was a PFS2 event becomes a censored event).

No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

4.2.5 Time to first subsequent therapy or death (TFST) and time to second subsequent therapy or death (TSST)

Time to first subsequent therapy or death (TFST) and time to second subsequent therapy or death (TSST) will be analysed in the non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population at the same time as the primary analysis of PFS and final OS analysis using the same methodology and model as that used for the analysis of PFS. The HRs for the treatment effect together with 95% CIs will be presented. KM plots will be presented by treatment arm. In addition, time between progression and starting subsequent therapy will be assessed.

Summary tables of first and second subsequent therapies by treatment arm will be provided, as well as response to first and second subsequent therapy by treatment arm.

No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

4.2.6 Time to study treatment discontinuation or death (TDT)

Time to permanent study treatment discontinuation or death (TDT) will be analysed in the non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population at the same time as the primary analysis of PFS and using the same methodology and model. The HR for the treatment effect together with 95% CIs will be presented. A KM plot will be presented by treatment arm. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

4.2.7 Objective response rate

The ORR will be based on the site investigator RECIST data, and using all scans regardless of whether they were scheduled or not.

For each treatment group in the non-*tBRCAm* cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population), the objective response rate (ORR) will be the number of patients with a CR or PR post-baseline divided by the number of patients in this treatment group in the FAS with evaluable disease at baseline.

For the *tBRCAm* cohort, the objective response rate (ORR) is the number of patients with a CR or PR post-baseline in this cohort divided by the number of patients in the FAS for this cohort with evaluable disease at baseline.

In all the assessments, only patients in either cohort with evaluable disease at baseline can achieve an objective response of CR or PR. A summary of all responses will also be presented by treatment group for the non-*tBRCAm* cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population) and for the single arm in the *tBRCAm* cohort.

In addition, the overall response rate will be summarised prior to surgery in those patients planned to have IDS with evaluable disease at baseline.

In addition, for patients who have an objective response, the duration and onset of the response will be summarised.

Time to onset of objective response is defined as the time from the date of allocation to cohort or randomisation until the date of first documented response. Time to response will not be defined for those patients who do not have documented response.

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

4.2.8 Duration of response

Duration of response (DoR) will be defined as the time from the date of first documented response until date of documented progression or death in the absence of disease progression. 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 PR or CR. If a subject does not progress following a response, then their duration of response will use the PFS censoring time.

For the non-*tBRCAm* cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population), descriptive data will be provided for the duration of response in responding patients, including the associated Kaplan-Meier curves (without any formal comparison or p-value attached).

Kaplan Meier plots of DoR based on the investigator assessment of RECIST will be presented. Median DoR will also be summarised calculated from the KM curve.

For *tBRCA*m cohort, descriptive data will be provided for the duration of response in responding patients, including the associated Kaplan-Meier curves (without any formal comparison or p-value attached).

Kaplan Meier plots of DoR based on the investigator assessment of RECIST will be presented. Median DoR will also be summarised calculated from the KM curve. Additionally, swimmer plots that clearly show the profile of each patient who responds will also be produced.

4.3 Patient Reported Outcomes (PROs)

For PRO summary by timepoint, all data including unscheduled assessment will be included according to the time window defined as described in section 3.6.8.

EORTC QLQ-C30 & QLQ-OV28

Change from baseline

A summary of absolute and unadjusted change from baseline for each EORTC QLC-C30 (except for the single item scale financial difficulties) and EORTC QLQ-OV28 scale will be reported by timepoint for each treatment group.

For the non-tBRCA cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population), mean change from baseline for each scale (14 scales of EORTC QLQ-C30 and 8 scales of EORTC QLQ-OV28) will be analysed using a piecewise linear modelling analysis of the change from baseline (where baseline is defined as prior to dosing on Cycle 1 Day 1) in each scale for each visit for Durvalumab+SoC versus SoC as well as Durvalumab+Olaparib+SoC versus SoC. The primary analysis will be to compare the average treatment effect from the point of randomisation for the first 24 months (which will include analysis visits obtained within the first 24 months, i.e. baseline, day 43 (week 6), weeks 12, 18, 24, 30, 36, etc.) unless there is excessive missing data (defined as > 75% missing data).

Graphical plots of the mean of each EORTC QLQ-C30 scale/item scores (except financial difficulty score) and each EORTC QLQ-OV28 scale/score, including change from baseline, and associated 95% CI by scheduled visits/timepoints in the study will be produced.

Response (improvement, no change, deterioration)

At each post-baseline assessment, the change from baseline in each EORTC QLQ-C30 scale/item score (global health status/QoL, functions, and all symptoms) or each EORTC QLQ-OV28 scale/score will be categorized as improvement, no change, or deterioration as defined in **Table 8**. A summary table of the number (%) of subjects in each response category (improvement, no change, deterioration) for each visit will be presented by treatment group.

Table 8 EORTC QLQ-C30 and QLQ-OV28 change categories

Score	Change from baseline	Visit response
Health status/QoL scale and all functional scales	≥+10 (increase of at least 10)	Improvement
	≥-10 (decrease of at least 10) or "Subject too sick to complete the questionnaires (disease under investigation)"	Deterioration
	Otherwise	No change
Symptom scales/scores	≥+10 (increase of at least 10) or "Subject too sick to complete the questionnaires (disease under investigation)"	Deterioration
	≥-10 (decrease of at least 10)	Improvement
	Otherwise	No change

EORTC European Organisation for Research and Treatment of Cancer; QLQ C30 30-item core quality-of-life questionnaire; OV28 ovarian-cancer specific module.

In addition, a summary table of improvement for global health status/QoL, 5 functions, and 3 symptoms (fatigue, pain, nausea/vomiting) will be presented by treatment group.

Improvement for each scale/score will be calculated as the number (%) of subjects with improvement defined as meeting one of the following criteria:

1. Has 2 consecutive visit responses of "improvement" at least 21 days apart
2. Has 1 visit response of "improvement" with no further assessments and did not die within 2 PRO assessment visits

PGIS

Responses on the PGIS will be summarized descriptively as the number (%) of subjects with each response at each visit by treatment group.

PRO-CTCAE

PRO-CTCAE data will be summarized descriptively as the number (%) of subjects with each level of response for each PRO-CTCAE item at each visit by treatment group.

EQ-5D-5L

Descriptive statistics will be calculated for each scheduled visit/time point in the study, for each trial arm and as a total. These will report the number of patients, the number of EQ-5D questionnaires completed at each visit, the number and proportion responding to each dimension of the EQ-5D-5L. Additionally summary statistics (e.g. n, mean, median, sd, min, max) will 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.

Additionally, for the EQ-5D, graphs and listings will be reported for health state utility values and visual analogue scale by visits as well as change in these scores from baseline.

To assess the amount of missing data and the reasons for missing data, plots including the number of questionnaires completed/partially completed/missing at each visit may be produced.

Graphical plots of the mean EQ-5D index score and EQ-VAS score, including change from baseline, and associated 95% CI by scheduled visits/time points in the study will be produced. To support submissions to payers, additional analyses may be undertaken, and these will be outlined in a separate Payer Analysis Plan.

QAPFS

For each treatment group, Quality adjusted PFS will be calculated as the product of the mean EQ-5D index score from randomisation to progression and the restricted mean survival time for PFS. For the non-tBRCA cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population), the between-group difference in quality-adjusted PFS for SoC+Durvalumab+Olaparib vs SoC and SoC+Durvalumab vs SoC will be presented with corresponding 95% confidence intervals (CIs) and p-values which will be calculated using bootstrap resampling methods. Results for the single-arm tBRCA cohort will be summarised.

The restricted mean survival time for PFS will be calculated as the area under the Kaplan-Meier curve for time to progression up to the last time point where there is an observation in all arms of the study.

The mean EQ-5D single index utility score will be estimated using a mixed model for repeated measures (MMRM) analysis. The model will include all EQ-5D measures up to the last measurement prior to the date of progression The MMRM model for the non-*tBRCAm* cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population) will include the fixed effects of treatment, baseline EQ-5D index score, visit and treatment by visit interaction. For the *tBRCAm* cohort, the MMRM model will include baseline EQ-5D index and visit as fixed effects. SUBJECT will be included as a random effect in the MMRM. Mean adjusted EQ-5D index scores for each treatment group will be estimated using least squares mean estimation.

The MMRM analyses will be performed using Restricted maximum likelihood estimation (REML), and an unstructured covariance matrix to model the within-subject error. Kenward-Roger approximation will be used to estimate the degrees of freedom. Alternative covariance structures will be explored if the analysis fails to converge using the unstructured covariance matrix. Also, SUBJECT may be treated as a fixed effect if the model fails to converge with all other possible covariance structures.

TWiST and QTWiST

For both the *tBRCAm* and non-*tBRCAm* cohorts, the time without symptoms of disease progression or toxicity (TWiST) and the Quality-adjusted time without symptoms of disease progression or toxicity (QTWiST) will be calculated to determine the duration of 'good quality of life' for each treatment group. In the non-*tBRCAm* cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population), the between-group difference in TWiST and QTWiST for SoC+Durvalumab+Olaparib vs SoC and SoC+Durvalumab vs SoC will be presented with corresponding 95% confidence intervals (CIs) and p-values. Results for the single-arm *tBRCAm* cohort will be summarised.

The TWiST and Q-TWiST analyses will follow standard approaches, in which the survival time for patients is partitioned into three health states:

1. Toxicity (TOX): is the period with "significant symptoms" of toxicity during the period up to protocol defined disease progression (or censoring for progression)
2. Time without significant symptoms of disease progression or toxicity (TWiST): is the period without "significant symptoms" of disease progression or toxicity during the period up to protocol defined disease progression (or censoring for progression)
3. Relapse (REL): is the period between protocol defined progression and death (or censoring for death).

- 'Significant symptoms' of toxicity will be defined as any period with a CTCAE grade≥3 AE. For each patient, the total duration of time spent with at least one of these toxicities after randomisation and before progression will be estimated. Any periods

spent with more than one type of toxicity will be counted once, such that overlapping toxicity intervals are not counted twice. Patients with an unresolved toxicity will be assumed to have toxicity until progression or censoring for PFS. Patients that do not experience toxicity prior to progression will be assigned a toxicity duration of zero, and will be censored for toxicity on day 1 of randomisation.

- Note a sensitivity analysis may be undertaken whereby 'Significant symptoms' of toxicity will be defined as any period with CTCAE grade≥3 AE or CTCAE grade 2 for nausea, vomiting or fatigue.

TWiST will be estimated using the restricted mean survival time for PFS minus the period spent in the TOX state:

TWiST duration = Mean PFS-Mean TOX

Survival curves corresponding to the period with TOX and for PFS will be estimated separately for each treatment group.

The Q-TWiST will be estimated by combining the TWiST and TOX durations with the restricted mean survival time in the relapse state (derived as mean OS minus the mean PFS) and the mean EQ-5D single index utility scores assigned to each state, as follows:

$$\text{Q-TWiST}=(u_{TOX} \text{ x TOX}) +(u_{TWiST} \text{ x TWiST}) + (u_{REL} \text{ x REL})$$

Where TOX, TWiST and REL are the mean health state durations and u_{TOX}, u_{TWiST}, u_{REL} are the EQ-5D single index scores for each state, averaged across treatment arms. The mean EQ-5D utility index for the states of TOX, TWiST and Relapse will be estimated using a MMRM analysis. The model will include all EQ-5D measures collected in the study.

For the non-*tBRCAm* cohort (non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population), the between-group difference in TWiST and Q-TWiST will be estimated for SoC + Durvalumab versus SoC and for SoC+Durvalumab+Olaparib versus SoC. A 95% bootstrap CI and two-sided p-value will be obtained for the differences in mean TWiST and Q-TWiST for the treatment comparisons in non-*tBRCAm* cohort. Note the Q-TWIST analysis will only performed at time of the final OS analysis.

Additional PRO analyses will be described in a separate data analysis plan.

4.4 Safety

Safety data will be summarised and listed only. No formal statistical analyses will be performed on the safety data. Similar summary tables and listings will be reported/created for both the *non-tBRCAm* and *tBRCAm* safety population separately. In addition a select number of safety outputs maybe repeated in the non-*tBRCAm* HRD positive safety population, as required.

4.4.1.1 Adverse events

All AEs, both in terms of Medical Dictionary for Regulatory Activities (MedDRA) preferred term and common toxicity criteria for adverse events (CTCAE) grade, will be listed and summarised descriptively by count (n) and percentage (%) for each treatment arm. MedDRA dictionary will be used for coding. The AE summaries, unless otherwise stated, will be based on treatment-emergent AEs up until the start of subsequent anti-cancer therapy (including non-palliative radiotherapy) or until the end of safety follow-up period (latest of either 30 days following discontinuation of olaparib/placebo or 90 days following discontinuation of durvalumab/placebo), whichever occurs first. This will more accurately depict AEs attributable to only the study treatment as a number of AEs following discontinuation of the study treatment are likely to be attributable to subsequent therapy. In order to assess the longer-term toxicity profile, a selection of AE summaries will also be produced containing AEs observed up until the end of follow-up period (i.e. without taking subsequent therapy into account).

Any events in this period that occur after a patient has received further therapy for cancer will be flagged in the data listings.

Any AE occurring before Cycle 2 Day 1 of study will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'.

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

Handling of missing data

Generally, the imputation of dates is used to decide if an observation is treatment emergent for adverse events (AE) or concomitant medications (CM). The imputed dates are not advised to be used to calculate durations where the results would be less accurate.

- 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.

- Partial or missing AE or CM start dates will be imputed as follows:

  For date of onset:
  - Missing day - Impute the 1st of the month unless month is same as month of first dose of randomised/allocated 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 – 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 i.e. is prior to the end date of the CM/AE.

- Partial or missing AE or CM end dates will be imputed as follows:

  For date of resolution:
  - Missing day - Impute the last day of the month. If the subject died in the same month, then set the imputed date as the death date
  - Missing day and month – impute 31st December . If the subject died in the same year, then set the imputed date as the death date
  - Completely Missing – No imputation

Summary information will be tabulated overall and where required by treatment phase (initial induction and maintenance phase (see Section 2.1 for definition) for each treatment group for:

- All AEs
- All AEs causally related to study medication
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, causally related to study medication
- AEs with CTCAE grade 3 or 4
- AEs with CTCAE grade 3 or 4, causally related to study medication
- AEs with outcome of death
- AEs with outcome of death causally related to study medication
- All serious adverse events (SAEs)
- All SAEs causally related to study medication
- AEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication, causally related to study medication
- Other significant AEs
- Immune mediated AEs (as determined by the reporting investigator).
- Infusion reaction AEs (as determined by the reporting investigator)
- Other significant AEs causally related to study medication

An overall summary of the number and percentage of patients in each category will be presented, as will an overall summary of the number of episodes in each category. In addition, a truncated adverse event table of most common AEs, showing all events that occur in at least 5% of patients in the safety analysis set will be summarised by preferred term, by decreasing frequency in the total column. 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 of 4.9% will not appear if a cut-off is 5%).

Each AE event rate (per 100 patient years) will also be summarised by preferred term within each system organ class for the output summarizing all AEs. For each preferred term, the event rate will be presented and will be defined as the number of patients with that AE divided by the total safety follow up (days) of any randomized/allocated treatment summed over patients in each group multiplied by 365.25 x 100 to present in terms of per 100 patient years.

Total safety follow up is defined as date of first dose of any of the investigational study treatments, including durvalumab/placebo or olaparib/placebo up until the start of subsequent anti-cancer therapy (including non-palliative radiotherapy) or until the end of safety follow-up period (latest of either 30 days following discontinuation of olaparib/placebo or 90 days following discontinuation of durvalumab/placebo), whichever occurs first.

AEs will be assigned CTCAE grades (National Cancer Institute (NCI) CTCAE version 5.0 and summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, system organ class, preferred term and treatment group. Fluctuations observed in CTCAE grades during study will be listed for those AEs which are CTCAE ≥ 3.

Summaries of the number and percentage of patients with AEs leading to dose modification of study medication by preferred term and treatment group will be presented for the following:

- AEs leading to a dose reduction of study medication where allowable
- AEs leading to a dose interruption of study medication
- AEs leading to a dose modification, defined as a dose interruption, dose delay and/or dose reduction of study medication.

Adverse events with outcome of death, SAEs, AEs leading to discontinuation of treatment, AEs causally related to study medication, and OAEs will be listed.

A summary of deaths will be provided with number and percentage of patients by treatment group, categorised as:

- Total number of deaths (regardless of date of death)
- Related to disease under investigation only
- AE with outcome of death only
- Death related to disease under investigation and an AE outcome of death
- Deaths after end of safety follow up period (as defined above), unrelated to disease under investigation
- Patients with unknown reason for death
- Other deaths

A corresponding listing will also be produced.

Summary of long term tolerability of AEs

For each AE, median time to first onset of the AE will be presented in patients in the safety analysis set by treatment group. Patients who did not experience the AE will not be included in the summaries. Summary tables of time to first onset for each AE will also be produced (e.g. 1-28 days, 29-56 days, 57-84 days, 85-112 days, >112 days). Median duration of the AE will be presented in patients who experienced each AE.

4.4.1.2 Adverse events of special interest (AESI) and adverse events of possible interest (AEPI)

Preferred terms used to identify AESI and adverse events possible interest (AEPIs are for durvalumab only) will be listed before DBL and documented in the Study Master File.

For Durvalumab: Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI/AEPI grouping. 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.

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

- Any AESI/AEPI
- Any AESI/AEPI by PT and maximum CTCAE grade
- Any AESI/AEPI of maximum CTCAE grade 3 or 4
- Any serious AESI/AEPI
- Any AESI/AEPI with outcome of death
- Any AESI/AEPI causally related to study medication
- At least one AESI/AEPI leading to discontinuation of study medication
- 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)

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

For Olaparib: A separate summary table will also be produced capturing any toxicities of interest. Note the summary table capturing the toxicities of interest of MDS/AML and new primary malignancy includes events from first dose of study drug until the end of the study (i.e. not restricted to treatment emergent events).

The toxicities of interest are:

- Myelodysplastic syndrome/Acute Myeloid Leukemia
- New Primary Malignancy
- Pneumonitis

Immune-mediated Adverse events (imAEs)

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

4.4.1.3 Laboratory assessments

Data obtained up until the end of safety follow-up period (latest of either 30 days following discontinuation of olaparib/placebo or 90 days following discontinuation of durvalumab/placebo) or until the initiation of the first subsequent therapy (including non-palliative radiotherapy) following discontinuation of 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 immunotherapy agents or 30 days following discontinuation of olaparib are likely to be attributable to subsequent therapy. However, to assess the longer-term toxicity profile, summaries of laboratory data may also be produced containing data collected up until 90 days following discontinuation of durvalumab/placebo or up until 30 days following discontinuation of olaparib/placebo (i.e., without taking subsequent therapy into account).

A small selection of summaries of laboratory data may also be produced containing data from initiation of the first subsequent therapy following discontinuation of study treatment until 90 days following discontinuation of durvalumab/placebo or until 30 days following discontinuation of olaparib/placebo (i.e. summarising the laboratory data collected on patients taking subsequent therapy during the safety collection follow-up window post discontinuation of study treatment). These outputs will only be produced if the number of laboratory toxicities observed warrant the inclusion of such outputs for interpretational purposes. Any data post 90 days last dose for durvalumab/placebo agents or post 30 days last dose for olaparib/placebo will not be summarised.

Data summaries will be provided in International System (SI) of units.

Box-plots of absolute values and change from baseline for a selection of continuous laboratory assessments will be presented.

For all continuous laboratory assessments, absolute value, change from baseline and percentage change from baseline will be summarised using descriptive statistics at each scheduled assessment time by treatment group. For categorical laboratory assessments, shift from baseline will be summarised using frequency and proportion at each scheduled assessment time by treatment group.

Shift tables for laboratory values by worst common toxicity criteria (CTC) 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 CTC grade shift outputs will be produced are:

- Haematology: Haemoglobin, Leukocytes, Lymphocytes, Neutrophils, Platelets, MCV
- Clinical chemistry: ALT, AST, ALP, LDH, Amylase, TSH, BUN, Total bilirubin, Albumin–hypo and – hyper, Sodium–hypo and–hyper, Potassium–hypo and–hyper, Corrected calcium–hypo and–hyper, Glucose–hypo and–hyper, Creatinine.

Liver Function

For liver biochemistry, the following summaries will include the number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study
  - ALT ≥ 3x –≤ 5x, > 5x – ≤8x, > 8x - ≤ 10x, >10x - ≤ 20x and >20x Upper Limit of Normal (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, and >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 patients who have ALT ≥ 3x ULN plus Total bilirubin ≥ 2x ULN or AST ≥ 3x ULN plus Total bilirubin ≥ 2x ULN at any visit.

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

Plots of ALT and AST vs. Total bilirubin all expressed as multiples of ULN by treatment group will also be produced with reference lines at 3×ULN for ALT, AST, and 2×ULN for Total bilirubin. In each plot, Total bilirubin will be in the vertical axis.

Assessment of Thyroid Function Test Results

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

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

A separate summary will present:

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

- Grade change from baseline to on treatment minimum and maximum

Absolute value and change from baseline of TSH, free T3 and free T4 will be summarized using descriptive statistics at each scheduled assessment time.

Assessment of Renal Function Test Abnormalities

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

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

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

Urinalysis results (categorical data collected at baseline and only if clinically indicated post-baseline) will be listed only.

Clinically significant laboratory results will be flagged and listed. Reference ranges will also be listed. All laboratory summaries and listings will be presented by treatment group.

4.4.1.4 Concomitant medications

Concomitant medications will be summarized by the coded terms. The number of subjects receiving a medication will be summarized for the full analysis set by Anatomical Therapeutic Chemical category name (ATC level 4) and generic term (preferred term) by treatment group. A medication taken during the course of the study is considered concomitant. A subject is only counted once if receiving the medication more than once. Medications received prior to, concomitantly, or post-treatment will be coded using the latest version of WHODrug.

Disallowed medications will be listed.

4.4.1.5 ECGs

ECG data obtained up until the 30-day (for olaparib/placebo) or 90-day (for durvalumab/placebo) safety follow-up visit will be listed by treatment group.

4.4.1.6 Vital signs

Vital signs (pulse rate, systolic blood pressure (SBP), diastolic blood pressure (DBP), temperature, weight and height [baseline only]) will be summarised over time in terms of absolute values and changes from baseline at each scheduled visit. Flags will be applied to values falling outside the project reference ranges (which will be explicitly noted on these listings where applicable).

Data obtained up until the end of safety follow-up period (90 days following discontinuation of durvalumab/placebo or 30 days following discontinuation of olaparib/placebo) or until the initiation of the first subsequent therapy (including non-palliative radiotherapy) following discontinuation of treatment (whichever occurs first) will be used for reporting.

Vital signs data will be listed by treatment group.

4.4.1.7 Demographic and baseline characteristics

The following will be listed and summarised by randomised treatment arm:

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations
- Inclusion in analysis populations
- Demographics (age, age group (<65 years vs ≥65 years), race and ethnicity);
- Patient characteristics at baseline (height, weight, weight group, body mass index (BMI) and body mass index group ('Underweight (<18.5)','Normal (≥18.5 - <25)', 'Overweight (≥25 - 30)' and 'Obese (≥30)'))
- Stratification factors as per randomisation and eCRF
- Patient recruitment by country and centre
- Previous therapy for other cancer
- Disease characteristics at baseline (ECOG performance status, primary tumour location, histology type, tumour grade, International Federation of Gynecology and Obstetrics (FIGO) stage, Cytoreductive surgery and Baseline CA-125 value)
- Disease related medical history
- Relevant surgical history (including history and outcome of debulking surgery)
- Number of blood transfusions
- Physical examination at baseline
- Disallowed and/or allowed concomitant medications
- Post-discontinuation cancer therapy
- HRD status (HRD positive vs HRD negative vs HRD unknown);
- Homologous recombination repair related gene mutation (HRRm) status (HRRm positive vs HRRm negative vs HRRm unknown);
- PD-L1 expression (PD-L1 high vs PD-L1 low vs PD-L1 unknown).

Patient disposition data will again be summarised at the time of final OS analysis.

4.4.1.8 Treatment exposure

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

- Total exposure of olaparib/durvalumab/placebo and SoC
- Actual exposure of olaparib/durvalumab/placebo and SoC
- Number of days on 300 mg olaparib/placebo bd = actual exposure for the dose assigned.

- Reasons for dose reductions, dose interruptions, and dose modifications of olaparib/placebo. Dose reductions and dose interruptions will be based on investigator initiated dosing decisions. Dose interruptions due to "Subject Forgot to Take Dose" will be omitted from these summaries.
- Reasons for dose interruption or delay of durvalumab/placebo. Dose interruptions or delays will be based on investigator initiated dosing decisions.
- For patients on study treatment at the time of the PFS analysis, the data cut off (DCO) date will be used to calculate exposure.
- All treatment information data will be listed for the safety analysis set by treatment group.
- Selected exposure data will also be summarised and listed at the time of updated OS analysis .
- RDI of olaparib/durvalumab/placebo (entire intended treatment period).

Subsequent Therapy

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

4.4.2 Pharmacokinetics and Immunogenicity

The plasma concentration-time data for patients receiving olaparib and the serum concentration-time data for patients receiving durvalumab will be summarized as part of the CSR. A separate PK SAP will be produced, and the results described in a separate report

All data received will be presented in data listings. Pharmacokinetic summaries will be presented for patients in the PK analysis set. Descriptive statistics will be given for all individual concentration data per drug (olaparib or durvalumab) and per time points.

- Data from patients excluded from the PK analysis set will be included in the data listings, but not in the summaries.
- Extra measurements (such as unscheduled or repeat assessments) will also not be included in summary tables but will be included in patient listings.
- Concentration data will be summarized using descriptive statistics, including n, n<lower limit of quantification, arithmetic mean, SD, coefficient of variation (%CV), geometric mean (Gmean), Geometric coefficient of variation (%GCV), median, minimum, and maximum values. The geometric mean is calculated as the exponential of the arithmetic mean calculated from data on a log scale. The %GCV is calculated as $100 \cdot \sqrt{(\exp(s2)-1)}$ where s is the standard deviation of the data on a log scale.
- For all data, descriptive statistics will follow the rounding convention of the individual data. Coefficients of variation (%CV and %GCV), where reported, will always be reported to 1 decimal place.

The PK data will be presented by treatment arm and per sample time points. Individual concentration data will be reported as they are. Descriptive statistics may be rounded for reporting purposes.

For descriptive statistics of plasma/serum concentrations, non-quantifiable (NQ) values of plasma concentrations will be handled as follows:

- If, at a given time point, 50% or less of the plasma/serum concentrations are NQ, the mean, standard deviation, geometric mean, geometric standard deviation, %CV, and %GCV will be calculated by substituting the limit of quantification (LOQ) for values which are NQ
- If more than 50%, but not all, of the concentrations are NQ, the mean, geometric mean, standard deviation, geometric standard deviation, %CV, and %GCV will be reported as not calculable (NC)
- If all the concentrations are NQ, the geometric mean and mean will be reported as NQ and the standard deviation, geometric standard deviation, %CV, and %GCV as NC.

In addition, olaparib PK data may be explored graphically. The following plots may be presented:

- Individual patient plots of plasma concentration versus time. Sampling time on the x-axis with linear scale, concentration on the y-axis with log scale. The above plots will be repeated with concentration on the y-axis with linear scale
- Geometric mean (± geometric standard deviation, if appropriate) or combined individual line plots of plasma concentration versus time. Sampling time on x-axis with linear scale and concentration on y-axis with log scale. The above plots will be repeated with concentration on y-axis with linear scale

Immunogenicity analysis

Immunogenicity results will be analysed descriptively by summarizing the number and percentage of patients who develop ADA to durvalumab by ADA categories (Section 3.5). ADA titer and neutralizing antibody results will be reported for samples confirmed positive for the presence of ADA. Summaries will be based upon all patients in the ADA analysis set. Immunogenicity results will be listed for all patients in the safety analysis set regardless of ADA-evaluable status. AEs in ADA positive patients by ADA positive category will be listed.

The effect of immunogenicity as well as the effect of its neutralizing properties on PK, pharmacodynamics, efficacy, and safety will be evaluated, if the data allow.

4.4.3 Exploratory analyses

The following exploratory endpoints will be examined (note; some of these may be reported in a separate document outside of the CSR):

- The proportion of patients with NED at 15, 24 and 48 months after randomization/allocation will be summarised.
- Resource use will be summarised descriptively.
- Biomarkers will be analysed using appropriate summaries of the exploratory outcome variables and data listings will be produced and compared across the treatment arms.

Graphical methods will be widely used in exploring the characteristics and relationships of outcome variables.

5 INTERIM ANALYSES

5.1 Interim Analyses

One interim analysis will be performed for PFS and one interim analysis will be performed for OS:

- The first DCO, an interim PFS analysis, will occur when approximately 86% of the target number of PFS events is expected to be reached for the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population **and** non-*tBRCAm* ITT population (ie, approximately 128 of 149 PFS events across Arm 3 and Arm 1 in non-*tBRCAm* HRD-positive population and 390 of 453 PFS events in the non-*tBRCAm* ITT population). It is anticipated that approximately 86% of the target number of PFS events (ie, approximately 414 of 480 PFS events) will be available for the comparison of Arm 2 vs Arm 1 in the non-*tBRCAm* ITT population at that time. In addition, a descriptive analysis of OS will also occur at this time. This interim will occur approximately 43.5 months after the first patient is randomised
- The second DCO, primary analysis of PFS and an interim OS analysis, will occur when approximately 149 PFS events have occurred (58% maturity) for the comparison of Arm 3 vs Arm 1 in non-*tBRCAm* HRD-positive population **and** approximately 453 PFS events have occurred (62% maturity) for the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* ITT population (approximately 52 months after the first patient is randomised). It is anticipated that 480 PFS events will have occurred (65% maturity) for the comparison of Arm 2 vs Arm 1 in the non *tBRCAm* ITT population at that time.
- The third and final DCO, the final OS analysis, is planned to occur at approximately 50% OS maturity across the 3 treatment arms in the non-*tBRCAm* ITT population or 5 years following randomization of the last non-*tBRCAm* patient (approximately 86 months after the first patient is randomised), whichever occurs sooner.

For PFS, the 2.5%/5% alpha (two-sided) for the secondary PFS endpoints will be controlled at the interim and the final analysis timepoints separately for each PFS comparison by using a bespoke spending function, where a fixed significance level will be assigned at the interim for the Arm 3 vs Arm 1 non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population and for Arm 2 vs Arm 1 non-*tBRCAm* ITT population of 0.0022, 0.008 and 0.005, respectively and the remaining significance level assigned to the final analysis, taking account of correlation.

If 86% of the target events are available at the time of the interim analysis for each respective treatment comparison/population, the two-sided significance level to be applied for the final analysis for the Arm 3 vs Arm 1 non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population and for Arm 2 vs Arm 1 non-*tBRCAm* ITT population would be ~0.05, 0.0497, 0.0246, respectively.

Note, both the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population and in the non-*tBRCAm* ITT population will be required to have met their respective statistical threshold in order for the study to be unblinded at the time of the interim analysis of PFS.

For OS, the 2.5%/5% alpha (two-sided) would be controlled at the interim and the final OS analysis timepoints separately for each comparison by using Lan De Mets O'Brien-Fleming spending function, where the significance level applied at the interim analysis depends upon the proportion of information (ie, information fraction) available.

With an alpha level of 2.5%, if 63% of the target events are available at the time of the interim OS analysis (ie, 81 of 129 OS events have occurred), then the two-sided significance levels of 0.00324, and 0.02396 will be applied to the interim and final analysis for OS for the Arm 3 vs Arm 1 non-*tBRCAm* HRD-positive population, respectively.

With an alpha level of 2.5%, if 63% of the target events are available at the time of the interim OS analysis (ie, 231 of 369 OS events have occurred), then the two-sided significance levels of 0.00319, and 0.02397 will be applied to the interim and final analysis for OS for the Arm 3 vs Arm 1 non-*tBRCAm* ITT population, respectively.

With an alpha level of 2.5%, if 63% of the target events are available at the time of the OS interim analysis (ie, 237 of 379 OS events have occurred), then the two-sided significance levels of 0.00317, and 0.02398 will be applied to the interim and final analysis for OS for the Arm 2 vs Arm 1 non-*tBRCAm* ITT population, respectively.

The final PFS and the interim/final OS analysis boundaries will ultimately be derived based on the actual number of events observed in the study; those referenced above are provided as examples only.

The significance level for the PFS and OS analyses will be calculated using the statistical software package EAST by specifying the information fraction for each analysis. The information fraction is calculated as the number of events at the analysis time-point divided by the total number of events at the final analysis time-point.

Additional analyses of PFS and/or OS may also be performed to meet Regulatory Agency requests, as required.

5.2 Data monitoring committee (DMC)

This study will use an external IDMC to assess ongoing safety of study participants by periodic reviews of accumulated safety data. Based on the review of data, IDMC will provide AZ and the TSC with recommendations for action with respect to study conduct and the management of all patients treated in the study. This committee will be composed of 2 physicians with expertise in ovarian cancer and its treatment options including small molecule inhibitors and immunotherapies and a statistician, who are independent of AstraZeneca and do not have any major conflict of interest.

The committee will meet within a month of First Subject Initiated to review safety data from the MEDIOLA (D081KC00001) triplet (triplet combination of olaparib, durvalumab and bevacizumab) cohort study (Phase 2, single arm, open label) which should have at least 10 patients with advanced gBRCAwt ovarian cancer enrolled with at least 4 weeks exposure. The second IDMC meeting to review data will occur when 30 patients in the non-*tBRCAm* cohort of DUO-O have completed at least 2 cycles of chemotherapy, bevacizumab and durvalumab/placebo. Patients from the *tBRCAm* cohort will be reviewed at the same time. Subsequent meetings through the end of the study and until the AZ study team is unblinded, unless the IDMC decide there is a reason to meet more frequently, is provided in the IDMC charter.

During regular data reviews, the IDMC will also separately assess the safety of the combination therapy in Japanese patients (specifically in relation to ILD and pneumonitis) as well as safety data for patients in the China cohort. The schedule of review is detailed in the IDMC charter.

An IDMC-DCO data cut-off date will be established for each IDMC meeting. Once the IDMC-DCO date is established, all data with a visit prior or equal to the IDMC-DCO date will be included in the statistical summaries for the IDMC meeting. All efforts will be made to enter and clean the data with visit dates equal to or earlier than the IDMC-DCO date within the allocated pre-agreed time.

Following each meeting, the IDMC will report to the sponsor and TSC chair and may recommend changes in the conduct of the study. The IDMC will recommend whether the study should continue as planned, be terminated for safety reasons or continue with appropriate modifications (See IDMC charter for suggested reasons). Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca and TSC chair. 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 and TSC.

In addition, the IDMC will meet for the interim PFS analysis, which will occur when approximately 86% of the target number of PFS events for the comparison of Arm 3 vs Arm 1 in the non-*tBRCAm* HRD-positive population and non-*tBRCAm* ITT population have been reached (approximately 43.5 months after the first patient has been randomised). For the interim analysis, the IDMC will review unblinded interim data and inform the sponsor whether the interim boundaries specified in Section 5.1 are met.

Full details of the IDMC procedures and communication process concerning all safety reviews and the PFS interim analysis can be found in the IDMC Charter.

6 CHINA COHORT

The global recruitment into the non-*tBRCAm* cohort of this study will close to all sites apart from China when approximately 1104 patients have been randomised. Any patient from China, randomised before the global recruitment is closed (ie, last subject randomised from a

non-Chinese site) will be included in both the global ITT population and the China cohort ITT population. A patient randomised in China after the global recruitment closure will be included only in the China cohort ITT population.

Approximately 120 patients from sites in China will be recruited and randomised in a 1:1:1 ratio to the study treatments in the non-*tBRCAm* cohort and will follow the same study plan and procedures as patients recruited to the global study. The safety and efficacy data collected will be summarised and analysed separately to the global study safety and ITT analysis sets (as defined in Section 2.1).

The primary analysis of efficacy for the China cohort will be an assessment of programmatically derived PFS based on investigator assessments (RECIST 1.1) in the China cohort ITT population (China FAS). The China FAS comprises all patients from sites in China who are randomised regardless of whether they receive treatment or not. Analyses of the China cohort will be performed at the same calendar time when approximately 17 PFS events have occurred (61% maturity) for the comparison of Arm 3 vs Arm 1 in the non-tBRCAm HRD-positive population and approximately 51 PFS events (64% maturity) for the comparison of Arm 3 vs 1 in the non-tBRCAm ITT population. At this time approximately 54 PFS events (68% maturity) are expected to have occurred for the comparison of Arm 2 vs 1 in the non-tBRCAm ITT population. Where data permit, summaries and analysis of secondary supportive efficacy endpoints (including at least but not limited to OS) will be performed for the China cohort. The detailed analysis plan will be documented in the China supplementary SAP.

When assessing safety and tolerability, summaries will be produced separately for the China cohort based on the China Safety Analysis Set. The China safety analysis set includes all subjects from sites in China who receive at least one dose of randomised investigational treatment (ie, durvalumab/olaparib), including placebo. The China safety data will be summarised descriptively and will not be formally analysed.

7 CHANGES OF ANALYSIS FROM PROTOCOL

Not applicable

8 REFERENCES

Aaronson et al 1993

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JCJM, Kaasa S, Klee MC, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw KCA, Sullivan M, Takeda F. The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute 1993; 85: 365-376.

Amit et al 2010

Amit O, Bushnell W, Dodd L, Roach N, Sargent D. Blinded independent central review of the progression-free survival endpoint. *Oncologist,* 2010, *15*(5), 492-495.

Berry et al 1991

Berry G, Kitchin RM, Mock PA. A comparison of two simple hazard ratio estimators based on the logrank test. Statistics in Medicine 1991; 10:749-755.

Burger et al 2013

Burger RA, Brady MF, Rhee J, Sovak MA, Kong G, Nguyen HP, et al. Independent radiologic review of the Gynecologic Oncology Group Study 0218, a phase III trial of bevacizumab in the primary treatment of advanced epithelial ovarian, primary peritoneal, or fallopian tube cancer. Gynecol Oncol 2013;131:21-26.

Collet 2003

Collet D. (2003) *Modelling Survival Data in Medical Research 2nd edn.* Chapman & Hall/CRC.

Eisenhauer et al 2009

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009 Jan;45(2):228-47.

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.

Fayers et al 2000

Fayers PM, Machin D. Quality of Life: Assessment, Analysis and Interpretation. J Wiley & Sons Ltd, Chichester, 2000. ISBN: 0-471-96861-7.

Fayers et al 2001

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.

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

Gail and Simon 1985
Gail M, Simon R. Testing for qualitative interactions between treatment effects and subject subsets. Biometrics. 1985; 41:361-72

Greimel et al 2003
Greimel E, Bottomley A, Cull A, et al. on behalf of the EORTC Quality of Life Group and the Quality of Life Unit. An International Field Study of the Reliability and Validity of a Disease-Specific Questionnaire Module (the QLQ-OV28) in Assessing the Quality of Life of Patients with Ovarian Cancer. Eur J Cancer 39:1402-1408, 2003.

Jackson et al 2014
Jackson D, White IR, Seaman S, Evans H, Baisley K, Carpenter J. Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation, Statist. Med. 2014; 33(27): 4681-94

Hsu and Taylor 2009
Hsu CH and Taylor J: Nonparametric comparison of two survival functions with dependent censoring via nonparametric multiple imputation, Statist. Med. 2009; 28:462–475

Newcombe 1998
Newcombe RG. Two-sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods, Statistics in Medicine, 1998, 17, 857–872.

Norquist et al 2018
Norquist BM, Brady MF, Harrell MI, Walsh T, Lee MK, Gulsuner S, et al. Mutations in Homologous Recombination Genes and Outcomes in Ovarian Carcinoma Patients in GOG 218: An NRG Oncology/Gynecologic Oncology Group Study. Clin Cancer Res. 2018 Feb 15;24(4):777-783.

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.

Perren et al 2011
Perren TJ, Swart AM, Pfisterer J, Ledermann JA, Pujade-Lauraine E, Kristensen G, et al; ICON7 Investigators. A phase 3 trial of bevacizumab in ovarian cancer. N Engl J Med. 2011 Dec 29;365(26):2484-96.

Ray-Coquard et al 2019
Ray-Coquard I, Pautier P, Pignata S, Pérol D, González-Martín A, Berger R, Fujiwara K, Vergote I, Colombo N, Mäenpää J, Selle F, Sehouli J, et al; PAOLA-1 Investigators. Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer. N Engl J Med. 2019 Dec 19; 381:2416-2428

Stone 2010
Stone A. The application of bespoke spending functions in group-sequential designs and the effect of delayed treatment switching in survival trials. Pharmaceut Statist 2010; 9:151–161.

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

van Hout et al 2012
van Hout B., Janssen,M.F., Feng,Y.S., Kohlmann,T., Busschbach,J., Golicki,D., Lloyd,A., Scalone,L., Kind,P., Pickard,A.S. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health. 2012 Jul-Aug;15(5):708-15

9 APPENDIX

9.1 Scoring algorithm for the EORTC-QLQC30

On the EORTC QLQ-C30 version 3 questionnaire, all items from 1 to 28 (Table 9) are four-point scales with response categories 1 = "Not at all", 2 = "A little", 3 = "Quite a bit" and 4 = "Very much." Items 29 and 30 that constitute the rating of a patient's overall health status and quality of life are assessed on a 7-point Likert scale 1 – 7 with 1 = "Very poor" and 7 = "Excellent".

All of the scales and single-item measures range in score from 0 to 100 (Fayers et al 2001) A high scale score represents a higher response level, that is, a higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms. Thus;

- A high score for a functional scale represents a high / healthy level of functioning,
- A high score for the global health status / QoL represents a high QoL,
- A high score for a symptom scale / item represents a high level of symptomatology / problems.

Table 9 Scoring the QLQ-C30 version 3.0

Domain/Scale	Number of items	Item numbers
Global health status / QoL		
Global health status/QoL	2	29, 30
Functional scales		
Physical functioning	5	1 to 5
Role functioning	2	6, 7
Emotional functioning	4	21 to 24
Cognitive functioning	2	20, 25
Social functioning	2	26, 27
Symptom scales / items		
Fatigue	3	10, 12, 18
Nausea and vomiting	2	14, 15

Pain	2	9, 19
Dyspnoea	1	8
Insomnia	1	11
Appetite loss	1	13
Constipation	1	16
Diarrhoea	1	17
Financial difficulties	1	28

The principle for scoring these scales is the same in all cases:

Step 1: Estimate the average of the items that contribute to the scale; this is the raw score.

If items $Q_1, Q_2, \ldots, Q_n$ are included in a scale, then for that scale, the $RawScore, RS$, is derived as the mean of the component items;

$$RawScore \ = \ RS \ = \frac{(Q_1 + Q_2 + \cdots + Q_n)}{n}$$

Step 2: Use a linear transformation to standardise the raw score, so that scores range from 0 to 100;

Then for Functional scales;

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

and for Symptom scales/items and Global health status / QoL;

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

Note: $range$ is the difference between the maximum possible value of the raw score RS and the minimum possible value. Therefore, the range of RS equals the range of the item values. For most items (1 to 28) scored 1 to 4, the $range$ is 3 (=4-1). The exceptions are the items contributing to the global health status / QoL, which are 7-point questions with $range$ = 6

Example:

Emotional functioning which is derived from items Q_{21}, Q_{22}, Q_{23} and Q_{24},

$$RawScore \ = \ RS \ = \frac{(Q_{21} + Q_{22} + Q_{23} + Q_{24})}{4}$$

and then the Emotional functioning transformed score

$$Score \ = \left\{1 - \frac{(RS - 1)}{3}\right\} \times 100$$

Fatigue which is derived from items $Q_{10}, Q_{12},$ and Q_{18},

$$RawScore \ = \ RS \ = \frac{(Q_{10} + Q_{12} + Q_{18})}{3}$$

and then the Fatigue functioning transformed score

$$Score = \left\{\frac{(RS-1)}{3}\right\} \times 100$$

Missing data

Missing data may be classified as either missing items (one or more missing answers to questions within a questionnaire), or missing forms (the whole questionnaire is missing for a patient).

Fayers and Machin (2000) (Fayers et al 2000) describe methods of analysis to use when data on response to items are missing, including imputation techniques. This involves applying the equations already given under "Scoring procedures" for calculating the scale scores; the missing items in multi-item scales are simply ignored when making the calculations when less than half the items are completed.

The protocol for missing items then can be summarized as;

- Have at least half of the items from the multi-item scale been answered?
  - If Yes, use all the items that were completed, and apply the standard equations given on the previous pages for calculating the scale scores; ignore any items with missing values when making the calculations.
  - If No, set scale score to missing.
- For single-item measures, set score to missing

No imputation or analytical techniques will be applied to account for data from missing forms.

9.2 Scoring algorithm for the EORTC-QLQ-OV28

The EORTC-QLQ-OV28 like the core questionnaire (EORTC QLQ-C30 version 3), has all items from 1 to 28 (when used in conjunction with the core questionnaire is numbered 31 to 58; Table 10) are four-point scales with response categories 1 = "Not at all", 2 = "A little", 3 = "Quite a bit" and 4 = "Very much." (Greimel et al 2003).

Table 10 Scoring the QLQ-OV28

Multi-item scales	Number of items	Item numbers
Abdominal / GI symptoms	7	31 - 37
Peripheral neuropathy	3	41 - 43
Other chemotherapy side effects	7	38 - 40, 44 - 47
Hormonal / menopausal	2	48, 49
Body image	2	50, 51
Attitude towards disease / treatment	3	52 - 54
Sexuality	2/4	55 - 58

Scoring procedure and adjusting for missing items and forms for the OV28 follows the procedure for the core questionnaire (EORTC-QLQ-C30) except for the following rules;

- For the scaling of the sexuality functioning, the following rules should be used:
  - If item 56 is answered with 1= ("not at all"), then the Sexuality scale is composed of items 55 and 56 only. Otherwise Sexuality is scored over items 55, 56, 57 and 58.
  - The response categories of the sexuality items are ordered:
    - Items 55, 56, 57: low to high
    - Item 58: high to low.

    Therefore, the last question should be reversed before calculating the raw mean scores. The resulting sexuality functioning scale (0-100) should be ordered with 0 indicating lowest functioning and 100 equal to the highest level of functioning.
- The 7 items constituting "other chemotherapy side effects" may not be highly intercorrelated depending on the treatment given. In case of doubt, it is advised to investigate these 7 items individually.

9.3 Scoring algorithm for the PRO-CTCAE

The items selected for this study will be assessed relative to one or more distinct attributes, including presence/absence, frequency, severity, and/or interference with usual or daily activities. For each item, responses are provided on a 5-point Likert scale with corresponding response choices for frequency, for severity and interference (Table 11). There are no guidelines yet established for how to combine attributes into a single score so each item will be evaluated separately based on the attributes specified in Table 11.

Table 11 Item structure and attribute

Item: Statement of request on attribute	Response	Description of response
Frequency: How **OFTEN** did you have…?	0	Never
	1	Rarely
	2	Occasionally
	3	Frequently
	4	Almost constantly
Severity: What was the severity of your…at its **WORST**?	0	None
	1	Mild
	2	Moderate
	3	Severe
	4	Very severe
Interference: How much did …**INTERFERE** with your usual or daily activities?	0	Not at all
	1	A little bit
	2	Somewhat
	3	Quite a bit
	4	Very much

Missing data

No imputation will be made for missing items. The proportion of missing data should also be summarized to aid interpretation.

9.4 COVID-19

9.4.1 Definitions and Derivations

The following definitions will be used to identify patients with coronavirus disease 2019 (COVID-19) adverse events:

- **Confirmed or Suspected COVID-19 AEs**: All AEs within the AE search criteria developed by the latest MedDRA MSSO guidance for COVID-19 have been reviewed by AstraZeneca Global Patient Safety and a list of terms pre-defined. A further review will take place prior to DBL to ensure any further terms not already included are captured.

- **COVID-19 Associated Adverse Events:** All confirmed and suspected COVID-19 AEs defined above, **plus** all other AEs occurring within <7 days before and <30 days after the start date of all the **confirmed** COVID-19 events.

9.4.2 Presentation

Depending on the extent of any impact, summaries of data relating to patients diagnosed with coronavirus disease 2019 (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, by treatment group, including:

- Disposition (discontinued treatment due to COVID-19 and withdrew study due to COVID-19)

- Deviations (indicating whether a deviation is due to COVID-19 or not)

- Summary of disruptions due to COVID-19 (visit impact, drug impacted)

- Listing for patients affected by the COVID-19 pandemic

- Listing for patients with reported issues in the Clinical Trial Management System due to the COVID-19 pandemic.

In addition, if there is sufficient number of patients (e.g. ≥5 and/or ≥ 2% of the patient population) with an event of interest then the following may be generated, else AE listings/narratives will only be generated:

Demographic and baseline characteristics

- Summaries of demographics and baseline characteristics repeated within the subset of patients with confirmed/suspected COVID-19 infection.

- Summaries of medical history repeated within the subset of patients with confirmed/suspected COVID-19 infection.

Efficacy

- **PFS:** A sensitivity analysis may be conducted to assess for the potential impact of COVID deaths on PFS. This will be assessed by repeating the primary PFS analysis except that any patient 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

- **OS:** A sensitivity analysis to assess for the potential impact of COVID deaths on OS. This will be assessed by repeating the OS analysis except that any patient 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.

Safety

- The number of patients with Confirmed/Suspected COVID-19 infection and Confirmed/Suspected COVID-19 deaths
- Summary in subjects with / without Confirmed/Suspected COVID-19 infection
    - Overall TEAE summary table
    - TEAEs by SoC and PT by max grade
- Summary of TEAEs associated with COVID-19
- Summary of TEAEs (excluding AEs associated with COVID-19 infection)
- Summary of Confirmed/Suspected COVID-19 TEAEs
- Summary of TEAEs (excluding Confirmed/Suspected COVID-19 infection AEs)
- Summary of TEAEs associated with COVID-19 infection resulting in death
- Summary of TEAEs resulting in death (excluding AEs associated with COVID-19 infection)
- Summary of TEAEs associated with COVID-19 infection leading to study treatment discontinuation
- Summary of TEAEs leading to study treatment discontinuation (excluding AEs associated with COVID-19 infection)

Furthermore, patient narratives in all patients with Confirmed/Suspected COVID-19 SAEs will be generated. Also a separate AE listing of patients with Confirmed/Suspected COVID infection will also be generated, as required.