AstraZeneca

Statistical Analysis Plan

Study Code	D0816C00010
Edition Number	4
Date	30 October 2018

A Phase III, Open Label, Randomised, Controlled, Multi-centre Study to assess the efficacy and safety of Olaparib Monotherapy versus Physician's Choice Single Agent Chemotherapy in the Treatment of Platinum Sensitive Relapsed Ovarian Cancer in Patients carrying germline *BRCA1/2* Mutations

Statistical Analysis Plan
Study Code D0816C00010
Edition Number 4
Date 30 October 2018

A Phase III, Open Label, Randomised, Controlled, Multi-centre Study to assess the efficacy and safety of Olaparib Monotherapy versus Physician's Choice Single Agent Chemotherapy in the Treatment of Platinum Sensitive Relapsed Ovarian Cancer in Patients carrying germline *BRCA1/2* Mutations

Study Statistician [redacted]

Statistical Analysis Plan
Study Code D0816C00010
Edition Number 4
Date 30 October 2018

A Phase III, Open Label, Randomised, Controlled, Multi-centre Study to assess the efficacy and safety of Olaparib Monotherapy versus Physician's Choice Single Agent Chemotherapy in the Treatment of Platinum Sensitive Relapsed Ovarian Cancer in Patients carrying germline *BRCA1/2* Mutations

Global Product Statistician [redacted]

A Phase III, Open Label, Randomised, Controlled, Multi-centre Study to assess the efficacy and safety of Olaparib Monotherapy versus Physician's Choice Single Agent Chemotherapy in the Treatment of Platinum Sensitive Relapsed Ovarian Cancer in Patients carrying germline *BRCA1/2* Mutations

[redacted]

[redacted]

TABLE OF CONTENTS PAGE

LIST OF ABBREVIATIONS

The following abbreviations and special terms are used in this study Statistical Analysis Plan.

Abbreviation or special term	Explanation
AE	Adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
Baseline	Refers to the most recent assessment of any variable prior to dosing with study treatment
bd	Twice daily
BICR	Blinded Independent Central Review
BoR	Best Overall RECIST Response
BP	Blood pressure
BRCA	Breast Cancer susceptibility gene
BRCA mutation or *BRCA*m	Breast Cancer susceptibility gene mutation (see *gBRCA* mutation or *gBRCAm*)
CA-125	Cancer Antigen – 125
CI	Confidence Interval
CR	Complete response
CRF / eCRF	Case Report Form (electronic)
CSR	Clinical Study Report
CT	Computed tomography
CTC / CTCAE	Common Terminology Criteria for Adverse Event
CTSQ-16	Cancer Therapy Satisfaction Questionnaire
DAE	Discontinuation of Investigational Product due to Adverse Event
DCO	Data Cut Off
DBL	Database Lock
DNA	Deoxyribonucleic acid
DoR	Duration of response
dp	Decimal places
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group: A performance status using scales and criteria to assess how a patient's disease is progressing
EQ-5D-5L / EQ-5D	EuroQoL five dimensions, five level (EQ-5D-5L) health state utility index

Abbreviation or special term	Explanation
EWB	Emotional well being
FACIT	Functional Assessment of Chronic Illness Therapy
FACT-O	Functional Assessment of Cancer Therapy – Ovarian: A multidimensional questionnaire for patients with ovarian cancer
FAS	Full Analysis Set
FIGO	International Federation of Gynecology and Obstetrics
FSI	First Subject In
FWB	Functional well being
g*BRCA*	Germline *BRCA*
gBRCA mutation or *gBRCAm*	The term "*gBRCA* mutation" is used to refer to a germline *BRCA1* or *BRCA2* mutation classified as "deleterious" or "suspected deleterious" in accordance with the American College of Medical Genetics and Genomics recommendations for standards for interpretation and reporting of sequence variants
gBRCA wt	*gBRCA* wildtype
GCIG	Gynecologic Cancer Intergroup
HDU	High Dependency Unit
HR	Hazard Ratio
HRQoL	Health-related Quality of Life
IDMC	Independent Data Monitoring Committee
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weighting
IVRS	Interactive Voice Response System
KM	Kaplan Meier
LD	Longest diameter
MDAS	Measurable disease analysis set
Mg	Milligram
MRI	Magnetic resonance imaging
MTP	Multiple Testing Procedure
NCI	National Cancer Institute
NE	Not evaluable
NONMEM	Non-Linear Mixed Effects Modelling
NTL	Non-target lesions
OAE	Other Significant Adverse Event

Abbreviation or special term	Explanation
OR	Odds Ratio
ORR	Objective response rates
OS	Overall survival
PARP	Polyadenosine 5'diphosphoribose [poly (ADP ribose)] polymerisation
PD	Progressive disease
PFS / PFS1	Progression Free Survival
PFS2	Time from randomisation to second progression
PK	Pharmacokinetic
PKPD	Pharmacokinetic pharmacodynamic
PLD	Pegylated liposomal doxorubicin
po	Per os (by mouth, orally)
PS	Performance Status
PR	Partial response
PWB	Physical well being
Q	Question
QoL	Quality of Life
RECIST	Response Evaluation Criteria In Solid Tumours. This study will use RECIST version 1.1
RPSFT	Rank Preserving Structural Failure Time
SAE	Serious adverse event
SD	Stable disease
SDTM	Study Data Tabulation Model
SWB	Social well being
Study treatment	Olaparib or chemotherapy
TA	Therapeutic Area
t*BRCA* mutation or t*BRCA*m	The term "t*BRCA* mutation" is used to refer to a somatic tumour *BRCA*1 or *BRCA*2 mutation classified as "deleterious" or "suspected deleterious" in accordance with the American College of Medical Genetics and Genomics recommendations for standards for interpretation and reporting of sequence variants
TDT	Time from randomisation to study treatment discontinuation or death
TFST	Time from randomisation to first subsequent therapy or death
TL	Target lesions
TOI	Trial Outcome Index

Abbreviation or special term	Explanation
TSST	Time from randomisation to second subsequent therapy or death
ULN	Upper limit of normal
wt	Wildtype (patients without evidence of *BRCA*1 or *BRCA*2 deleterious or suspected deleterious mutations)

AMENDMENT HISTORY

Date	Brief description of changes
13 November 2017	1. Update the rules for missing RECIST visits to cover the time between the 8 week and 12 week schedules 2. Added clarification of how the missing visit rule will be implemented for the time to RECIST 1.1 or CA-125 progression or death 3. Clarification that ORR will be summarised for all patients in the FAS with measurable disease and also for all patients in the FAS 4. Clarification of the rules for censoring for PFS2, TFST and TSST 5. Updated to reflect the missing data rules for TOI and FACT-O and to add in details of the analysis of total FACT-O 6. Update of the duration of a cycle from 21 to 28 days 7. Clarification added for the conversion of the intended dose units for the dose intensity derivations 8. Addition of details on the plan for PK data 9. Addition of the details on the precision to which the p-values, hazard rations, odds ratios and confidence intervals will be presented 10. Clarification added on the derivation of age 11. Prior cytoreductive surgery removed from subgroups as this cannot be derived 12. Kaplan-Meier plot for investigator assessed data added for PFS 13. Change in the number of events required for subgroup analyses to be performed 14. Removal of a selection of sensitivity analysis tables, figures and listings to reduce the volume of outputs generated 15. Update to the criteria for the analysis of TOI 16. Summary of PGIC added 17. Correction of CTSQ-16 to indicate that this is assessed at Week 24 18. Clarification that treatment emergent AEs include those that worsen after the first dose of study medication 19. Clarification of the data cut-off process 20. Primary objective and endpoint changed from PFS to ORR 21. Number of randomised patients changed from 411 to 250 22. The data cut-off for the primary analysis will occur in January 2019 or at a minimum of 6 months after LSI, whichever is sooner. 23. Time to response was added as a secondary objective 24. Figure 1, study flow chart was updated 25. Section 1.3 Number of subjects was updated to reflect the change in sample size from 411 to 250 26. Section 2.1, A measurable disease analysis set was added for the primary endpoint, ORR; A PK analysis set was added.

Date	Brief description of changes
	27. Section 3.2 Outcome variables was updated to reflect changes based on the new primary objective and endpoint and time to response was added 28. Sections 3.2.2 and 3.2.5 PFS; the rules surrounding missed visits were clarified 29. Section 3.2.9. Clarification was added around the censoring for TFST and TSST (section 3.2.11) 30. Section 3.4.3. Rules were added for calculating doses 31. Section 4: Analysis methods were updated to reflect the change in number of patients and primary endpoint. Additional sensitivity and subgroup analyses were added for the new primary endpoint. 32. Section 4: Rules were added for imputing age where only a partial date of birth has been collected 33. Section 4.1.10 HRQoL MMRM analysis was amended to include data from week 48. 34. Section 4.1.13 Demographic tables added by measurable disease 35. Section 4.1.15 Additional information was added for a PK analysis
14 April 2018	1. Addition of on-treatment analyses for ORR and PFS 2. Addition of an unadjusted logistic regression analysis for ORR 3. Addition of an unadjusted Cox proportional hazards model for PFS 4. Addition of a forest plot for ORR 5. Addition of a pooling strategy for the stratification factors and clarification that the resultant pooled factors are the factors to be used in the analyses 6. Change to use the IVRS factors in the subgroup analyses for ORR and PFS and only perform subgroup analyses using the eCRF values in the case of mis-stratifications 7. Change of subgroup analyses for ORR from adjusted to unadjusted for consistency with PFS subgroup analyses and correction to use the measurable disease analysis set 8. Change in the definitions for the early and late discrepancy rate for ORR and PFS 9. Removal of the missing data rules for FACT-O as the scoring manual will be followed 10. A baseline TOI score by visit interaction term was added to the MMRM model for TOI and removal of the random intercept 11. CA-125 response added to the secondary objective table as an outcome measure 12. Justification removed from the number of subjects section and analysis methods section for the duration of follow up as not needed for the SAP

Date	Brief description of changes
	13. Table 1 formatting updated 14. Violations and deviations section simplified, minor deviations text removed and details on the deviation bias sensitivity analysis population moved to the corresponding section for the analysis 15. Update from best overall response to best objective response 16. Clarification added for the handling of missing data in terms of the best objective response 17. Clarification added to the section on independent review on adjudication and the data provided to the central reviewers 18. Clarification added that the sensitivity analysis of confirmed CR or PR will be conducted for both BICR and Investigator data 19. The formulae for the calculation of the PFS, TFST and TSST times have been added 20. Clarification added that if a patient has a missing baseline scan and the die within 2 visits of baseline, then this will be counted as a PFS event. 21. Clarification added for the censoring of DoR 22. Clarification added that alternative analysis methods may be explored if the FACT-O data has evidence of systematic missing values 23. Details adding from the scoring manual for CTSQ-16 to document how the domains will be derived 24. Stratified removed from the description of the logistic regression analysis in Table 8 25. CA-125 response moved to a separate row as a secondary endpoint in Table 8 26. Clarification added that if there are patients who have not had at least 2 prior lines of chemotherapy, then they will be grouped with patients who have had 2 prior lines of chemotherapy for the ORR analysis 27. Clarification added to define stratum for the subgroup analyses for ORR for use in the rule for performing analyses based on the number of responses 28. Clarification added to indicate that additional analyses may be performed to explore the impact of confounding factors for ORR 29. Clarification added for the analysis method for sensitivity analyses for ORR in terms of the use of stratification factors 30. Clarification added to explicitly state that BICR will be used for specific sensitivity analyses 31. Details of the multiplicity approach removed from the PFS2 and OS sections as this is repeated

Date	Brief description of changes
	32. Reference to summary of unblinding removed as this is an open label study
30 October 2018	1. Clarification added on the classification of the measurable disease analysis set in case of adjudication 2. Detail added for the two missed visit rule for PFS2 3. Clarfication on deriving CA-125 progression and response 4. Rules for missing TOI and total FACT-O scores 5. Update to handle missing date of birth 6. Subgroup added for PFS: measurable verus non-measurable disease at baseline 7. Update to the FACT-O mixed model for repeated measures anlaysis to only include visits performed whilst the patient is on treatment.

1. STUDY DETAILS

1.1 Study objectives

Primary Objective:	***Outcome Measure:***
• *To determine the efficacy of olaparib vs. physician's choice single agent chemotherapy by assessment of Objective Response Rate (ORR) using blinded independent central review (BICR)*	• *Objective Response Rate (ORR) by BICR using RECIST 1.1 criteria*

Secondary Objective:	**Outcome Measure:**
• *To compare the efficacy of single agent olaparib versus physician's choice single agent chemotherapy*	• *Progression Free Survival (PFS) by BICR using RECIST 1.1 criteria* • *Time from randomisation to second progression (PFS2) by investigator assessment of radiological, clinical or CA-125 progression* • *Overall Survival (OS)* • *Time to earliest progression by BICR RECIST 1.1 or CA-125 or death* • *Time from randomisation to first subsequent therapy or death (TFST)* • *Time from randomisation to second subsequent therapy or death (TSST)* • *Time from randomisation to study treatment discontinuation or death (TDT)* • *Duration of response (DoR) by BICR using RECIST 1.1 criteria for evaluable patients* • *Time to Response (TTR) by BICR using RECIST 1.1 criteria for evaluable patients* • *CA-125 response*
• *To compare the efficacy of single agent olaparib versus physician's choice single agent chemotherapy on the Health-related Quality of Life (HRQoL) as measured by the trial outcome index (TOI) of the Functional Assessment of Cancer Therapy – Ovarian (FACT-O)*	• *Mean change from baseline in TOI score* • *Proportion improved (in the absence of subsequent cancer therapy) in TOI score*

• *To assess efficacy of olaparib in patients identified as having a deleterious or suspected deleterious variant in either of the BRCA genes using variants identified with current and future BRCA mutation assays (e.g. gene sequencing and large rearrangement analysis)*	• *ORR (by BICR), PFS (by BICR), PFS2, OS, TDT, TFST and TSST, analyses will be performed in those patients whose gBRCAm status is confirmed by the central Myriad test (only required if population differs from the MDAS (for ORR) or FAS (for PFS))* • *Development and delivery of a BRCA mutation companion diagnostic*
• *To determine exposure to olaparib following dosing at the 300 mg bd tablet dose and explore exposure-response relationships*	• *Olaparib plasma concentration data. Population PK and PK-Pharmacodynamic (PD) analyses will be completed and reported separately from the clinical study report.*

Safety Objective:	**Outcome Measure:**
• *To assess the safety and tolerability of single agent olaparib vs. physician's choice single agent chemotherapy*	• *Adverse Events (AE), physical examination, vital signs including blood pressure (BP), pulse, electrocardiogram (ECG) and laboratory findings including clinical chemistry and haematology*

Exploratory Objective:	**Outcome Measure:**
• *To assess the effect on patient self-reported feelings about side-effects of single agent olaparib versus physician's choice of single agent chemotherapy using the 'Feelings about side-effects' domain of the Cancer Therapy Satisfaction Questionnaire (CTSQ-16)*	• *Treatment satisfaction score (as measured by the Satisfaction with Therapy scale of the CTSQ-16)* • *Patient-reported feelings measured by the 'feelings about side-effects' domain of the Cancer Therapy Satisfaction Questionnaire (CTSQ-16)*
• *To investigate the health economic impact of treatment and the disease on hospital related resource use and health state utility*	• *Number, type and reason of hospitalisations and hospital attendances, procedures undertaken and hospital length of stay* • *Health state utility derived from the HRQoL instrument, the EuroQoL EQ5D-5L*
• *To explore methods of estimating overall survival (OS) adjusting for the impact of the control arm receiving subsequent Polyadenosine 5'diphosphoribose [poly (ADP ribose)] polymerise (PARP) inhibitors or imbalances between the*	• *Overall survival adjusted for impact of subsequent PARP inhibitors (or other potentially active investigational agents (if appropriate, to support reimbursement appraisals)*

treatment arms for other potentially active agents	
• *To determine the frequency of and describe the nature of BRCA mutation/s in tumour samples and to compare this with germline BRCA mutation status*	• *BRCA1 and/or BRCA2 mutation status in tumour*
• *To explore whether resistance mechanisms to olaparib can be identified through analysis of tumour and blood samples – archival tumour (mandatory if available), blood samples at baseline and on disease progression (mandated) and serial biopsies at baseline and disease progression (optional)* • *Future exploratory research into factors that may influence development of cancer and/or response to treatment (where response is defined broadly to include efficacy, tolerability or safety) may be performed on the collected and stored archival tumour samples (mandatory if available), blood samples at baseline and on disease progression (mandated) and serial biopsies at baseline and disease progression (optional)* • *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)*	• *Potential retrospective tissue biomarker research*

The exploratory analyses may not be reported in the clinical study report (CSR). If not, they will be reported separately.

1.2 Study design

This open label, randomised, controlled, multi-centre study will assess the efficacy and safety of single agent olaparib vs. standard of care, based on physician's choice of single agent chemotherapy (i.e. weekly paclitaxel, topotecan, pegylated liposomal doxorubicin or gemcitabine) in relapsed ovarian cancer patients who have received at least 2 prior lines of platinum based chemotherapy, who have progressed at least 6 months after their last platinum based chemotherapy and who carry a germline deleterious or suspected deleterious *BRCA*

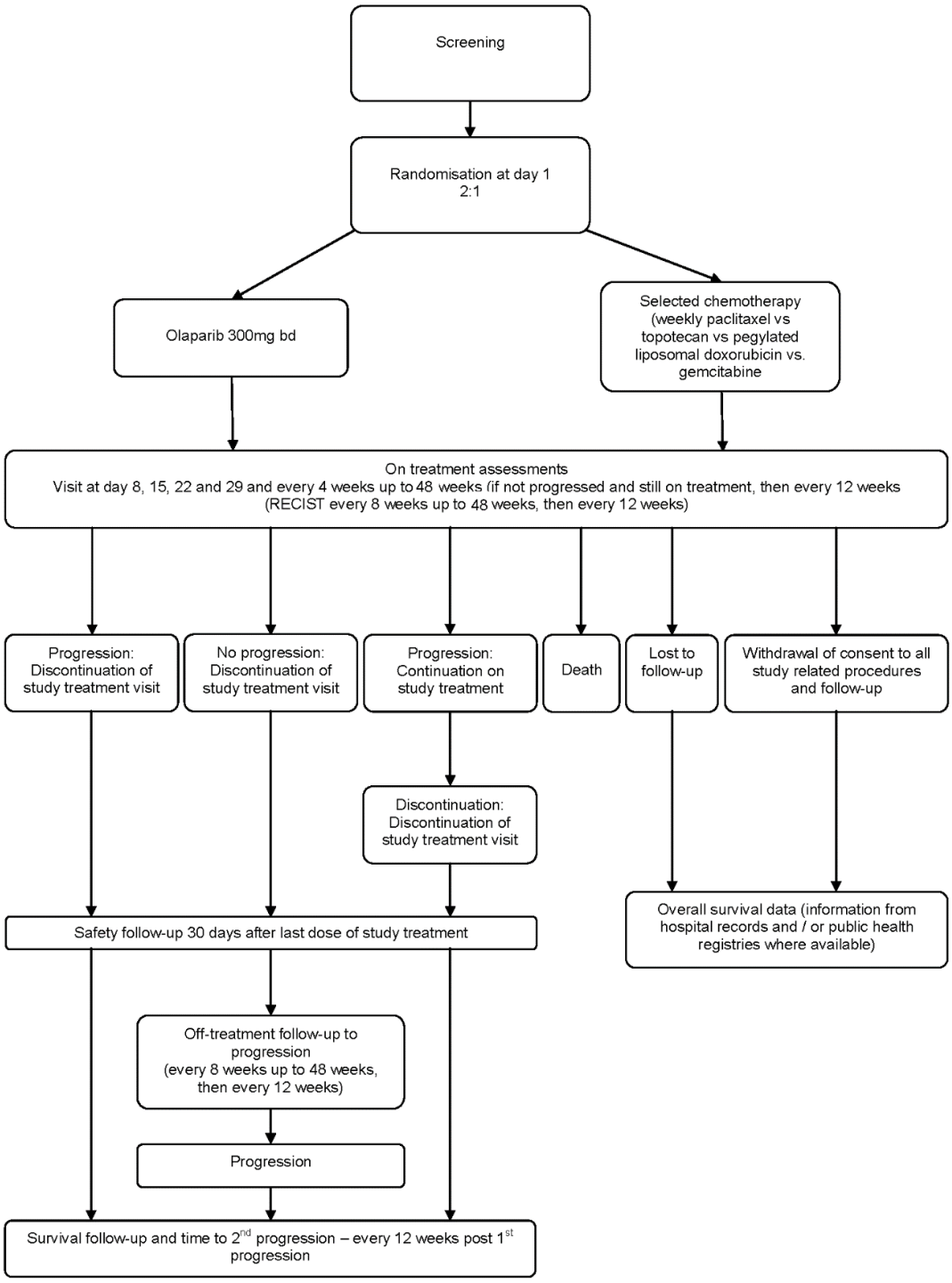
mutation. Non-platinum based chemotherapy in this setting can be given to prolong the platinum free interval and can be followed by further platinum treatment at a later relapse or can be considered for patients who are not warranted for further platinum treatment. Due to different routes and schedules of administration of the study treatments and different toxicity profiles, the study is not feasible to be blinded. Given the open label design of the study, rigorous methodology will be employed to ensure robustness of the primary endpoint assessment with a primary analysis of ORR based on blinded independent central review (BICR) of all patient scans for patients with measurable disease at baseline. Secondary endpoints will include progression free survival (PFS) by BICR, time from randomisation to second progression assessed by the investigator (PFS2), overall survival (OS), CA-125 response, safety assessments and health related quality of life (HRQoL).

The treatment groups include olaparib 300 mg po twice daily tablet continuously, or physician's choice of chemotherapy. The investigator must declare prior to randomisation their choice of chemotherapy, i.e. weekly paclitaxel, topotecan, pegylated liposomal doxorubicin or gemcitabine. The randomisation scheme will be stratified based on:

- Selected chemotherapy (weekly paclitaxel vs. topotecan vs. pegylated liposomal doxorubicin vs. gemcitabine)
- Received prior chemotherapy regimens for ovarian cancer (2 or 3 prior lines of chemotherapy vs. 4 or more)
- Time to disease progression after the end of the last platinum based chemotherapy (6-12 mo vs. > 12 mo)

The study flow diagram is presented in Figure 1.

Figure 1 Study Flow Chart



1.3 Number of subjects

The sample size for this study was selected to be consistent with the research hypothesis as described below.

Olaparib, administered as monotherapy improves objective response rate compared to physician's choice of single agent standard of care chemotherapy (weekly paclitaxel, topotecan, pegylated liposomal doxorubicin (PLD) or gemcitabine) in patients with relapsed platinum sensitive ovarian cancer who have received at least 2 prior platinum based lines of chemotherapy and carry *gBRCA* mutation.

The primary endpoint of the study is ORR. With at least 223 subjects with measurable disease at baseline, randomised 2:1 olaparib: chemotherapy, the study will have >80% power to show a statistically significant difference in ORR at the two-sided 5% level, assuming a response rate of 25% on the chemotherapy arm and at least 45% on the olaparib arm for subjects with measurable disease at baseline according to BICR. It is anticipated that approximately 90% of subjects will have measurable disease at baseline according to BICR and therefore to ensure adequate power, the sample size will have at least 250 subjects.

It is anticipated that the study recruitment period will be approximately 40 months and that the data-cut off for the primary analysis will occur in January 2019 or at a minimum of 6 months after LSI, whichever is sooner. No further analyses of ORR or PFS are planned beyond this point unless requested by Health Authorities.

2. ANALYSIS SETS

2.1 Definition of analysis sets

Table 1 gives a summary of outcome variables and analysis populations.

2.1.1 Full analysis set (FAS)

The full analysis set (FAS) will include all randomised patients and will compare the treatment groups on the basis of randomised treatment, regardless of the treatment actually received. Patients who were randomised but did not subsequently go on to receive study treatment are included in the Full Analysis Set (FAS).

2.1.2 Measurable disease analysis set (MDAS)

The measurable disease analysis set (MDAS) includes all patients in the FAS with measurable disease at baseline (as per RECIST 1.1)

- Measurable disease at baseline for the primary analysis is determined using BICR,

- If the patient was not adjudicated, then the primary reader will be used to determine whether or not the subject had target lesions at baseline
- If the patient was adjudicated, then the reader who the adjudicated selected will be used to determine whether or not the subject had target lesions at baseline

- For reporting the investigator response, the measurable disease analysis set will be determined using the investigator reported measurable disease at baseline

2.1.3 Safety analysis set

All patients who received at least one dose of randomised study treatment, olaparib or chemotherapy, will be included in the safety analysis set. If a patient receives at least one dose of olaparib study treatment they will be summarised in the olaparib arm for safety summaries (e.g. olaparib arm will include patients randomised to olaparib who receive at least one dose of olaparib or chemotherapy patients who receive at least one dose of olaparib study treatment in error at any time). If a patient randomised to olaparib receives only chemotherapy treatment then they will be summarised as part of the chemotherapy arm.

2.1.4 PK Analysis Set

The PK analysis set includes all patients who receive an olaparib dose and provide evaluable plasma concentration data. Examples of events which may affect PK data evaluability include study drug not taken according to protocol, disallowed surgical procedure, vomiting within 3 hours of dosing and use of disallowed concomitant medication. The population will be defined by the study team clinical pharmacology scientist, pharmacometrician and statistician prior to any analyses being performed.

Table 1 Summary of Outcome Variables and Analysis Populations

Outcome Variable	Populations
Efficacy Data	
- Primary : ORR by BICR	Measurable disease (MDAS)
- Secondary : PFS (by BICR), PFS2, OS, time to earliest progression by BICR RECIST 1.1, CA-125 or death, , CA-125 response, TFST, TSST, TDT, symptom/HRQoL endpoints	Full analysis set (FAS)
- Duration of response (DoR) by BICR	Measurable disease (MDAS)
Demography	Full analysis set (FAS)
Safety Data	
- Exposure	Safety
- Adverse Events	Safety
- Lab measurements	Safety

Table 1 Summary of Outcome Variables and Analysis Populations

Outcome Variable	Populations
- Vital Signs	Safety
Plasma concentration data	PK

2.2 Violations and deviations

Important protocol deviations will be listed and summarised by randomised treatment group. None of the deviations will lead to any patients being excluded from any of the analysis sets described in Section 2.1. The following general categories will be considered important deviations . This list is not exhaustive and the study team may highlight additional important protocol deviations at their discretion: :

- Patients randomised but who did not receive olaparib or chemotherapy.
- Patients who deviate from key entry criteria:
  - Female patients with histologically diagnosed relapsed high grade serous ovarian cancer (including primary peritoneal and/or fallopian tube cancer) or high grade endometrioid cancer (Inclusion criterion 3)
  - Documented mutation in BRCA1 or BRCA2 that is predicted to be deleterious or suspected deleterious (known or predicted to be detrimental/lead to loss of function) (Inclusion criterion 4)
- Baseline RECIST scan > 28 days before randomisation
- Baseline RECIST scan after study treatment is started
- Patients who receive a different chemotherapy regimen to that stated prior to randomisation. (Note that the sites are asked to specify intended chemotherapy ahead of randomisation taking place.)
- Patients who at some point receive the incorrect treatment (i.e. not their randomised treatment). The likelihood of this is expected to be low given the open nature of the trial and the different modes of administration of the randomised treatments.

  The summary will include all patients with a dispensing error but will also include information on how many of those patients received at least one dose of the wrong treatment (olaparib/chemotherapy) at any time. Patients who receive the wrong treatment at any time will be included in the safety analysis set as described in Section 2.1.

All patients who failed any inclusion/exclusion criteria will be listed along with details of the failed criteria. This information will also be summarised in terms of the number (%) of patients failing any of the inclusion/exclusion criteria and will be based on the FAS.

Other deviations may occur during the trial, which are not considered important and not believed to have any significant impact on the interpretation of the study results. All of these deviations will be recorded by the study monitors but will not be listed or summarised as part of the CSR.

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

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST visit responses

Patients with measurable or non-measurable disease assessed at baseline by CT/MRI will be entered in this study.

For all patients, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. It will also be used to determine if and when a patient has progressed in accordance with RECIST and also their best objective response.

Baseline radiological tumour assessments are to be performed no more than 28 days before randomisation and ideally should be performed as close as possible to the start of study treatment. Tumour assessments are then performed every 8 weeks (±1 week) up to 48 weeks and then every 12 weeks (±1 week) following randomisation until objective disease progression as defined by modified RECIST 1.1.

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

At each visit, an overall visit response will be determined by the BICR or programmatically derived from the data provided by the investigator (i.e. not the investigator opinion) - using the information from target lesions (TL), non-target lesions (NTL) and new lesions.

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 (except lymph nodes which must have short axis ≥ 15 mm) with CT or MRI and which is suitable for accurate repeated measurements.

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

All other lesions (or sites of disease) not recorded as TL should be identified as NTL 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).

Table 2 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 if any of the target lesions were not assessed or not evaluable or had a lesion intervention at this visit. 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

Rounding of TL data

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

Missing TL data

For a visit to be evaluable, all 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.

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 lesions 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 is also met i.e. if a lymph node LD increases by 20% but remains < 10 mm.
- Step 2: If some lesion 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 lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD
- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR

TL too big to measure

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

Irradiated lesions/lesion intervention

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

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

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD this will remain as a valid response category.

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and scale up as described below as long as there remain ≤ 1/3 of the TLs with missing measurements. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then if appropriate, a scaled sum of diameters will be calculated (as long as ≤ 1/3 of the TLs with interventions), and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention also has a value of 0 recorded. If scaling-up is not appropriate due to too few non-missing sizes then the visit response will be set as NE.

If ≤ 1/3 of the target lesion measurements have interventions 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-intervention lesions to the nadir sum of diameters excluding the lesions with interventions.

Example of scaling

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

Lesion 5 has had an intervention at the follow-up visit.

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

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

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

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 lesion that split.

Lesions that merge

If two target lesions 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 and MRI are the only methods of assessment that can be used within the trial. 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.

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.

Non-target lesion response will be derived based on the Investigator's overall assessment of NTLs as follows:

Table 3 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 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' on the basis of non-target lesions, there must be an overall level of substantial worsening in non-target disease such that, even in presence of SD or PR in target lesions, the overall tumour burden has increased sufficiently to merit discontinuation of therapy. A modest 'increase' in the size of one or more non-target lesions is usually not sufficient to qualify for unequivocal progression status.

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

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

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

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

A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.

If the question 'Any new lesions 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.

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

Table 4 Overall Visit Responses

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

Table 4 Overall Visit Responses

Target Lesions	Non-target lesions	New Lesions	Overall Response
NA	NE	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

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

3.1.4 Independent review

The independent review charter contains the details of the blinded independent central (BICR) review conducted by the AstraZeneca-appointed central Core Imaging Laboratory and will be developed in advance of the start of the study. The independent data review will provide RECIST measurements and response for each visit for each patient at the time of primary Data Cut Off (DCO). Prior radiotherapy will also be provided to the BICR to allow the selection of appropriate TLs. The imaging scans will be reviewed by two independent radiologists using RECIST 1.1 and will be adjudicated, if required (i.e. two reviewers' review the scans and adjudication is performed by a separate reviewer in case of a disagreement). After the primary analysis, BICR review of scans will no longer be required.

For each patient, the independent review will provide time point response data and the relevant scan dates for each time point (i.e. for visits where progression is/is not identified) with supporting measurements, assessments and clinical data (e.g. previous radiation reports) and no programmatic derivation of visit response is necessary.

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

3.2 Outcome variables

For the analyses that use BICR response data (CR, PR, SD, PD) the RECIST 1.1 assessments per BICR will be directly used. For secondary and sensitivity analyses based on investigator assessments at each visit, patients will be programmatically assigned a RECIST visit response of CR, PR, SD, PD, NE depending on the status of their disease compared to baseline and previous assessments.

3.2.1 Objective Response Rate (ORR)

Best overall RECIST response (BoR) is calculated based on the overall visit responses from each RECIST assessment (Table 4). It is the best response a patient has had during their time in the study up until RECIST progression or the last evaluable assessment in the absence of RECIST progression.

Categorisation of best overall response will be determined programmatically based on the RECIST criteria using the following response categories: complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD) and not evaluable (NE).

Best overall response will be determined programmatically from the time-point response from the BICR data. In addition, this will also be reported using investigator-recorded assessment.

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 8 weeks +/- 1 week, i.e. at least 49 days (to allow for 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 towar ds a particular overall visit assessment.

For patients whose progression event is death, BoR will be calculated based on data up until the last evaluable RECIST assessment prior to death.

For patients who die with no evaluable RECIST assessments, if the death occurred ≤ 17 weeks (i.e. 16 weeks ±1 week) after randomisation then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurred > 17 weeks (i.e. 16 weeks ±1 week) after randomisation, then BoR will be assigned to the non-evaluable (NE) category.

Progression events that have been censored due to them being >126 days (i.e. 16 weeks ± 7 days) after the last evaluable assessment the visit response of PD will not contribute to the BoR derivation..A patient will be classified as a responder if the RECIST 1.1 criteria for a CR or PR are satisfied at any time up to and including the defined analysis cut-off point. For each treatment group, the objective response rate (ORR) is the number of responders (patients with a CR or PR) divided by the number of patients in the measurable disease analysis set (MDAS). Only patients with measurable disease at enrolment can achieve an objective response of CR or PR which will not require confirmation for the primary outcome due to the randomised controlled study design as per the RECIST guidelines. However, a sensitivity analysis of confirmed CR or PR will be conducted for both BICR and Investigator data. A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit. Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue 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 (i.e. both visits contributing to a response must be prior to subsequent therapy for the patient to be considered as a responder).

In the case where a patient has two non-consecutive visit responses of PR, then, as long as the time between the 2 visits of PR is greater than 4 weeks and there is no PD between the PR visits, the patient will be defined as a responder. Similarly, if a patient has visit responses of CR, NE, CR, then, as long as the time between the 2 visits of CR is greater than 4 weeks, then a best response of CR will be assigned.

3.2.2 Progression free survival (PFS)

PFS is defined as the time from randomisation until the date of objective radiological disease progression according to RECIST 1.1 or death (by any cause in the absence of disease progression) regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy prior to disease progression (i.e. date of RECIST progression/death or censoring – date of randomisation + 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 evaluable RECIST assessment. However, if the patient progresses or dies after two or more missed visits, the patient will be censored at the time of the latest evaluable RECIST assessment, prior to progression.

Given the scheduled visit assessment scheme and change in scanning frequency after 48 weeks then the following rules will be used to define two missed visits:

- If the latest evaluable assessment was on or prior to Week 33/Day 231 (Week 32 + one week) then two missed visits will equate to more than 18 weeks (8 x 2 + 2)

- If the latest evaluable assessment was post Week 33/ Day 231 and on or prior to Week 47/Day 379 (Week 48 - one week) then two missed visits will equate to more than 22 weeks (8 + 12 + 2)

- If the latest evaluable assessment was post Week 47/Day 379 (Week 48 - one week) then two missed visits will equate to more than 26 weeks (12 x 2 + 2)

If the patient has no evaluable visits or does not have a baseline assessment they will be censored at day 1 unless they die within two visits of baseline (17 weeks allowing for visit window).

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:

(a) The 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 either reviewer where both select PD as time-point response and there is no adjudication for central review (BICR) data.

(b) 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.

(c) When censoring a patient for PFS the patient will be censored at the latest of the RECIST assessment/scan dates contributing to the last evaluable overall visit assessment

Overall visit assessments will be determined for each assessment (scheduled or unscheduled) and will contribute to the derivation of PFS.

Objective progression is defined as at least a 20% increase in the sum of the diameters of the target lesions (compared to previous minimum sum) and an absolute increase of >= 5 mm or an overall non-target lesion assessment of progression or a new lesion.

The PFS analysis will be based on the blinded independent central review (BICR) of the radiological scans. The BICR will be based on all RECIST assessment scan data provided by investigators, including the final RECIST assessment obtained from patients after progression has been determined according to RECIST 1.1 criteria by the investigator. A charter for the BICR will be developed in advance of the start of the study. A sensitivity analysis based on the programmatically derived PFS based on investigator-recorded assessments will be carried out.

The baseline RECIST assessment should be performed prior to randomisation but if an evaluable RECIST assessment occurs after randomisation but before treatment then this assessment will be used as the baseline assessment. If a patient does not have a baseline RECIST scan performed prior to the date of first dose of study treatment (olaparib/chemotherapy) then the patients will be censored at Day 1 in the analysis, unless they die within two visits of baseline (17 weeks allowing for visit window)

3.2.3 Time from randomisation to second progression (PFS2)

Time from randomisation to second progression is defined as the time from the date of randomisation to the earliest of the progression event subsequent to that used for PFS endpoint or death. The date of second progression will be recorded by the Investigator and defined according to local standard clinical practice and may involve any of; objective radiological, clinical, CA-125 progression or death. Second progression status will be reviewed every 12 weeks following the progression event used for the PFS endpoint (the first progression) and status recorded. Patients alive and for whom a second disease progression has not been observed should be censored at the last time known to be 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. Patients who die in the absence of progression will have their death recorded as a progression event for both PFS and PFS2. However, if the patient experiences a second progression or dies after two or more missed visits, the patient will be censored at the time of the last PFS2 assessment prior to the two missed visits. Two missed visits will equate to more than 26 weeks (12 x 2 + 2).

3.2.4 Overall survival (OS)

Overall survival is defined as the time from the date of randomisation until death due to any cause. 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 data cut-off (DCO) for the analysis, and if patients are confirmed to be alive or if the death date is post the DCO date these patients will be censored at the date of DCO. Death dates may be found by checking publicly available death registries.

3.2.5 Time to earliest progression by RECIST 1.1 or CA-125 or death

Progression or recurrence based on serum CA-125 levels will be defined on the basis of a progressive serial elevation of serum CA-125, according to the following modified GCIG criteria:

- For patients with elevated CA-125 on or before the date of randomisation (i.e. greater than the upper limit of normal (ULN)):-

  (a) If CA-125 does not fall to within the normal range after the date of randomisation then there must be evidence of CA-125 greater than, or equal to, 2 times the nadir value in the 28 day period before day 1 on 2 or more consecutive visits at least 1 week apart

  (b) Where CA-125 does fall to within the normal range after the date of randomisation (and the patient has not already progressed by way of (a) above) then there must be evidence of CA-125 greater than, or equal to, 2 times the ULN on 2 or more consecutive visits at least 1 week apart

- Patients with CA-125 in the normal range on or before the date of randomisation and no results greater than ULN on or before the date of randomisation must show evidence of CA- 125 greater than, or equal to, 2 times the ULN on 2 or more consecutive visits at least 1 week apart after the date of randomisation.

- CA-125 progression will be assigned the date of the first measurement after the date of randomisation that meets the above criteria.

Time to progression by RECIST or CA-125 progression or death is defined as the time from randomisation to the earlier date of RECIST (based on BICR) or CA-125 progression or death by any cause. Patients without a CA-125 progression or a RECIST progression who are still alive at the time of analysis will be censored at the time of their last evaluable RECIST assessment or their last available CA-125 measurement, whichever is the earliest at the time of the analysis. Since CA-125 is assessed more frequently than RECIST the two missed visit rule is based upon the RECIST schedule. Therefore if a patient dies, has RECIST progression or has CA-125 progression after two or more missed RECIST assessments, then the patient will be censored using the last evaluable RECIST assessment where CA-125 was also collected. This will be defined as a RECIST assessment where the date of CA-125 sample is +/- 11 days

(note the earliest date of the RECIST/CA-125 assessment will be used.). If only one assessment is missing during this period, no censoring is required.

3.2.6 Duration of Response (DoR)

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.

Patients who have not progressed or died following a response, will be censored at the time of the latest date of assessment from their last evaluable RECIST assessment. However, if the patient progresses or dies after two or more missed visits, the patient will be censored at the time of the latest evaluable RECIST assessment, prior to progression. If the patient has no evaluable visits or does not have a baseline assessment they will be censored at day 1.

3.2.7 Time to Response (TTR)

The time to response is defined as the time from randomisation until the date of first documented response. The date of first documented response should coincide with that used for the DoR endpoint.

Time to response will not be defined for those patients who do not have a documented response.

3.2.8 CA-125 response

Patients will be evaluable for CA-125 response (GCIG criteria; Rustin et al 2004; http://ctep.cancer.gov/resources/gcig/respdef_nov2005.doc**)** if:

- a pre-randomisation CA-125 level (taken within 2 weeks prior to the date of randomisation or on the date of randomisation) is at least twice the upper limit of normal, and
- there is no more than a 10% fall in CA-125 between the final 2 pre-randomisation samples (if 2 are taken)
- the same assay method is used for each sample from the same patient

A pre-randomisation sample will be defined to be a sample taken up to and including the day of randomisation. The protocol states that CA-125 samples on this day should be taken prior to dosing of study treatment and consequently prior to randomisation.

A response according to CA-125 will be considered to have occurred if there is at least a 50% reduction in CA-125 levels from the last pre-randomisation sample. The response must be confirmed and maintained for at least 28 days with no intervening change of >50%. Any intervening sample and the confirmatory sample must either be <= ULN or by <=1.1 x the initial response sample, if the initial response sample is > ULN. In addition, the confirmatory sample must be prior to any CA-125 progression event. The date when the CA-125 level is first reduced by 50% is the date of the CA-125 response. Note the GCIG criteria are not validated for this trial population.

A variable will be derived to indicate patients who have had a RECIST response or a CA-125 response.

3.2.9 Time to first subsequent chemotherapy or death (TFST)

As a supportive summary to PFS, time to start of first subsequent chemotherapy or death (TFST) will be assessed. TFST is defined as the time from the date of randomisation to the earlier of first subsequent chemotherapy start date or death (i.e. date of first subsequent cancer therapy / death or censoring – date of randomisation +1). Subsequent chemotherapies 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 therapy or death will be censored at the last known time to have not received subsequent chemotherapy, i.e. the last follow-up visit where this was confirmed. Patients still on study medication who have not received a subsequent therapy will be censored on the last recorded date on which the patient was known to be alive. Patients who have permanently discontinued study medication who have not received a subsequent therapy will be censored on the date of the last assessment reported on the TTSCAPRX form that indicates that the first subsequent therapy has not been received or the date of last dose if the patient had no TTSCAPRX form completed.

3.2.10 Time to study treatment discontinuation or death (TDT)

Time to study treatment discontinuation or death (TDT) will be assessed. TDT is defined as the time from randomisation to the earlier of the date of study treatment discontinuation or death. 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: Survival calls will be made in the week following the date of DCO for the analysis, and if patients are confirmed to be alive or if the death date is post the DCO date these patients will be censored at the date of DCO.

3.2.11 Time to second subsequent chemotherapy or death (TSST)

As a supportive summary to PFS2, time to start of second subsequent chemotherapy or death (TSST) will be assessed. TSST is defined as the time from the date of randomisation to the earlier of the second subsequent chemotherapy start date or death. (i.e. date of second subsequent cancer therapy / death or censoring – date of randomisation +1). Any patient not known to have had a further second subsequent therapy or death will be censored at the last known time to have not received second subsequent chemotherapy, i.e. the last follow-up visit where this was confirmed. Patients still on study medication who have not received a second subsequent therapy will be censored on the last recorded date on which the patient was known to be alive. Patients who have permanently discontinued study medication who have not received a second subsequent therapy will be censored on the date of the last assessment reported on the TTSCAPRX form that indicates that the second subsequent therapy has not been received or the date of last dose if the patient had no TTSCAPRX form completed.

3.3 Patient reported outcome variables

3.3.1 FACT-O

Patient-reported health-related quality of life (HRQoL) will be assessed using the FACT-O questionnaire (Basen-Enquist K et al 2001). The FACT-O is composed of the following sub-scales: physical, social/family, emotional, and functional well-being as well as the additional concerns ovarian cancer scale consisting of specific ovarian cancer symptoms.

The endpoint for HRQoL analysis will be the Trial Outcome Index (TOI) (Cella D et al 1993), an established single targeted index derived from the FACT-O questionnaire and it is considered to target the most relevant symptoms together with function and physical well-being and can be directly related to signs and symptoms and AEs. The TOI is composed of the following scales of the FACT-O: physical and functional well-being and additional concerns.

Data relating to the FACT-O will be self-reported through patient questionnaires according to the study plan. Patients will be asked to report their HRQoL over the course of the previous 7 days. All patients will be asked to complete the FACT-O. The FACT-O questionnaire will be administered at baseline, at Day 29 then in line with the RECIST assessments every 8 weeks (+/- 1 week) up to Week 48 regardless of treatment discontinuation or disease progression.

The Trial Outcome Index (TOI) score will be derived from the sum of the scores of the 25 items included in the physical well-being (7 items), functional well-being (7 items), and additional concerns ovarian cancer subscale (11 items [the subscale consists of 12 items, but item BMT7, asking about concerns about the ability to have children, should not be included in the calculations) of the FACT-O questionnaire version 4. The total FACT-O score will also be calculated which is made up of the sum of the individual subscale scores: physical well being (PWB), social well being (SWB), emotional well being (EWB), functional well being (FWB) and ovarian cancer subscale (Additional Concerns).

The scores will be derived in accordance with the FACT-O Scoring Manual. A number of items are negatively stated and need to be reversed by subtracting the response from “4”. The

scoring manual identifies that the following items need to be reversed prior to summarizing: GP1-7, GE1, GE3-6, O1-3, C2, and B5. After reversing proper items, the scores are derived.

For each subscale, if less than 50% of the subscale items are missing, the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscale. If at least 50% of the items are missing, that subscale also will be treated as missing.

For the TOI score, if more than 80% of all items within the subscales within the domains included have non-missing responses, the TOI score will be divided by the number of non-missing items and multiplied by 25 (the total number of items which could be included in the derivation of the endpoint). If less than or equal to 80% of items have non-missing responses, the TOI score will be treated as missing.

Similarly, for the total FACT-O score, if more than 80% of all items within the subscales within the domains included have non-missing responses, the score will be divided by the number of non-missing items and multiplied by the total number of items which could be included in the derivation of the endpoint. If less than or equal to 80% of items have non-missing responses, the score will be treated as missing.

The reason for any missing assessment will be identified. If data are missing at random, the above techniques will be used. If there is evidence that the missing data are systematic, missing values will be handled to ensure that any possible bias is minimised and alternative analysis methods may be explored. The TOI score ranges from 0-100 and the FACT-O from 0-152. For all Functional Assessment of Chronic Illness Therapy (FACIT) scales and symptom indices, a higher score indicates a higher HRQoL.

The actual change from baseline in TOI score will be derived for each visit where there is available data. For example; at visit X, the calculation will be (TOI score at visit X – Baseline TOI score). Actual change from baseline for the individual domain scores will be calculated in a similar way.

A change of at least 10 points in TOI score will be considered as a clinically relevant or a minimally important difference (Osoba et al 2005). In addition to deriving the visit response for TOI, the visit response for total FACT-O will be derived using a change of at least 15 points.

The population for analyses of HRQoL (TOI) will include a subset of the FAS population who have baseline TOI score. Similarly, the population for analyses of total FACT-O will include a subset of the FAS population who have baseline FACT-O score.

The definitions of the visit response for HRQoL are outlined below (Table 5):

Table 5 Health Related QoL Visit Response

Score	Change from baseline	Visit response
TOI	≥ +10	Improved
	≤ -10	Worsened
	Otherwise	No change
Total FACT-O	≥ +15	Improved
	≤ -15	Worsened
	Otherwise	No change

Best Overall TOI improvement (improvement in the absence of subsequent cancer therapy) will be defined as a change from baseline in the TOI of +10 points or more (Osoba et al 2005) sustained for at least 28 days, the denominator consisting of a subset of the FAS population who have baseline TOI. It will be derived as the best symptom improvement response the patient achieved, based on evaluable HRQoL data collected from randomisation up to the earliest of starting any subsequent cancer therapy or death. Therefore, the following criteria will be used to assign a best overall score response for each subject based on the individual visit responses (Table 6). Similarly, the best overall total FACT-O improvement will be derived.

Table 6 Health Related Quality of Life: Change rates - overall score.

Best Overall TOI (or total FACT-O) score response	Criteria
Improved	Two visit responses of "improved" a minimum of 28 days apart without an intervening visit response of "worsened"
No change	Does not qualify for overall score response of "improved". Two visit responses of either "no change" or "improved and "no change" a minimum of 28 days apart without an intervening visit response of "worsened"

Table 6 Health Related Quality of Life: Change rates - overall score.

Best Overall TOI (or total FACT-O) score response	Criteria
Worsened	Does not qualify for overall score response of "improved" A visit response of "worsened" without a response of "improved" or "no change" within 28 days
Other	Does not qualify for one of the above.

A TOI improvement rate (in the absence of subsequent cancer therapy) will be calculated as the % of all analysed patients with a best overall score response of improved. In the calculation of the proportion of patients that have a response of Improved, No Change or Worsened, the denominator used in the calculation will use the number evaluable for TOI. Similarly, a FACT-O improvement rate will be calculated.

Summary measures of overall compliance and compliance over time will be derived for the FACT-O form. These will be based upon:

- Received forms = number of FACT-O forms received back plus the number not received back where the reason was 'Subject too heavily affected by symptoms of disease under investigation'
- Expected forms = number of patients still under HRQoL follow-up at the specified assessment time excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under HRQoL follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.
- Evaluable forms = subset of expected FACT-O forms with at least one subscale that can be determined; or where REVPRDI form is ticked 'Subject too heavily affected by symptoms of disease under investigation'

Thus the overall compliance rate is defined as number of patients with an evaluable baseline and at least one evaluable follow-up form (as defined above), divided by the number of patients expected to have completed at least a baseline FACT-O form.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable baseline form and a form at the time point (as defined above), divided by number of patients still expected to complete forms at that visit. Similarly the evaluability rate over time will be calculated separately for each visit, including baseline,

as the number of evaluable forms (per definition above), divided by the number of received forms.

3.3.2 Cancer Therapy Satisfaction Questionnaire (CTSQ-16)

The CTSQ-16 is a 16-item measure that assesses satisfaction with and preference for chemotherapy treatment, and for biological therapy in either pill or IV form, across three domains: Feelings about Side Effects, Satisfaction with Therapy, and Expectations of Therapy (Trask et al 2008). The CTSQ-16 questionnaire is in Appendix G of the Protocol. The questions should be scored as 5 for the leftmost response and 1 for the rightmost response. In the CRF this is coded with a score of 1 for the leftmost response and 5 for the rightmost response. Therefore, all the scores will be reversed by subtracting the score from 6 prior to following the scoring instructions from the CTSQ scoring manual detailed below.

Questions (Q) 5, 6, 9 and 11 will be reversed again to calculate the domains by subtracting the score from 6. These reversed values will be referred to as Q5R, Q6R, Q9R and Q11R respectively. This means that Q5R, Q6R, Q9R and Q11R will be the values collected in the eCRF for this study and the remaining questions will be reversed compared to the value collected in the eCRF.

For each domain, if the number of completed items is greater than or equal to the minimum number indicated in Table 7, then the domain is derived using the following formula:

Domain score = [(sum of completed item responses/number of completed items)-1]x25
However, if fewer items are completed than the minimum number indicated in Table 7, then the domain is not scored and a missing value is assigned.

Table 7 Summary of domain scoring for CTSQ domains

CTSQ domain	Description of content of items in domain	Item numbers	Total number of items	Minimum number of completed items required to score
Expectations of therapy	Return to normal life, Get rid of cancer, Prevent cancer from coming back, Stop cancer from spreading, Help you live longer	Q1, Q2, Q3, Q4, Q8	5	3
Feelings about side effects	Cancer therapy limited daily activities, Upset about side effects, Taking cancer therapy as difficult as expected, Were side effects as expected	Q5R, Q6R, Q11R, Q13	4	4

Table 7 Summary of domain scoring for CTSQ domains

CTSQ domain	Description of content of items in domain	Item numbers	Total number of items	Minimum number of completed items required to score
Satisfactions with therapy	Worth taking even with side effects, Think about stopping cancer therapy, How worthwhile was cancer therapy, Benefits meet expectations, Satisf. with form of cancer therapy, Satisf. with recent cancer therapy, Would you take this cancer therapy again	Q7, Q9R, Q10, Q12, Q14, Q15, Q16	7	5

3.4 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.

3.4.1 Adverse events (AEs)

AEs and SAEs will be collected throughout the study, from informed consent until 30 days after the last dose of study treatment. Any untoward event occurring subsequent to the 30-day follow-up AE reporting period that the investigator assesses as possibly related to the study treatment should also be reported as an AE.

Other significant adverse events (OAEs)

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

Examples of these are marked 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.

3.4.2 Treatment exposure

The exposure to olaparib will be calculated for all patients who received olaparib, where the duration of exposure will be based on the planned administration as per protocol. The following will be calculated:

Total (or intended) exposure of olaparib

- Total (or intended) exposure = last dose date – first dose date + 1

Actual exposure of olaparib

- Actual exposure = intended exposure – total duration of dose interruptions, where intended exposure will be calculated as above. Dose interruptions will include those where a patient forgot to take a dose.

Number of days on 300 mg olaparib bd

- Number of days on 300 mg olaparib bd = actual exposure for the dose assigned.

Exposure to chemotherapy for patients in the chemotherapy arm will be measured by the number of cycles received. For all four choices of chemotherapy regimen, a cycle corresponds to a period of 28 days. If a cycle is prolonged due to toxicity, this should still be counted as one cycle.

3.4.3 Dose intensity

The dose intensity will be calculated for both patients who receive olaparib and for those who receive chemotherapy. Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to treatment discontinuation. Percentage intended dose (PID) is the percentage of the actual dose delivered relative to the intended dose through to progression. Both will be derived using study treatment data up to the date of objective disease progression as defined by RECIST using the investigator site assessments. If the investigator considered that it was in the patient's best interest to continue study treatment past this time, this will not be included in the derivation of dose intensity.

RDI and PID will be defined as follows:

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up to the earlier of progression (or a censoring event) or the actual last day of dosing and D is the intended cumulative dose up to the earlier of progression (or a censoring event) or the actual last day of dosing plus the protocol-defined post-dose rest period.

- PID = 100% * d/D, where d is the actual cumulative dose delivered up to progression (or a censoring event) and D is the intended cumulative dose up to progression (or a censoring event). D is the total dose that would be delivered, if there were no modification to dose or schedule.

For Chemotherapy, actual dose is the dose received reported in mg. Intended dose is the dose planned for the patient reported in mg/m^2 and converted to mg by multiplying by their body surface area (m^2) calculated from height and weight at baseline using the Mosteller formula. Due to the possibility of rounding of the BSA calculations at the sites, the following rules will be applied:

- For Paclitaxel, if the actual dose in mg/m^2 rounded to the nearest 10 mg/m^2 is equal to the planned dose in mg/m^2 then the planned dose will be assumed to be equal to the actual dose
- For PLD, if the actual dose in mg/m^2 rounded to the nearest 10 mg/m^2 is equal to the planned dose in mg/m^2 then the planned dose will be assumed to be equal to the actual dose
- For gemcitabine, if the actual dose in mg/m^2 rounded to the nearest 100mg/m^2 is equal to the planned dose in mg/m^2 then the planned dose will be assumed to be equal to the actual dose
- For topotecan, if the actual dose in mg/m^2 rounded to the nearest 1mg/m^2 is equal to the planned dose in mg/m^2 then the planned dose will be assumed to be equal to the actual dose

Example of Dose Intensity

			Study day									
RDI	**PID**	**patient**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
86%	60%	1	X	X	X	X	X	X	D			PD
70%	70%	2	X	X	X	X	X	X	X	O	O	O PD D
100%	100%	3	X	X	X	X	X	X	X	X PD D		

X: 300mg dose taken, O: dose not taken, PD: Progressive Disease, D: Discontinued

Patients 1-2 progressed on Day 10, so the intended dose through to progression was 10 * 600 mg of olaparib = 6000 mg and Patient 3 progressed on Day 8 and intended dose was 600*8=4800 mg.

Patient 1: received 600 mg of olaparib daily for 6 days, discontinued treatment for reasons other than progression on day 7, then progressed or died on day 10.

RDI = (6 * 600 mg)/ 4200 mg = 86%

PID = (6 * 600 mg)/ 6000 mg = 60%

Patient 2: received 600 mg of olaparib daily for 7 days, did not discontinue treatment, but nevertheless did not receive another dose, then progressed or died on day 10.

RDI = PID = (7 * 600 mg)/ 6000 mg = 70%

Patient 3: received 600 mg of olaparib daily for 8 days, and then progressed or died on same day 8.

RDI = PID = (8 * 600 mg)/ 4800 mg = 100%

3.4.4 Laboratory data

Blood and urine samples for determination of clinical chemistry, haematology, coagulation, and urinalysis will be taken at the times indicated in the study schedule (see Tables 1, 2 and 3 of the CSP). For derivation of post baseline visit values considering visit window and to handle multiple records, derivation rules as described in Section 3.4.7 below will be used.

3.4.5 Electrocardiograms (ECGs)

ECGs are required during screening within 7 days prior to starting study treatment and when clinically indicated afterwards.

All ECGs should be assessed by the investigator as to whether they are clinically significantly abnormal / not clinically significantly abnormal. If there is a clinically significant abnormal finding, the Investigator will record it as an AE on the eCRF.

3.4.6 Vital signs

Height will be assessed at screening only. Weight will be assessed at screening and as clinically indicated at any other time. Any changes in vital signs should be recorded as an AE, if applicable.

3.4.7 General considerations for safety assessments

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

- The time windows should be exhaustive so that data recorded at any time point has the potential to be 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 laboratory data are:

  – Day 8, visit window 2 – 11

  – Day 15, visit window 12-18

  – Day 22, visit window 19-25

  – Day 29, visit window 25 – 42

  – Day 57, visit window 43 - 71

  ...

- 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.

  – To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group visit data should only be 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 will be defined as the last non-missing measurement prior to dosing with study treatment (olaparib or chemotherapy). For laboratory data and vital signs data, any assessments made on 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 treatment (olaparib or chemotherapy)

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

3.5 Health Economic Endpoints

3.5.1 Hospital Resource Use

Hospital resource use variables include the following:

- Length of hospital stay
- Reasons for hospitalisation
- Length of any time spent in ICU/HDU
- Outpatient/daytime attendances

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

Sum of total duration of hospital stay will be considered for analysis if patient was admitted to hospital more than one time during study period.

The length of ICU stay will be calculated using the same method as detailed above for the length of hospital stay.

3.5.2 EQ-5D-5L (exploratory analysis)

The EQ-5D is a standardised measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care.

The EQ-5D-5L index comprises five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). For each dimension, respondents select which statement best describes their health on that day from a possible five options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems and unable to/ extreme problems). A unique EQ-5D health state is referred to by a five digit code allowing for a total of 3125 health states. For example, state 11111 indicates no problems on any of the five dimensions. This data will be converted into a weighted health

state index by applying scores from EQ5D value sets elicited from general population samples (the base case will be the UK valuation set, with other country value sets applied in scenario analyses). Where values sets are not available, the EQ-5D-5L to EQ-5D-3L crosswalk will be applied. In addition to the descriptive system, respondents also assess their health today on a visual analogue scale, ranging from 0 (worst imaginable health) to 100 (best imaginable health). This score is reported separately.

The evaluable population will comprise all patients who receive study treatment as per safety analysis set and have a baseline EQ-5D-5L assessment.

Further details of the analysis of EQ-5D will be given in the payer analysis plan.

3.6 Pharmacokinetic Endpoints

Pharmacokinetic (PK) sampling is to be performed in a subset of approximately 65 evaluable patients from the olaparib treatment group. The sampling times are: Day 1 pre-dose & 1 hour post-dose; Day 29 Pre-dose, 0.5-1 hour, 1-3 hours, 3-6 hours and 6-12 hours post dose.

The plasma concentration data will be listed and summarised (for Day 1 pre-dose & 1 hour post-dose and Day 29 Pre-dose time points only). Data from patients excluded from the PK analysis set will be included in the data listings, but not in the summary. Extra measurements (such as unscheduled or repeat assessments) will also not be included in summary table, but will be included in patient listings.

Plasma concentration data will be summarised using descriptive statistics, including n, arithmetic mean, SD, coefficient of variation (%CV), median, minimum, and maximum values. Additionally, geometric means and geometric %CV (%GCV) will be reported, the geometric mean will be calculated as the exponential of the arithmetic mean calculated from data on a log scale. The %GCV was calculated as $100\sqrt{(\exp(s^2)-1}$ where s was the SD of the data on a log scale. The plasma 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 were NQ, the mean, SD, geometric mean, %CV, and %GCV were calculated by substituting the limit of quantification (LOQ) for values which are NQ.

- If more than 50%, but not all, of the concentrations were NQ, the mean, geometric mean, SD, %CV, and %GCV were reported as not calculable (NC).

- If all the concentrations were NQ, the geometric mean and mean were reported as NQ and the SD, %CV, and %GCV as NC.

The PK analysis of the plasma concentration data for olaparib will be performed at AstraZeneca R&D or by a CRO identified by AstraZeneca R&D. The actual sampling times will be used in the PK calculations. Non-linear mixed effects modelling (NONMEM) will

evaluate the pharmacokinetic characteristics of olaparib, quantify variability in the pharmacokinetics, identify demographic or pathophysiological covariates which may explain the observed variability, estimate steady state C_{max}, AUC and C_{min} and explore exposure-response relationships. The pharmacokinetic analysis will be reported separately from the clinical study report.

4. ANALYSIS METHODS

The data cut-off for the primary analysis will occur in January 2019 or at a minimum of 6 months after LSI, whichever is sooner. No further analyses of ORR or PFS are planned beyond this point unless requested by Health Authorities.

An initial PFS, PFS2, OS, TFST, TSST and TDT analysis will be performed at the same time as the primary analysis of ORR. In addition, change from baseline in TOI and the TOI improvement rate will be analysed. At this time DoR and CA-125 response will be summarised descriptively. No adjustment will be applied for multiplicity for TFST, TSST, TDT and TOI as these are viewed as supportive endpoints.

4.1 Analysis methods

The treatment comparison is olaparib 300 mg bd vs chemotherapy.

ORR and DoR analyses will be performed on the measurable disease population (MDAS). All other efficacy analyses, including PFS, PFS2 and OS, will be performed on the FAS population. In addition to the main analyses of ORR, PFS, PFS2, OS, TDT, TFST and TSST, analyses of these endpoints will be performed in those patients whose gBRCAm status is confirmed by the Myriad test unless populations are the same.

Results of all statistical analysis will be presented using a 95% confidence interval and 2-sided p-value. Exact p-values should be presented using 3 decimal places for all values ≥ 0.001. The only categorical presentation to be used is when $p<0.001$. Hazard ratios, odds ratios and confidence interval limits are to be presented to 2 decimal places.

The following table details which endpoints are to be subject to formal statistical analysis, together with pre-planned sensitivity analyses making clear which analysis is regarded as primary for that endpoint.

Table 8 Formal Statistical Analyses to be Conducted and Pre-Planned Sensitivity Analyses

Endpoints Analysed	Notes	Analysis population
Objective Response Rate (Number of patients who have a CR or PR determined using RECIST 1.1 criteria divided by the	Primary analysis: Logistic regression using BICR assessment in the MDAS	MDAS
	Sensitivity / supportive analyses:	

Table 8 Formal Statistical Analyses to be Conducted and Pre-Planned Sensitivity Analyses

Endpoints Analysed	Notes	Analysis population
number of patients with measurable disease.)	1) Ascertainment bias analysis: Logistic regression using investigator assessments in the MDAS	MDAS
	2) Logistic regression based on BICR for all patients with evaluable disease. Patients with evaluable disease include all of those in the MDAS and patients with non-target lesions at baseline and non-measurable disease at baseline. (Only required if the evaluable disease population differs from the MDAS.).	Evaluable disease subset of FAS
	3) Logistic regression based on BICR for all patients in the FAS. (Only required if the FAS population differs from the MDAS).	FAS
	4) Logistic regression based on confirmed BICR responses in the MDAS.	MDAS
	5) Logistic regression based on confirmed investigator responses in the MDAS.	MDAS
	6) Deviation bias (if meaningful to do); Logistic regression using BICR data	MDAS (excluding patients with deviations that may affect the efficacy of the trial)
	7) On-treatment analysis: Logistic regression using BICR assessment for patients in the MDAS who received at least one dose of randomised treatment	All patients in the MDAS who received at least one dose of randomised treatment
	8) Unadjusted logistic regression using BICR assessment in the MDAS	MDAS
PFS (Time from randomisation to first progression or death)	Primary analysis: stratified log-rank test using BICR assessments	FAS
	Sensitivity/supportive analyses:	
	1) Evaluation time bias analysis: stratified log-rank test using BICR assessments	FAS
	2) Attrition bias analysis (using alternative censoring rules): stratified log-rank test using BICR assessments	FAS

Table 8 Formal Statistical Analyses to be Conducted and Pre-Planned Sensitivity Analyses

Endpoints Analysed	Notes	Analysis population
	3) Ascertainment bias analysis: stratified log-rank test using Investigator assessments	FAS
	4) On-treatment analysis: Stratified log-rank test using BICR assessment for patients in the FAS who received at least one dose of randomised treatment	All patients in the FAS that recevied at least one dose of randomised treatment
	5) Deviation bias (if meaningful to do); stratified log-rank test using BICR data	FAS (excluding patients with deviations that may affect the efficacy of the trial)
	6) A Cox proportional hazards model	FAS
	7) Unadjusted Cox proportional hazards model	FAS
PFS2 (Time from randomisation to second progression or death)	Primary analysis: stratified log rank test based on investigator assessment of second progression	FAS
Overall Survival (Time from randomisation to death due to any cause)	Primary analysis: stratified log-rank test Supportive analysis: KM plot of time to censoring for OS	FAS
Time to earliest progression by RECIST 1.1 or CA-125 or death	Primary analysis: stratified log-rank test using BICR data	FAS
TFST (Time to first subsequent therapy or death)	Stratified log rank test using eCRF data	FAS
TSST (Time to second subsequent therapy or death)	Stratified log rank test using eCRF data	FAS
TDT (Time to study treatment discontinuation or death)	Stratified log rank test using eCRF data	FAS
CA-125 response	CA-125 response per treatment arm (descriptive only); CA-125 and/or RECIST response per treatment arm (descriptive only) based on BICR for RECIST	FAS

Table 8 Formal Statistical Analyses to be Conducted and Pre-Planned Sensitivity Analyses

Endpoints Analysed	Notes	Analysis population
Change from baseline in TOI score	Primary : Mean change from baseline in TOI score analysed by MMRM Supportive : Proportion improved (in the absence of subsequent cancer therapy) in TOI score analysed using logistic regression	FAS

For the analysis of response rates and time to event endpoints, the following pooling strategy will be applied. If the number of responses / events in the individual stratum are too small for a meaningful analysis (less than 5 responses / events per stratum; *a stratum is defined as strata1*strata2*...strataX*treatment; so with 2 stratification factors of each 2 levels and two treatments we have 2*2*2=8 stratum)* stratification factors will be removed in the following order until there are at least 5 responses / events in each stratum: selected study chemotherapy (paclitaxel / topotecan / pegylated liposomal doxorubicin / gemcitabine); number of prior chemotherapy regimens received for ovarian cancer (2 or 3 vs. 4 or more); time to disease progression on last platinum based chemotherapy received prior to randomisation (6 – 12 months / > 12 months). This will be done for each individual endpoint to be analysed and consequently the strata used in the analysis may vary by endpoint. All the sensitivities for each endpoint, will use the same strata as the primary model, for that endpoint, unless there are <5 events per stratum and then an andjusted model will be used. If required, unadjusted sensitivity analyses of each of the endpoints may also be performed.

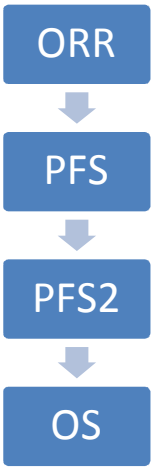
4.1.1 Multiplicity

In order to describe the nature of the benefits of olaparib maintenance treatment, ORR, PFS, PFS2, and OS will be tested at a 2-sided significance level of 5%.

However, in order to strongly control the type I error at 2.5% 1-sided, a multiple testing procedure will also be employed across the primary endpoint and secondary endpoints intended for key label claims (i.e. ORR, PFS, PFS2 and OS). There is no requirement to adjust for multiplicity due to ORR or PFS interim analyses, since there are no planned interim ORR or PFS analyses with the opportunity to make an early claim of efficacy.

A hierarchical testing strategy will be employed where ORR is tested first using the full test mass (full test mass = alpha 5% 2 sided) and key secondary endpoints of PFS, PFS2 and OS will then be tested using a multiple testing procedure with a recycling strategy (i.e., the MTP will recycle the test mass to the endpoint not yet rejected in the hierarchy outlined in Figure 2).

Figure 2 Multiple Testing Procedure



PFS will only be tested if the null hypothesis (of no difference) for ORR is rejected. PFS2 will only be tested if the null hypothesis (of no difference) is rejected for both ORR and PFS. OS will only be tested if the null hypothesis (of no difference) is rejected for ORR, PFS and PFS2. An additional PFS2 and OS analysis will only be conducted with further follow up (~60% OS events) if both ORR and PFS are statistically significant based on the primary analysis and the null hypotheses for PFS2 and/or OS are not rejected at the time of the primary analysis. If an additional analysis is conducted for PFS2 and OS, to control for multiple testing due to an interim and final analysis a Lan DeMets spending function (Lan and DeMets 1983) that approximates an O'Brien Fleming approach will be used to account for multiplicity.

If the null hypothesis is rejected for ORR and PFS then the 5% (2 sided) alpha level will be allocated to PFS2 for the interim and final analysis. Subsequently if the null hypothesis for PFS2 is rejected then OS will be controlled at the interim and primary time point by using the Lan DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the significance level applied at the interim depends upon the proportion of information available.

If both ORR and PFS are statistically significant based on the primary analysis and the null hypotheses for PFS2 and/or OS are not rejected at the time of the primary analysis, a final analysis of PFS2 and OS will take place when there are approximately 60% deaths (approximately 150 events). It is not known how many PFS2 events there will be at this final cut off. Therefore, for the calculation of the significance level for the interim analysis of PFS2 it will be assumed that the proportion of information available for PFS2 is the same as that for OS.

For example, if 50% of OS events required at the time of the primary OS analysis are available at the time of the interim (i.e., 75/150 events have occurred), the 2-sided significance level to be applied for the OS interim analysis would be 0.31% and the 2-sided significance

level to be applied for the final OS analysis would be 4.7%. The same 2-sided significance level would be used for PFS2.

If the number of OS events observed at the time of the interim is higher or lower than 75 the interim alpha will be adjusted accordingly.

4.1.2 Primary variable - Objective Response Rate (ORR)

The data cut-off for the primary analysis will occur in January 2019 or at a minimum of 6 months after LSI, whichever is sooner. No further analyses of ORR are planned beyond this point unless requested by Health Authorities.

For each treatment arm, Best Overall Response (BoR) will be summarised by n (%) for each category (CR, PR, SD, NED, PD, NE) based on the BICR and separately for the investigator's assessment. No formal statistical analyses are planned.

The objective response rate (ORR) based on BICR and the ORR based on the investigator's assessment will be summarised (i.e., number of patients (%)) by treatment group, in patients in the Measurable Disease Analysis Set(MDAS).

ORR (BICR and investigator data) will be analysed using logistic regression adjusted by the stratification factors decided from the pooling strategy. Results of the analysis will be presented in terms of an odds ratio (olaparib vs. chemotherapy) together with its associated 95% CI and 2-sided p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). CIs will be profile likelihood CIs (e.g. using the option 'LRCI' in SAS procedure GENMOD).

Stratification variables will be defined according to data from the interactive voice/web response system (IVRS/IWRS). Although not anticipated, if patients are randomised in error when they have not previously had at least 2 prior lines of chemotherapy, then they will be grouped with patients who have had 2 prior lines of chemotherapy. If there are any patients who were miss-stratified, a sensitivity analysis of the primary ORR analysis will be carried out using the (correct) baseline data collected in the eCRF.

A response of CR or PR will not require confirmation due to the randomised controlled study design as per the RECIST guidelines. However, a sensitivity analysis of confirmed CR or PR will be conducted for both BICR and Investigator data.

4.1.2.1 Subgroup analysis for the primary endpoint (ORR)

Subgroup analyses will be conducted comparing ORR between treatments. The purpose of the subgroup analyses is to assess the consistency of treatment effect across potential or expected prognostic factors. If there are too few events available for a meaningful analysis of a particular subgroup (it is not considered appropriate to present analyses where there are less than 5 events in each stratum; *a stratum is defined as strata1*strata2*...strataX*treatment; so with 2 stratification factors of each 2 levels and two treatments we have 2*2*2=8 stratum),*), the relationship between that subgroup-level and ORR will not be formally analysed. In this

case, only descriptive summaries will be provided (response count and percentage). Options for pooling subgroups may be explored if clinically meaningful.

The subgroup analyses for the stratification factors will be based on the values entered into the IVRS, all other factors will be based on values recoreded on the eCRF as indicated below. If there are cases where patients are mis-stratified, then subgroups by stratification factors will be repeated using the (correct) data collected on the eCRF.

The following subgroups of the measurable disease analysis set will be analysed for ORR

- Stratification factorsNumber of prior chemotherapy regimens received for ovarian cancer (2 or 3 vs. 4 or more)
- Selected study chemotherapy (paclitaxel/topotecan/pegylated liposomal doxorubicin/gemcitabine)
- Time to disease progression on last platinum based chemotherapy received prior to randomisation (6 – 12 months / > 12 months)

Additional subgroups of interest include:

- Age at randomisation (<65, ≥65) - This will be determined from the date of birth (BIRTHDAT in the DEM module) and date of randomisation (RND_DAT in the CRIT1 module) on the eCRF at screening. Where a partial date of birth has been collected, the following imputation rules will be used:
    - If only the month and year of birth has been collected, the day of birth will be imputed as 15th
    - If only the year of birth has been collected, the day and month of birth will be imputed as 1st July
    - If the date of birth is completely missing, the age of the patient collected on the CRF will be used.
- Region: to account for regional differences in clinical practice
    - Region 1 – North America (US, Canada, Mexico) vs RoW (Argentina, Belgium, Brazil, Czech Rep, Hungary, Israel, Italy, Poland, Spain, South Korea)
    - Region 2 - N America and Western Europe/Australia (US, Canada, Belgium, Italy, Spain, Mexico) vs Asia/non Western Europe (Argentina, Brazil, Poland, Hungary, Czech Rep, Israel, South Korea)

- Race (White, Black/African-American, Asian or Native Hawaiian/Pacific Islander) - This will be determined from the response to "Race" (DEM module) on the eCRF at screening.

- ECOG performance status at baseline (ECOG PS 0 [PSTAT=0] versus ECOG PS ≥ 1 [PSTAT≥1]) - This will be determined from the response to "Performance status" (PSTAT module) on the eCRF at screening.

- Prior use of bevacizumab (Yes or No) - Patients with no prior bevacizumab reported will be included in the "No" category. Prior bevacizumab use will be identified using the previous ovarian cancer therapy eCRF page

Other baseline variables may also be assessed if there is clinical justification.

For each subgroup, the ORs (olaparib: physician's choice of chemotherapy) and associated CIs will be calculated from an unadjusted logistic regression model. The ORs and 95% CIs will be presented on a forest plot including the OR and 95% CI from the overall population (using the unadjusted logistic regression model). If necessary, additional analyses will be performed to explore the impact of confounding factors. No adjustment to the significance level for testing will be made since all these subgroup analyses will be considered exploratory and may only be supportive of the primary analysis of ORR. If necessary additonal analyses will be performed to explore the impact of confounding factors.

The primary ORR analysis will be repeated excluding any patients who did not have a gBRCA mutation status confirmed by the Myriad test, using an adjusted logistic regression model with the same ties and stratification factors as the primary ORR analysis. If there are < 5 responses per stratum, the logistic regression will be unadjusted.

4.1.2.2 Sensitivity analyses for the primary endpoint (ORR)

(a) Ascertainment bias

An adjusted logistic regression will be repeated using the programmatically derived RECIST using investigator assessed ORR, using the same ties and stratification factors as the primary analysis, provided there are enough responses for a meaningful analysis. If there are not ≥ 5 responses per stratum, an unadjusted logistic regression will be performed. The OR and 95% Confidence Interval will be presented.

If there is an important discrepancy between the primary analysis using BICR assessments and this sensitivity analysis using investigator assessments, then the proportion of patients with site but no central confirmation of objective response will be summarised.

Disagreements between investigator and central reviews of RECIST objective response will be presented for each treatment group. The summary will include the early discrepancy rate which is the frequency of investigator review declared responses before the BICR review (≥2 weeks earlier and including responses declared by investigator but not BICR) as a proportion of all investigator review responses and the late discrepancy rate which is the frequency of

investigator review declared responses after the BICR review (≥2 weeks later and including responses declared by BICR but not investigator) as a proportion of all discrepancies.

(b) Subjects with evaluable disease at baseline

As a sensitivity analysis to the primary ORR analysis, the analysis will be repeated including all patients with evaluable disease at baseline using BICR data. This will include all patients in the Measurable Disease Analysis Set (MDAS) and patients with non-measurable disease at baseline, but presenting with non-target lesions at baseline. The adjusted logistic regression model will use the same ties and stratification factors as the primary analysis, provided there are enough responses for a meaningful analysis. If there are not ≥ 5 responses per stratum, an unadjusted logistic regression will be performed.

(c) Full Analysis Set

As a sensitivity analysis to the primary ORR analysis, the analysis will be repeated including all randomised patients using BICR data. (All patients in the FAS). The adjusted logistic regression model will use the same ties and stratification factors as the primary analysis, provided there are enough responses for a meaningful analysis. If there are not ≥ 5 responses per stratum, an unadjusted logistic regression will be performed.

(d) Confirmed response

A sensitivity analysis of confirmed CR or PR will be conducted for both BICR and investigator assessments. A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit. Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue 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 (i.e. both visits contributing to a response must be prior to subsequent therapy for the patient to be considered as a responder).

In the case where a patient has two non-consecutive visit responses of PR, then, as long as the time between the 2 visits of PR is greater than 4 weeks and there is no PD between the PR visits, the patient will be defined as a responder. Similarly, if a patient has visit responses of CR, NE, CR, then, as long as the time between the 2 visits of CR is greater than 4 weeks, then a best response of CR will be assigned.

The adjusted logistic regression model will use the same ties and stratification factors as the primary analysis, provided there are enough responses for a meaningful analysis. If there are not ≥ 5 responses per stratum, an unadjusted logistic regression will be performed.

(e) On-treatment analysis

As a sensitivity analysis of the primary endpoint of ORR, the primary analysis will be repeated in all patients in the Measurable Disease Analysis Set who received at least one dose of randomised treatment.

An adjustedlogistic regression will be repeated using the BICR RECIST data, using the same ties and stratification factors as described for the primary analysis of ORR, provided there are enough responses for a meaningful analysis. If there are not ≥ 5 responses per stratum, an unadjusted logistic regression will be performed. The OR and 95% CI will be presented.

(f) Deviation bias (if meaningful to do)

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

- Were randomised but did not receive olaparib or chemotherapy.
- Deviated from key entry criteria:
  - Female patients with histologically diagnosed relapsed high grade serous ovarian cancer (including primary peritoneal and/or fallopian tube cancer) or high grade endometrioid cancer (Inclusion criterion 3)
  - Documented mutation in BRCA1 or BRCA2 that is predicted to be deleterious or suspected deleterious (known or predicted to be detrimental/lead to loss of function) (Inclusion criterion 4)

An adjusted logistic regression will be repeated using the BICR RECIST data, using the same ties and stratification factors as described for the primary analysis of ORR. The adjusted logistic regression model will use the same ties and stratification factors as the primary analysis, provided there are enough responses for a meaningful analysis. If there are not ≥ 5 responses per stratum, an unadjusted logistic regression will be performed. The OR and 95% CI will be presented.

(g) Unadjusted logistic regression using BICR assessment in the MDAS

As a sensitivity analysis of the primary endpoint of ORR, the primary analysis will be repeated with an unadjusted logisitic regression, using the BICR data for, all patients in the Measurable Disease Analysis Set. The OR and 95% CI will be presented.

4.1.3 Progression free survival (PFS)

PFS will be analysed using a log-rank test stratified in accordance with the pre-defined pooling strategy.The HR and its confidence interval will be estimated from the U and V

statistics obtained directly from the LIFETEST model with inclusion of STRATA terms for the stratification variables, if applicable (and using the Breslow approach for handling ties).

The HR and its confidence interval will be estimated from the log-rank as follows (Berry et al 1991 and Sellke et al 1983)

HR = exp(U/V)

95% CI for HR = (exp{U/V – 1.96/ √V}, exp{U/V + 1.96/ √V})

Where U=Σ(d_{1i}-e_{1i}) is the log-rank test statistic (with d_{1i} and e_{1i} the observed and expected events in group 1) and √V the standard deviation of the log-rank test statistic as produced in the LIFETEST output.

Stratification variables will be defined according to data from the interactive voice/web response system (IVRS/IWRS). Although not anticipated, if patients are randomised in error when they have not previously had at least 2 prior lines of chemotherapy, then they will be grouped with patients who have had 2 prior lines of chemotherapy. If there are any patients who were miss-stratified, a sensitivity analysis of the primary PFS analysis will be carried out using the (correct) baseline data collected in the eCRF.

The HR (olaparib vs chemotherapy) together with its corresponding 95% CI and p-value will be presented (a HR less than 1 will favour olaparib).

A Kaplan-Meier (KM) plot of PFS will be presented by treatment group with tick marks to identify censored observations. 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.

The assumption of proportionality will be assessed. Note that in the presence of non-proportionality, the HR will be interpreted as an average HR over the observed extent of follow-up. Proportionality will be tested firstly by producing plots of complementary log-log (event times) versus log (time) and, if these raise concerns, a time dependent covariate would be fitted to assess the extent to which this represents random variation.

The PFS analysis will be based on BICR assessments, and using all scans regardless of whether they were scheduled or not.

The proportion of patients alive and progression free at 6, 12, 18 and 24 months will be summarised (using the KM curve) and presented by treatment group.

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 did not progress and the latest scan prior to DCO was more than one scheduled tumour assessment interval (+ 2 weeks) prior to the DCO date.

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.

As patients will be randomised, imbalances in demographic factors between the treatment groups are not anticipated. However if any imbalances should occur, the HR and associated confidence interval calculated from a Cox Proportional Hazards model containing treatment, stratification variables and these additional demographic variables, may be reported.

Subgroup analyses will be conducted comparing PFS between treatments. The purpose of the subgroup analyses is to assess the consistency of treatment effect across potential or expected prognostic factors. If there are too few events available for a meaningful analysis of a particular subgroup (it is not considered appropriate to present analyses where there are less than 20 events per subgroup level), the relationship between that subgroup and PFS will not be formally analysed. In this case, only descriptive summaries will be provided (progression count and percentage). Options for pooling subgroups may be explored if clinically meaningful.

The subgroup analyses for the stratification factors will be based on the values entered into the IVRS, all other factors will be based on values recorded on the eCRF as indicated below. If there are cases where patients are mis-stratified, then subgroups by stratification factors will be repeated using the (correct) data collected on the eCRF.

The following subgroups of the full analysis set will be analysed for PFS:

Stratification factors

- Number of prior chemotherapy regimens received for ovarian cancer (2 or 3 vs. 4 or more)
- Selected study chemotherapy (paclitaxel/topotecan/pegylated liposomal doxorubicin/gemcitabine)
- Time to disease progression on last platinum based chemotherapy received prior to randomisation (6 – 12 months / > 12 months)

Additional subgroups of interest include:

- Measurable versus non measurable disease based on the blinded independent central review – This will be determined from the RECIST transfer for the blinded independent central review. If the patient was not adjudicated, then the patient has no measurable disease if there are no target lesions for the primary reviewer. If the patient was adjudicated, then the patient has no measurable disease if there are no target lesions for the reviewer that the adjudicator selected.

- Measurable versus non measurable disease based on the investigator assessment – This will be determined from the RECIST CRF. A patient has no measurable disease if there are no target lesions (the response to the question "Any target lesions present?" is "No") in the investigator site assessment at screening

- Age at randomisation (<65, ≥65) - This will be determined from the date of birth (BIRTHDAT in the DEM module) and date of randomisation (RND_DAT in the CRIT1 module) on the eCRF at screening. Where a partial date of birth has been collected, the following imputation rules will be used:

    - If only the month and year of birth has been collected, the day of birth will be imputed as 15th

    - If only the year of birth has been collected, the day and month of birth will be imputed as 1st July

    - If the date of birth is completely missing, the age of the patient collected on the CRF will be used.

- Region: to account for regional differences in clinical practice

    - Region 1 – North America (US, Canada, Mexico) vs RoW
    - (Argentina, Belgium, Brazil, Czech Rep, Hungary, Israel, Italy, Poland, Spain, South Korea)

    - Region 2 - N America and Western Europe/Australia (US, Canada, Belgium, Italy, Spain, Mexico) vs Asia/non Western Europe (Argentina, Brazil, Poland, Hungary, Czech Rep, Israel, South Korea)

- Race (White, Black/African-American, Asian or Native Hawaiian/Pacific Islander) - This will be determined from the response to "Race" (DEM module) on the eCRF at screening.

- ECOG performance status at baseline (ECOG PS 0 [PSTAT=0] versus ECOG PS ≥ 1 [PSTAT≥1]) - This will be determined from the response to "Performance status" (PSTAT module) on the eCRF at screening.

- Prior use of bevacizumab (Yes or No) - Patients with no prior bevacizumab reported will be included in the "No" category. Prior bevacizumab use will be identified using the previous ovarian cancer therapy eCRF page

Other baseline variables may also be assessed if there is clinical justification.

For each subgroup, the HRs (olaparib: chemotherapy) and associated CIs will be calculated from a Cox proportional hazards model (ties = Efron) that contains the treatment term, factor

and treatment-by-factor interaction term. The treatment effect HRs for each treatment comparison along with their CIs will be obtained for each level of the subgroup from this single model. The HRs and 95% CIs will be presented on a forest plot including the HR and 95% CI from the overall population (using the unadjusted Cox model). If necessary, additional analyses will be performed to explore the impact of confounding factors.

No adjustment to the significance level for testing will be made since all these subgroup analyses will be considered exploratory and may only be supportive of the analysis of PFS.

The presence of quantitative interactions will be assessed by means of an overall global interaction test. This will be performed in the overall population 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 a covariate does not have more than 5 events per stratum (i.e. within each strata of the treatment*covariate interaction), then the covariate-by-treatment interaction term will be omitted. Moreover, if the covariate does not have more than 5 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.

The PFS analysis will be repeated excluding any patients who did not have a gBRCA mutation status confirmed by the central Myriad test. A KM plot of PFS in this subset of patients will be presented by treatment group. Summaries of the number and percentage of patients experiencing a PFS event and the type of event (RECIST or death) will be provided along with median PFS for each treatment arm. The HR and associated 95% CI will be reported using the same methodology as per the primary analysis of PFS.

PFS sensitivity analyses

Sensitivity analyses will be performed to assess the possible presence of time-assessment bias (i.e., differential assessment times between treatment groups).

Summary statistics for the number of weeks between the time of progression and the last evaluable 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 and the number of patients who miss one RECIST assessment will be presented for each treatment group.

(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 evaluable 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). To support this analysis, the mean of subject-level average inter-assessment times will be tabulated for each treatment. This approach will use the BICR 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, non-evaluable tumour assessments will be included. In addition, subjects who take subsequent therapy prior to progression or death will be censored at their last evaluable 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 will be presented.

(c) Ascertainment bias

A stratified log-rank test will be repeated using investigator assessed RECIST data to programmatically derive PFS. The HR and 95% CI will be presented. A Kaplan-Meier plot will also be presented for PFS based on the investigator assessed data.

If there is an important discrepancy between the primary PFS analysis using BICR assessments and this sensitivity analysis using Investigator assessments, then the proportion of subjects with site but no central confirmation of progression will be summarised. The approach of imputing an event at the next visit in the central review analysis may help inform the most likely HR value, but only if an important discrepancy exists.

Disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group. The summary will include the early discrepancy rate which is the frequency of investigator review declared progressions before the BICR review ($\geq$ 2 weeks earlier and including responses declared by investigator and not BICR) as a

proportion of all investigator review progressions and the late discrepancy rate which is the frequency of investigator review declared progressions after the BICR review (≥ 2 weeks later and including responses declared by BICR and not investigator) as a proportion of all discrepancies.

(d) On-treatment analysis

As a sensitivity analysis of the PFS endpoint, the analysis will be repeated in all patients in the full analysis set who recevied at least one dose of randomised treatment.

A stratified log-rank test will be repeated using the 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.

(e) Deviation bias (if meaningful to do)

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

- Were randomised but did not receive olaparib or chemotherapy.
- Deviated from key entry criteria:
  - Female patients with histologically diagnosed relapsed high grade serous ovarian cancer (including primary peritoneal and/or fallopian tube cancer) or high grade endometrioid cancer (Inclusion criterion 3)
  - Documented mutation in BRCA1 or BRCA2 that is predicted to be deleterious or suspected deleterious (known or predicted to be detrimental/lead to loss of function) (Inclusion criterion 4)

A stratified log-rank test will be repeated using the 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.

(f) Cox proportional hazards model

An additional sensitivity analysis of PFS will be performed based on a Cox proportional hazards model. The hazard ratio (HR) and confidence interval (CI) will be estimated from a Cox Proportional Hazards model (with ties=Efron and the same stratification variables as the primary PFS analysis as covariates) and the CI will be calculated using a profile likelihood approach.

(g) Unadjusted Cox proportional hazards model

An additional sensitivity analysis of PFS will be performed based on an unadjusted Cox proportional hazards model. The hazard ratio (HR) and confidence interval (CI) will be

estimated from a Cox Proportional Hazards model (with ties=Efron) and the CI will be calculated using a profile likelihood approach.

4.1.4 Progression free survival 2 (PFS2)

An initial PFS2 analysis will be performed at the same time as the primary analysis (subject to the hierarchy defined in 4.1.1) and will use the same methodology and model as the PFS analysis, stratified in accordance with the pre-defined pooling strategy (defined in 4.1). . If there are less than 20 events in total at the time of the primary analysis, descriptive summaries will be provided only. A further analysis of PFS2 will be performed when the OS data are approximately 60% mature only if both ORR and PFS are statistically significant based on the primary analysis and the null hypotheses for PFS2 and/or OS are not rejected at the time of the primary analysis.

The type of progression (objective progression by RECIST, progression by CA-125, symptomatic progression or other) will also be summarised by treatment arm.

The analysis of PFS2 will be repeated excluding any patients who did not have a *gBRCA* mutation status confirmed by the Myriad test. The HR and associated 95% CI from a log-rank will be reported as per the primary analysis of PFS2.

The sensitivity analysis outlined for PFS in Section 4.1.3 will not be repeated for PFS2. Descriptive summaries of time from second progression to previous assessment by treatment arm will be provided. . For subjects, who have no previous assessment for second progression prior to the second progression event, the date of the first progression will be used.

4.1.5 Overall survival (OS)

OS data will be analysed at the time of the primary analysis and will use the same methodology and model as the PFS analysis, stratified in accordance with the pre-defined pooling strategy (defined in 4.1). If there are less than 20 deaths in total at the time of the primary analysis, descriptive summaries will be provided only. A further analysis of OS will be performed when the OS data are approximately 60% mature (approximately 150 deaths) only if both ORR and PFS are statistically significant based on the primary analysis and the null hypotheses for PFS2 and/or OS are not rejected at the time of the primary analysis. Subject to the hierarchy defined in 4.1.1, to control for multiple testing due to an interim and final analysis a Lan DeMets spending function (Lan and DeMets 1983) that approximates an O'Brien Fleming approach will be used.

The analysis of OS will be repeated excluding any patients who did not have a *gBRCA* mutation status confirmed by the Myriad test. The HR and associated 95% CI from a log-rank will be reported as per the primary analysis of OS.

The sensitivity analysis outlined for PFS in Section 4.1.3 will not be repeated for OS.

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:

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

Exploratory analyses of OS

Exploratory analyses of OS adjusting for impact of subsequent PARP inhibitor trial or treatment may be performed if a sufficient proportion of patients switch. Methods such as Rank Preserving Structural Failure Time (RPSFT) (Robins et al 1991), Inverse Probability of Censoring Weighting (IPCW) (Robins 1993) and other methods in development will be explored. The decision to adjust and final choice of methods will be based on a blinded review of the data and the plausibility of the underlying assumptions. Baseline and time-dependent characteristics will be explored, and summaries of baseline characteristics will be summarised for chemotherapy patients, splitting between those that have and haven't switched at the time of the analyses.

4.1.6 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 at the same time as the primary analysis and will use the same methodology and model as the PFS endpoint, stratified in accordance with the pre-defined pooling strategy (defined in 4.1).. The HRs for the treatment effect together with 95% CIs will be presented. KM plots will be presented by treatment arm. In addition, the 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.

Further analyses of these endpoints may be performed when the OS data are approximately 60% mature only if both ORR and PFS are statistically significant based on the primary analysis and the null hypotheses for PFS2 and/or OS are not rejected at the time of the primary analysis.

The analysis of time to earliest progression by RECIST 1.1, CA-125 or death will be repeated excluding any patients who did not have a *gBRCA* mutation status confirmed by the Myriad test. The HR and associated 95% CI from a log-rank will be reported as per the primary analysis of TFST.

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

4.1.7 Time to study treatment discontinuation or death (TDT)

Time to study treatment discontinuation or death (TDT) will be analysed at the same time as the primary analysis and using the same methodology and model as for the PFS endpoint, stratified in accordance with the pre-defined pooling strategy (defined in 4.1). 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.

Further analysis of this endpoint may be performed when the OS data are approximately 60% mature only if both ORR and PFS are statistically significant based on the primary analysis and the null hypotheses for PFS2 and/or OS are not rejected at the time of the primary analysis.

The analysis of TDT will be repeated excluding any patients who did not have a *gBRCA* mutation status confirmed by the Myriad test. The HR and associated 95% CI from a log-rank will be reported as per the primary analysis of TDT.

4.1.8 Time to earliest progression by RECIST 1.1, CA-125 or death

Time to progression by BICR RECIST 1.1, CA-125 or death will be performed at the same time as the primary analysis and will use the same methodology and model as the PFS endpoint, stratified in accordance with the pre-defined pooling strategy (defined in 4.1)..

The number (%) of patients reporting a CA-125 progression and a combined objective progression and/or CA-125 progression will be tabulated.

No multiplicity adjustment will be applied as this is viewed as a supportive endpoint (to PFS).

The analysis of time to earliest progression by RECIST 1.1, CA-125 or death will be repeated excluding any patients who did not have a *gBRCA* mutation status confirmed by the Myriad test. The HR and associated 95% CI from a log-rank will be reported as per the primary analysis of Time to earliest progression by RECIST 1.1, CA-125 or death.

4.1.9 Duration of Response (DoR) and CA-125 Response

Any patients who experienced CR or PR which was first observed whilst receiving subsequent therapy after discontinuation of olaparib will be identified. In addition, the number and percentage of patients with a RECIST response based on investigator and/or a CA-125 response will be summarised by treatment arm.

Median duration of response (DoR) derived using Kaplan-Meier methodology and median time to response will be summarised by treatment arm. A Kaplan-Meier plot of DoR will be presented by treatment arm.

4.1.10 FACT-O

The analysis population for HRQoL data will be the subset of the FAS.

Change from baseline in TOI score will be regarded as the primary analysis of the FACT-O questionnaire and will be analysed using a mixed model for repeated measures (MMRM) analysis of the change from baseline in TOI score for each visit. The primary analysis will be to compare the average treatment effect from the point of randomisation for the first 8 months (which will include visit data obtained at baseline, weeks 4, 8, 16, 24, 32, 40 and 48). Other timepoints and the study discontinuation visit and the safety follow-up visit will be excluded from this analysis but may be included on supportive summaries and graphical displays as appropriate. Only visits performed whilst the patient is on treatment will be included in the model using APHASE.

The MMRM model will include patient, treatment, visit and treatment by visit interaction as explanatory variables and the baseline TOI score and baseline TOI score by visit interaction,as a covariate. Treatment, visit and treatment by visit interaction will be fixed effects in the model; patient will be included as a random effect. The treatment by visit interaction will remain in the model regardless of significance.

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. The following provides sample code for implementing the MMRM analysis:

```
proc mixed data=TOI method = reml;

  class TRT VISIT SUBJECT;

  model TOISC = TRT VISIT TRT*VISIT TOIBL TOIBL*VISIT / s ddfm=kr;

  repeated VISIT / type=UN subject=SUBJECT;

  lsmeans TRT*VISIT / slice=VISIT pdiff diff alpha=0.05 cl;
```

where TRT is the randomised treatment, , VISIT is the visit, TOISC is the change from baseline in the TOI score, and TOIBL is the baseline TOI score.

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

The adjusted mean estimates and corresponding 95% confidence intervals will be presented by visit for each treatment group.

Descriptive statistics and graphs will be reported for the TOI by visits as well as change in these scores from baseline. These will also be reported for physical well-being, functional well-being, additional concerns, the individual items of the additional concerns and the total

FACT-O. Summary tables of Trial Outcome Index (TOI) best change rates (improvement, worsening, and no change) will be provided.

TOI improvement rate will be analysed using a logistic regression model (improved vs not improved) and adjusted by the same covariates as used in the primary ORR analyses. The results of the analysis will be presented in terms of an odds ratio together with its associated profile likelihood CIs and p-values (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). However, if there are <5 patients with a response of improved in either of the two treatment groups, or if the model is not able to converge due to small numbers in the fitted covariate categories, then no logistic regression analysis will be performed and instead a Fisher's exact test will be considered and mid p-values used. Descriptive statistics will be shown irrespective of the number of responses in each treatment group.

FACT-O compliance (overall compliance and by visit compliance) will be summarised for each treatment group.

Summary tables of best change rates (improvement, worsening and no change, (Table 6)) and analysis of improvement rates will be performed for TOI and total FACT-O

4.1.11 Exploratory analyses

Cancer Therapy Satisfaction Questionnaire (CTSQ-16)

Data from the CTSQ-16 scores of treatment satisfaction (as measured by the Satisfaction with Therapy scale and the other sub-scales and items of the CTSQ-16) will be summarised by treatment arm. The questionnaire will be completed at Week 24 and at study treatment discontinuation (with no baseline assessment).

EQ-5D-5L

The evaluable population will comprise all patients who receive study treatment included in the safety analysis set and who have a baseline EQ-5D-5L assessment.

Descriptive statistics 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, and will be summarised by actual treatment group. To support future economic evaluations of olaparib, additional appropriate analyses may be undertaken, for example, mean health state utility pre and post treatment, and pre and post progression. These exploratory analyses may be carried out to support health authority appraisals and will consequently not be reported in the CSR. Further detail will be provided in the payer analysis plan.

Hospital Resource Use

Data will be listed on the number of outpatient (ambulatory and daytime) visits, hospitalisations (inpatient overnight admissions) by randomised treatment group for all patients who receive study treatment included in the safety analysis set. The total length of

hospital stay (number of nights) by ward type (e.g. general ward, intensive care/high dependency unit) by randomised group will also be listed.

To support future economic evaluations of olaparib, additional appropriate analyses may be undertaken. These exploratory analyses may be carried out to support health authority appraisals and will consequently not be reported in the CSR. Further detail will be provided in the payer analysis plan.

Subsequent therapy

Subsequent therapies received after discontinuation of olaparib will be summarised and listed by treatment group, together with number of regimens received. Patients who subsequently received a PARP inhibitor or entered a PARP inhibitor trial will be summarised and listed by treatment arm according to line of subsequent therapy, i.e. immediately after olaparib or as a later line, in addition to patients in the chemotherapy arm who subsequently received olaparib.

Patient Global Impression of Change

Patient Global Impression of Change will be summarized descriptively.

4.1.12 Safety

Safety data will be summarised and listed only. No formal statistical analyses will be performed on the safety data. All safety data will be summarised based on their treatment received (if a patient received at least one dose of Olaparib they will be summarised in the olaparib arm – see section 2.1.1). In the unlikely event that any errors in treatment dispensing occurred, these errors will be listed.

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. Any AE occurring before olaparib/chemotherapy treatment (i.e. before Study Day 1) and that did not worsen (have an increase in CTCAE grade) after first dose of olaparib/chemotherapy treatment will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'. Any untoward event occurring subsequent to the 30-day follow-up AE reporting period that the investigator assesses as possibly related to the study treatment will also be included in the AE listings, but not in the summary tables.

The summary tables will include all AEs that had a start date or a worsening (increase in CTCAE grade) after the start of treatment up until the end of the 30 day follow-up period. The 30 day follow-up period will be defined as 30 days following discontinuation of olaparib/chemotherapy 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 each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from on the episode level summaries).

Summary information (the number and percent of patients by treatment) will be tabulated 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 outcome of death
- AEs with outcome of death causally related to study medication
- All SAEs
- All SAEs causally related to study medication
- AEs leading to discontinuation of olaparib/chemotherapy
- AEs leading to discontinuation of olaparib/chemotherapy, causally related to olaparib/chemotherapy
- Other significant AEs
- Other significant AEs causally related to olaparib/chemotherapy

An overall summary of the number and percentage of patients in each category will be presented, as will an overall summary of the number of episodes in each category. In addition, a truncated AE table of most common AEs, showing all events that occur in at least 5% of patients overall will be summarised by preferred term, by decreasing frequency. This cut-off may be modified after review of the data.

Each AE event rate (per 1000 patient years) will also be summarised by preferred term within each system organ class. 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 sum of the duration of therapy plus 30 days for the safety follow up (for patients without the event) and the time to the AE (for patients with the event) in each group multiplied by 1000.

AEs will be assigned CTCAE grades (National Cancer Institute (NCI) CTCAE version 4.0) and summaries of the number and percentage of patients will be provided by maximum

reported CTCAE grade, system organ class, preferred term and actual treatment group. Fluctuations observed in CTCAE grades during study will be listed.

Summaries of the number and percentage of patients with AEs leading to dose change of olaparib/chemotherapy and also dose interruptions of olaparib/chemotherapy will be presented by preferred term and treatment group.

In addition, AEs with outcome of death, SAEs, OAEs, AEs leading to discontinuation of treatment, AEs causally related to olaparib/chemotherapy and other significant AEs will be listed.

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

- Related to disease under investigation,
- AE outcome=death,
- Both related to disease under investigation and with AE outcome=death,
- AE with outcome = death ≥ 30 days after last treatment dose,
- Deaths ≥ 30 days after last treatment dose, unrelated to AE or disease under investigation, and
- Patients with unknown reason for death.

A corresponding listing will also be produced.

Separate summaries will be produced that presents any events that occur prior to dosing or starting more than 30 days after discontinuing therapy.

Summary of long term tolerability

To assess long term tolerability, prevalence plots, life table plots and cumulative incidence plots will be presented for:

- Nausea
- Vomiting
- Any other events considered important after review of the safety data, provided there are ≥ 5% events overall.

A prevalence plot provides information on the extent to which the events may be an ongoing burden to patients. The prevalence at time t after first dose of study treatment is calculated as the number of patients experiencing the event divided by the number of patients receiving

study treatment or in safety follow-up at time t; generally, t is categorised by each day after dosing. The prevalence is plotted over time split by treatment arm. Multiple occurrences of the same event are considered for each patient but a patient is only counted in the numerator whilst they are experiencing one of the occurrences of the event.

For each of these AEs (i.e. Nausea, vomiting, other AEs as defined above), median time to first onset of the AE will be presented in patients with the specific AE of interest in the safety analysis set by actual treatment group. 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.

A life table plot can be used to describe the time to onset of the event and specifically when patients are at most risk of first experiencing the event. The hazard, or in other words, the probability of having an AE in a specified time period (e.g. 0-1 months, 1-3 months, 3-6 months, etc.) given that the patient reaches that time period without having an event is plotted for each time period split by treatment.

A cumulative incidence plot is a plot of the raw cumulative incidence and cumulative incidence function over time with the treatment arms presented on separate plots. The raw cumulative incidence is the actual probability that a patient will have experienced their first occurrence of the event by a given time point. The cumulative incidence function estimates the cumulative incidence if the DCO had not been imposed and all patients had completed safety follow-up (Pintilie M.).

Laboratory assessments

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 actual treatment group. For categorical laboratory assessments, shift from baseline will be summarised using frequency and proportion at each scheduled assessment time by actual 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. For parameters with no CTCAE grading, shift tables from baseline to worst value on-treatment will be provided (i.e. on-treatment is defined as data collected up until the last dose of olaparib/chemotherapy). Shift tables for appropriate urinalysis values by worst grade may be also provided as appropriate.

A scatter plot of alanine aminotransferase (ALT) versus total bilirubin, both expressed as multiples of the upper limit of normal (ULN), will be produced with reference lines at 3×ULN for ALT, and 2×ULN for total bilirubin. The scatter plot will be repeated for aspartate aminotransferase (AST) versus total bilirubin with reference lines at 3×ULN for AST, and 2×ULN for total bilirubin. In each plot, total bilirubin will be in the vertical axis.

Liver biochemistry test results over time for patients with elevated ALT or AST, and elevated total bilirubin (at any time) will be plotted (potential Hy's Law cases). Individual patient data where ALT or AST plus total bilirubin are elevated at any time will be listed also.

All laboratory summaries will be presented by actual treatment group.

Vital signs

Vital signs (SBP, DBP, pulse rate, body temperature and weight) will be summarised at baseline by actual treatment group.

4.1.13 Demographic and baseline characteristics data

The following will be summarised by randomised treatment group for the FAS (full analysis set) and separately for the Measurable disease analysis set (MDAS), unless indicated otherwise:-

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations
- Inclusion in analysis populations
- Demographics (age, age group, sex, race and ethnicity)
    - Age will be derived as age at last birthday in whole years using the date of randomisation and date of birth. Where a partial date of birth has been collected, the following imputation rules will be applied in order to calculate the patient's age for use in listings and summaries tables presenting age and/or age group and subgroup analyses based on age:
        - If only the month and year of birth has been collected, the day of birth will be imputed as 15th
        - If only the year of birth has been collected the day and month of birth will be imputed as 1st July
        - If the date of birth is completely missing, the derived age of the patient will also be missing.
- Patient characteristics at baseline (height, weight, weight group, body mass index (BMI) and body mass index group)
- Stratification factors recorded on the eCRF

- Stratification factors according to the IVRS
- Patient recruitment by country and centre (only for FAS)
- Previous ovarian cancer therapy
- Previous therapy for other cancer
- Disease characteristics at baseline (ECOG performance status, BRCA mutation status at screening [BRCA1, BRCA2 or BRCA1/2], gBRCA category by Myriad testing [gBRCAm, gBRCA wt, gBRCA VUS, Missing], primary tumour location, histology type, tumour grade, FIGO stage, time from diagnosis to randomisation and overall disease classification)
    - In the case of partial or missing dates of diagnosis for the calculation of the time from diagnosis:
        - If only the month and year has been collected, the day will be imputed as 15th
        - If only the year has been collected, the day and month will be imputed as 1st July
        - If the date is completely missing, the derived time from diagnosis will also be missing
- Extent of disease
- Disease related medical history (only for FAS)
- Relevant surgical history (only for FAS)
- Physical examination at baseline
- Time from completion of previous platinum chemotherapy to randomisation
- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer therapy
- Blood and related products, colony stimulating factors and other anti-anaemic preparations will be tabulated, including a summary only including patients with grade >=2 haemoglobin on treatment and the Number of concomitant blood transfusions over time.

AZ drug dictionary (AZDD) will be used for concomitant medication coding.

A listing containing both the eCRF BRCA results and the Myriad BRCA results will be produced.

If there is an additonal OS analysis, patients disposition data will also be summarised and listed at the time of updated OS analysis (approximately 60% maturity for OS).

4.1.14 Treatment exposure

The following summaries related to study treatment will be produced for patients in the safety analysis set who received at least one dose of olaparib:

- Total exposure of olaparib.
- Actual exposure of olaparib.
- Number of days on 300 mg olaparib bd = actual exposure for the dose assigned.
- Number of and reasons for dose reductions, dose interruptions, and dose modifications of olaparib. Dose reductions and dose interruptions will be based on investigator initiated dosing decisions. Dose interruptions/reductions due to "Subject Forgot to Take Dose" will be omitted from these summaries.
- PID and RDI of olaparib (entire intended treatment period).

The following summaries related to study treatment will be produced for patients in the safety analysis set who received chemotherapy by chemotherapy regimen:

- Total number of cycles of chemotherapy received.
- Number of and reasons for dose reductions and dose delays of chemotherapy
- PID and RDI of chemotherapy (entire intended treatment period).

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

All treatment information data will be listed for the safety analysis set by randomised treatment group.

4.1.15 Pharmacokinetic analysis

The olaparib plasma concentration data obtained from the collected samples will be included in a listing for the CSR. This listing will be based on patients in the safety analysis set. The plasma concentration data for Day 1 pre-dose, 1 hour post-dose and Day 29 Pre-dose time will be summarized based on patients in the PK analysis set,

The plasma concentration data will be analysed using nonlinear mixed effects modelling (NONMEM).The objectives and methods for population PK and PK-pharmacodynamic (PKPD) modelling will be described in a separate modelling analysis plan and the results of this analysis will be reported in a separate PK and PKPD modelling report.

4.1.16 Data cut-offs

The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the ORR analysis (initial OS analysis) and at the time of the final OS analysis (if applicable) 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.

5. INTERIM ANALYSES

No formal statistical interim analyses for ORR or PFS are planned for this trial. The data cut-off for the primary analysis will occur in January 2019 or at a minimum of 6 months after LSI, whichever is sooner. . PFS2 and OS will be analysed at the time of the primary analysis. An additional PFS2 and OS analysis will only be conducted with further follow up (~60% OS events) if both ORR and PFS are statistically significant based on the primary analysis and the null hypotheses for PFS2 and/or OS are not rejected at the time of the primary analysis.

This study will use an external Independent Data Monitoring Committee (IDMC) to perform interim reviews of accumulating study safety data. This committee will be composed of therapeutic area experts and a statistician, who are not employed by AZ, and do not have any major conflict of interest. Following the review, the IDMC will recommend whether the study should continue unchanged, be terminated, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca. The report will only include the recommendation and any potential protocol amendments and it will not contain any unblinded information or reference to the confidential considerations of the committee to have led to their recommendation. A separate IDMC charter will be developed which will contain any details of the IDMC members and clearly define the responsibilities of the IDMC.

In addition to the periodic review of safety data by an IDMC, the safety of all AstraZeneca clinical studies is closely monitored on an on-going basis by AstraZeneca representatives in consultation with the Patient Safety Department. Issues identified will be addressed; this could involve, for instance, amendments to the study protocol and letters to investigators.

6. CHANGES OF ANALYSIS FROM PROTOCOL

Section 10.5.3.2 in the protocol states: "The analysis of PFS2 will be repeated excluding any patients who did not have a *gBRCA* mutation status confirmed by the Myriad test. The same

methodology and model will be used and the HR and associated 95% CI from a Cox Proportional Hazards model will be reported."

Section 4.2.4 in the SAP has been amended to "The analysis of PFS2 will be repeated excluding any patients who did not have a *gBRCA* mutation status confirmed by the Myriad test. The HR and associated 95% CI from a log-rank will be reported as per the primary analysis of PFS."

The subset of patients with confirmed gBRCA mutation status will be analysed using the same methodology as the primary analysis for PFS and repeated for the following endpoints: PFS, PFS2, OS, time to earliest progression by RECIST1.1 or CA-125 or death, TFST, TDT and TSST.

Section 10.4.1.2.1 PFS in the CSP had a typo. "Objective progression is defined as at least a 20% increase in the sum of the diameters of the target lesions (compared to previous minimum sum) and an absolute increase of > 5 mm or an overall non-target lesion assessment of progression or a new lesion." This has been amended to " >=5mm " in the SAP section 3.2.2.

7. REFERENCES

Basen-Enquist K et al 2001
Basen-Engquist K, Bodurka-Bevers D, Fitzgerald MA, Webster K, Cella D, Hu S, et al. Reliability and validity of the functional assessment of cancer therapy-ovarian. J Clin Oncol. 2001 Mar 15;19(6):1809-17.

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

Cella D et al 1993
Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol. 1993 Mar;11(3):570-9.

Ellis et al 2008
Ellis S, Carroll K J, Pemberton K. Analysis of duration of response in oncology trials. Contemporary Clinical Trials 2008; 29: 456-465.

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

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

1983; 70; 3:659–63

Osoba et al 2005
Osoba D, Bezjak A, Brundage M, et al: Analysis and interpretation of health-related quality-of life data from clinical trials: Basic approach of The National Cancer Institute of Canada Clinical Trials Group. Eur J Cancer 2005; 41:280-287

Pintilie M.
Competing risks: A practical perspective. Wiley.

Robins et al 1991
Robins JM, Tsiatis AA. Correcting for noncompliance in randomised trials using rank preserving structural failure time models. Communications in Statistics - Theory and Methods 1991; 20:2609-2631

Robins 1993
Robins J. Information recovery and bias adjustment in proportional hazards regression analysis of randomised trials using surrogate markers. Proceedings of the Biopharmaceutical Section, American Statistical Association 1993; 24-33

Rustin et al 2004
Rustin, GJS et al. New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). J Natl Cancer Inst 2004; 96(6):487-488

Sellke et al 1983
Sellke, T. and Siegmund, D. Sequential analysis of the proportional hazards model. Biometrika 1983; 70: 315-326

Stone 2010
Stone, A. The application of bespoke spending functions in group sequential designs and the effect of delayed treatment crossover in survival trials. Pharm. Statistics 5 (2010) 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 Research 2:97-108, 2010

Trask et al 2008
Trask PC, Tellefsen C, Espindle D, Getter C, Hsu M-A, Psychometric validation of the cancer therapy satisfaction questionnaire, Value in Health 2008: 11(4): 669-79.

.