Statistical Analysis Plan	
Study Code	
	D081SC00001
Edition Number	4
Date	14th May 2021

A Randomised, Double-blind, Placebo-controlled, Multicentre Phase III Study of Olaparib Plus Abiraterone Relative to Placebo Plus Abiraterone as First-line Therapy in Men with Metastatic Castration-resistant Prostate Cancer

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse event
AESI	Adverse events of special interest
ALK-P	Alkaline phosphatase
AML	Acute myeloid leukaemia
AQA	Analgesic quantification algorithm
ATC	Anatomical Therapeutic Chemical
ATM	Ataxia-telangiectasia mutated
AUC_{ss}	Area under the plasma concentration-time curve across the dosing interval at steady state
$AUC_{(0-8)}$	Area under the plasma concentration-time curve from time zero to 8 hours post-dose
BICR	Blinded independent central review
bid	Twice daily
BLQ	Below Limit of Quantification
BoR	Best objective response
BP	Blood pressure
BPI-SF	Brief Pain Inventory-Short Form
BRCA1	Breast Cancer 1 gene
BRCA2	Breast Cancer 2 gene
CI	Confidence interval
CL_{ss}/F	Apparent total body clearance of drug from plasma after extravascular administration at steady state
$C_{max,ss}$	Maximum observed plasma (peak) drug concentration at steady state
$C_{min,ss}$	Minimum observed plasma (peak) drug concentration at steady state
CR	Complete response
CRO	Contract Research Organisation
CSP	Clinical Study Protocol
CT	Computed tomography
CTC	Circulating tumour cells
CTCAE	Common Terminology Criteria for Adverse Event
ctDNA	Circulating tumour DNA
cfDNA	Cell free DNA
CV	Coefficient of variation

Abbreviation or special term	Explanation
DAE	Discontinuation of investigational product due to adverse event
DBL	Database lock
DCO	Data cut-off
DCR	Disease control rate
DoR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EFR	Evaluable for response
EQ-5D-5L	EuroQol 5 dimension, 5 level, health state utility index
EWB	Emotional well-being subscale
FACIT	Functional assessment of chronic illness
FACT-G	Functional Assessment of Cancer Therapy - General
FACT-P	Functional Assessment of Cancer Therapy - Prostate Cancer
FAPSI-6	FACT Advanced Prostate Symptom Index 6
FAS	Full Analysis Set
FPR	First patient randomised
FWB	Functional well-being subscale
gCV(%)	Geometric Coefficient of Variation
gmean	Geometric mean
gSD	Geometric standard deviation
H_0	Null hypothesis
HR	Hazard ratio
HRR	Homologous recombination repair
HRQoL	Health-Related Quality of Life
IDMC	Independent data monitoring committee
ITT	Intention-to-treat
LD	Longest diameter
LDH	Lactate dehydrogenase
LLOQ	Lower limit of quantification
KM	Kaplan-Meier
mCRPC	Metastatic castration-resistant prostate cancer
MDS	Myelodysplastic syndrome
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation or special term	Explanation
mHSPC	Metastatic hormone-sensitive prostate cancer
MMRM	Mixed model for repeated measures
$MRAUC_{(0\text{-}8)}$	Metabolite to parent ratio for $AUC_{(0\text{-}8)}$
$MRC_{max,ss}$	Metabolite to parent ratio for $C_{max,ss}$
$MRC_{min,ss}$	Metabolite to parent ratio for $C_{min,ss}$
MRI	Magnetic resonance imaging
NA	Not applicable
NC	Not Calculable
NCA	Non compartmental analysis
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NHA	New hormonal agent
NTL	Non-target lesion
NQ	Not Quantifiable
NR	Not Reported
NRS	Numeric rating scale
NS	No Sample
OAE	Other significant adverse event
OME	Oral morphine equivalents
ORR	Objective response rate
OS	Overall survival
PCS	Prostate cancer subscale
PCWG-3	Prostate Cancer Working Group 3
PD	Progressive disease
PFS2	Time from randomisation to second progression or death
PK	Pharmacokinetics
PR	Partial response
PRO	Patient Reported Outcome
PSA	Prostate-specific antigen
PWB	Physical well-being subscale
qd	Once daily
REML	Restricted maximum likelihood
RDI	Relative dose intensity

Abbreviation or special term	Explanation
RECIST 1.1	Response Evaluation Criteria in Solid tumours version 1.1
rPFS	Radiological progression-free survival
SAE	Serious adverse event
SD	Stable disease
SDev	Arithmetic Standard Deviation
SoA	Schedule of Activities
SOC	System organ class
SSRE	Symptomatic skeletal-related event
SWB	Social/family well-being subscale
TFST	Time to start of first subsequent anticancer therapy or death
TL	Target lesion
t_{last}	Time of last observed (quantifiable) concentration
$t_{max,ss}$	Time to reach peak or maximum observed plasma concentration at steady state
TOI	Trial Outcome Index
TTPP	Time to pain progression

AMENDMENT HISTORY

Date	Brief description of change
10th May 2021	• Added formal testing of key OS secondary efficacy objective at DCO1 and alpha spending for the 3 analyses of OS adjusted to control the overall 1-sided type I error rate at 2.5%. • Definition of disease control rate (DCR) and the analysis timepoint updated. • Analysis of confirmed ORR, BoR, DoR and DCR removed. • New analysis of COVID-related AE's and cardiac/thromboembolic AE grouped terms. • Graphical analysis of long term tolerability of AESI grouped terms removed.
15th January 2021	[redacted] • Section 6 has been updated. • The important protocol deviations have been further defined in Section 2.2. • Handling of missing dates for efficacy analyses has been further defined in Section 3. • In Section 3.3.6, second progression has been further defined by adding clinical symptomatic progression and PSA progression. To aid programming, it's clarified in this section that for the second progression, patients have to receive next-line anticancer therapy. • Further clarification of ORR derivation has been added in Section 3.4.1. • Time to PSA progression has been further defined in Section 3.4.6. • In Section 4.2.3.1, the subgroup analyses described for rPFS will be repeated for OS. • In Section 4.2.5, the analysis of concordance between investigator and BICR assessments for rPFS has been updated according to the current Corporate Standards. • For the summary output of the efficacy endpoints, the number of visits to reported has been clarified in Sections 4.2.2, 4.2.3.1-6, and 4.2.4.
31st March 2020	• Clarified throughout that the primary endpoint rPFS will be programmatically derived using investigator assessments (RECIST 1.1 and PCWG-3). [redacted] • Pain palliation and time to pain severity endpoints have been removed throughout SAP. • Evaluable for response analysis set has been added and the safety analysis set, and PK analysis set have been updated. • Updated Section 2.2 Violations and deviations. • General considerations for safety and efficacy endpoints moved to Section 4.1 Analysis methods and Section 4.2.7.1 General considerations for safety assessments, along with handing partial dates in AEs and concomitant medications.

Date	Brief description of change

	• Section 3.1: footnotes have been added to Table 9 for clarity and sentence added to clarify the first bone scan completed after baseline will be considered the '8-week scan' regardless if taken at week 8 or at an unscheduled assessment.
	• Section 3.2 Primary endpoint – Radiological progression free survival (rPFS)
	  • Updated rPFS censoring rules to censor at the time of the earliest date of their last evaluable RECIST 1.1 assessment (taking the latest target lesion, non-target lesion or new lesion scan date) or bone scan assessment that showed Non-PD.
	  • Footnotes have been added to Table 13 for clarity.
	  • Additional row has been added to Table 13.
	  • Text added describing how RECIST 1.1 and PCWG-3 assessments will be merged for BICR and investigator assessments
	• Section 3.3.2 (Time to pain progression)
	  • The requirement for 2 consecutive subsequent assessments is updated to need at least 2 weeks between the end of the initial visit and the start of the subsequent visit.
	  • Clarity added for an increase in opioid use definition.
	  • Added additional information regarding how the average BPI-SF worst pain [Item 3] and the average AQA score are derived.
	  • Added additional information regarding baseline, assessments on or before the date of first treatment will be considered screening.
	• Clarified patients with no SSRE event will be censored at the last SSRE assessment.
	• Clarified that for the time to second progression endpoint, the progression must occur on next-line anticancer therapy, immediately after study treatment.
	• Moved CTC conversion rate to Section 3.4, clarified which analysis set will be used for CTC conversion rate and added text to show patients data will only be included until the start date of the subsequent anti-cancer therapy for patients who receive a subsequent anti-cancer therapy.
	• Removed time to pain severity endpoint. Change in pain severity will be assessed via a MMRM.
	• Clarified the definition for average pain interference and pain severity scores.
	• Clarified that for PRO endpoints (FACT-P total score and sub scores, BPI-SF and AQA), data will only be included until the start date of the subsequent anti-cancer therapy and added censoring rules for patients who receive subsequent therapy.
	• Added missing data imputation rules for AQA scores and a summary of patients with imputed values.
	• Removed EQ-5D-5L endpoint and analysis.
	• Clarified definition of first and last dose for combination therapy.
	• Section 3.6.1 Exposure: added paragraph on missed or forgotten doses, removed compliance definition and added PID definition.
	• Clarified the definition of a concomitant medication.
	• Updated Table 16.

Date	Brief description of change
	• Clarified that the pooling strategy will be used on all sensitivity analyses and secondary endpoints throughout. • Clarified that values collected on the eCRF will be used to define subgroups for stratification factors and added consistency of treatment effect between subgroups. • Removed cumulative distribution function and logistic regression analysis for time to pain progression. • Moved CTC conversion rate to Section 4.2.4 and removed CTC count changes on a continuous scale. • Added MMRM to pain severity and pain interference Section. • Added text to clarify the logistic regression analyses for FACT-P. • Amended Section 4.2.4.7 Concordance between BICR and investigator assessments for rPFS to display early and late discrepancy rate. • Clarified which information will be summarised for demographics and baseline characteristics. • Added overall AE summaries for grouped AE terms. • Updated changes of analysis from protocol.

1 STUDY DETAILS

1.1 Study objectives

1.1.1 Primary objective

The primary objectives of the study and associated outcome measures are summarised in Table 1.

Table 1 Primary objective

Primary Objective:	Outcome Measures:
To determine the efficacy of the combination of olaparib and abiraterone vs placebo and abiraterone by assessment of rPFS in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage.	rPFS, defined as the time from randomisation to 1) radiological progression, assessed by investigator per RECIST 1.1 (soft tissue) and PCWG-3 criteria (bone), or 2) death from any cause, whichever occurs first.

mCRPC, Metastatic castration-resistant prostate cancer; NHA, New hormonal agent; PCWG-3, Prostate Cancer Working Group 3; rPFS, Radiological progression-free survival; RECIST 1.1, Response Evaluation Criteria in Solid tumours version 1.1.

1.1.2 Secondary objectives

The secondary objectives of the study and associated outcome measures are summarised in Table 2.

Table 2 Secondary objectives

Secondary Objectives:	Outcome Measures:
To determine the efficacy of the combination of olaparib and abiraterone vs placebo and abiraterone by assessment of OS in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage.	OS, defined as the time from randomisation to death from any cause.
To determine the efficacy of the combination of olaparib and abiraterone vs placebo and abiraterone as assessed by time to start of first subsequent anticancer therapy or death (TFST) in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage.	TFST, i.e., the time from randomisation to; 1) the start of the first subsequent anticancer therapy or 2) death from any cause.[a]
To determine the efficacy of the combination of olaparib and abiraterone vs placebo and abiraterone as assessed by time to pain progression (TTPP) in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage.	TTPP is defined as the time from randomisation to pain progression based on the Brief Pain Inventory-Short Form (BPI-SF) Item 3 "worst pain in 24 hours" and opiate analgesic use (analgesic quantification algorithm [AQA] score).[b]

Secondary Objectives:	Outcome Measures:
To further evaluate the efficacy of the combination of olaparib and abiraterone vs placebo and abiraterone by assessment of time to opiate use, time to an SSRE, CTC conversion, and PFS2 in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage.	• Time to opiate use: The time from randomisation to the first opiate use for cancer-related pain. • Time to an SSRE: the time from randomisation to the first SSRE. An SSRE is defined as use of radiation therapy to bone in order to prevent or relieve skeletal complications, occurrence of new symptomatic pathological bone fractures (vertebral or non-vertebral, resulting from minimal or no trauma), occurrence of radiologically confirmed spinal cord compression or a tumour-related orthopaedic surgical intervention. • PFS2: The time from randomisation to second progression on next-line anticancer therapy by investigator assessment of radiological progression, clinical symptomatic progression, PSA progression or death.
To assess the effect of the combination of olaparib and abiraterone vs placebo and abiraterone on disease related symptoms and HRQoL using BPI-SF and Functional Assessment of Cancer Therapy (FACT) - Prostate Cancer (FACT-P) questionnaires in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage.	• BPI-SF: change in pain severity domain, and change in pain interference domain • FACT-P total score, FACT-G total score, trial outcome index, functional well-being, physical well-being, prostate cancer subscale, and FACT Advanced Prostate Symptom Index-6 (FAPSI-6)
To evaluate tumour and blood samples collected from patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage for mutations in *BRCA1, BRCA2, ATM, BRIP1, BARD1, CDK12, CHEK1, CHEK2, FANCL, PALB2, RAD51B, RAD51C, RAD51D, RAD54L.*	HRR gene mutation status.
To determine steady-state exposure to abiraterone and its active metabolite Δ4-abiraterone in the presence and absence of olaparib. To determine steady-state exposure to olaparib when co-administered with abiraterone. [redacted]	• Plasma concentration data at steady state for olaparib, abiraterone, and Δ4-abiraterone in the subset of patients evaluable for PK. • If sufficient data are available, PK parameters at steady state (eg, maximum concentration [$C_{max,ss}$], time to $C_{max,ss}$ [$t_{max,ss}$], minimum concentration [$C_{min,ss}$], and partial area under the concentration-time curve [$AUC_{0-\tau}$]) will be calculated in the PK patient subset. In addition, the area under the curve at steady state (AUC_{ss}) and the apparent clearance (CL_{ss}/F) for olaparib and the metabolite to parent ratios for $C_{max,ss}$, $C_{min,ss}$ and $AUC_{0-\tau}$ for Δ4-abiraterone will be determined. The time of last concentration (t_{last}) will also be determined as a diagnostic parameter.

[a] Subsequent systemic anticancer therapies (excluding radiotherapy) will be reviewed prior to data unblinding to assess which represent clinically important treatments intended to control prostate cancer. TFST is defined as the time from randomisation to the earlier of 1) the first subsequent anticancer therapy start date following study treatment discontinuation or 2) death from any cause. Any patient not known to have died at the time of the analysis and not known to have had a further anticancer therapy will be censored at the last known time to have not received subsequent therapy, i.e., the last follow-up visit where this was confirmed.

[b] Pain progression is defined as follows: 1) for patients who are asymptomatic at baseline, a ≥2-point change from baseline in the average (4-7 days) Brief Pain Inventory-Short Form (BPI-SF) Item 3 score observed at 2 consecutive evaluations (with at least 2 weeks between the end of the initial visit and the start of the subsequent visit) OR initiation of opioid use for pain; 2) for patients who are symptomatic at baseline (average BPI-SF Item 3 score >0 and/or currently taking opioids), a ≥2-point change from baseline in the average BPI-SF Item 3 score observed at 2 consecutive visits and an average worst pain score ≥4, and no decrease in average opioid use (≥1-point decrease in analgesic quantification algorithm [AQA] score from a starting value of 2 or higher) OR any increase in opioid use (1-point change in AQA score or ≥2-point increase if the starting value is 0) at 2 consecutive follow-up visits (with at least 2 weeks between the end of the initial visit and the start of the subsequent visit). Any patient who has more than 2 visits that are not evaluable for pain progression will be censored at the last evaluable assessment.

AQA, Analgesic quantification algorithm; *ATM*, Ataxia-telangiectasia mutated; $AUC_{0\text{-}8}$, Area under the plasma concentration-time curve in 0-8 h; BPI-SF, Brief Pain Inventory-Short Form; *BRCA1*, Breast Cancer 1 gene; *BRCA2*, Breast Cancer 2 gene; $C_{max,ss}$, Maximum plasma concentration at steady state; $C_{min,ss}$, Minimum plasma concentration at steady state; CTC, Circulating tumour cells; FACT-G, Functional Assessment of Cancer Therapy - General; FACT-P, Functional Assessment of Cancer Therapy - Prostate Cancer; FAPSI-6; FACT Advanced Prostate Symptom Index-6; HRR, Homologous recombination repair; HRQoL, Health-Related Quality of Life; mCRPC, Metastatic castration-resistant prostate cancer; NHA, New hormonal agent; PFS2, Time from randomisation to second progression or death; PK, Pharmacokinetics; OS, overall survival; SSRE, Symptomatic skeletal-related event; TFST, Time to start of first subsequent anticancer therapy or death; $t_{max,ss}$, Time to $C_{max,ss}$; TTPP, Time to pain progression.

1.1.3 Safety objective

The safety objective of the study and associated outcome measures are summarised in Table 3.

Table 3 Safety objective

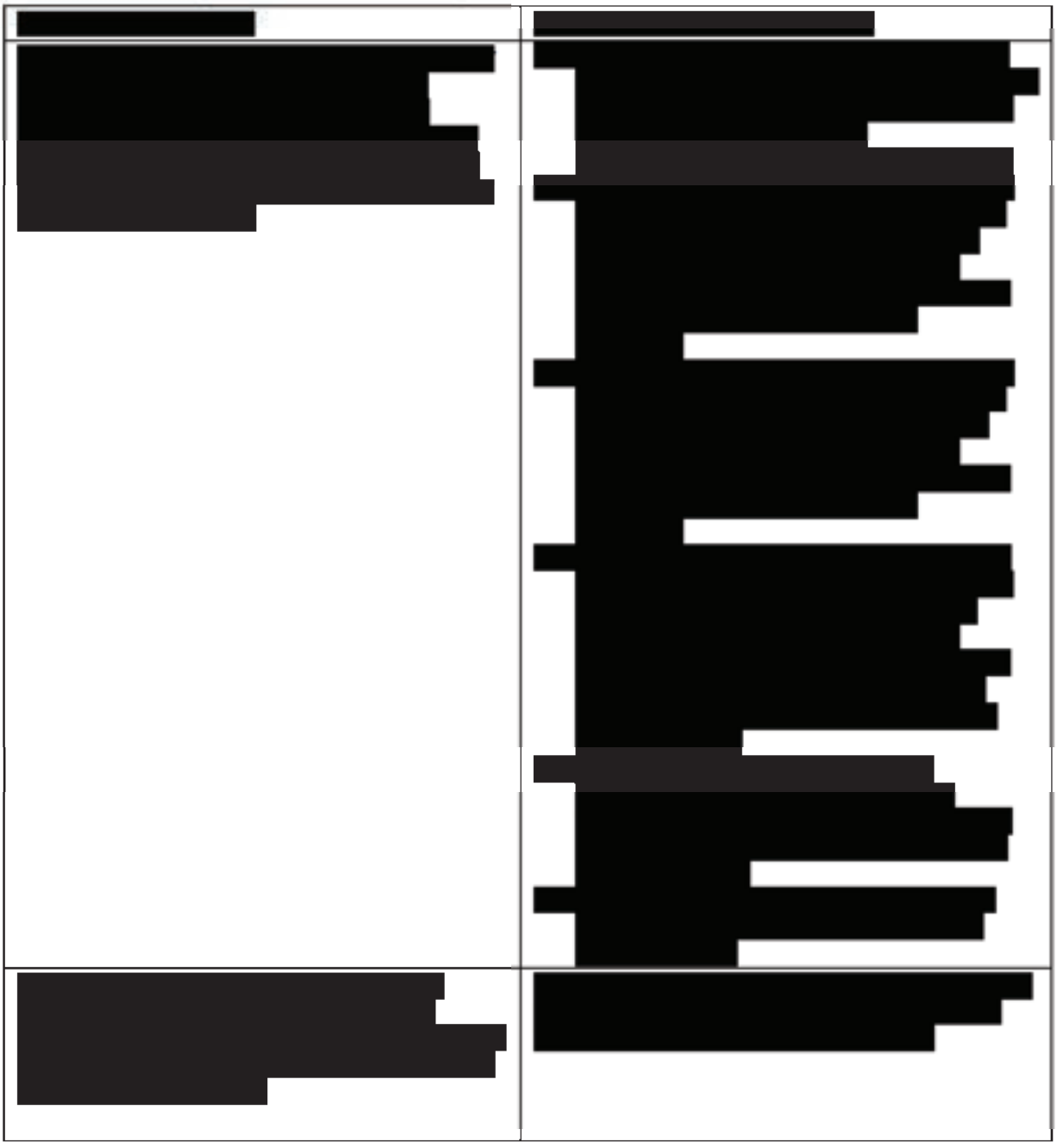
Safety Objective:	Outcome Measures:
To evaluate the safety and tolerability of the combination of olaparib and abiraterone vs placebo and abiraterone in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA at mCRPC stage.	AEs and SAEs, physical examination findings, vital signs (including BP and pulse rate), ECG findings and laboratory test results (including clinical chemistry and haematology parameters).

AE, Adverse event; BP, Blood pressure; ECG, Electrocardiogram, mCRPC, Metastatic castration-resistant prostate cancer; NHA, New hormonal agent; SAE, Serious adverse event.

1.1.4 [redacted]

[redacted]

Table 4 [redacted]



1.2 Study design

This is a randomised, double-blind, placebo-controlled, multicentre Phase III study evaluating the efficacy and safety of the combination of olaparib and abiraterone in patients with metastatic castration-resistant prostate cancer (mCRPC).

Approximately 720 patients were planned to be randomised in a 1:1 ratio to treatment with either olaparib in combination with abiraterone or placebo in combination with abiraterone.

Randomisation occurred within 28 days of screening. Patients will receive oral treatment with olaparib 300 mg twice daily + abiraterone 1000 mg once daily or placebo twice daily + abiraterone 1000 mg once daily. At the time of the protocol amendment, dated January 5, 2021, enrolment had completed with a total of 796 patients randomised.

[redacted]

Abiraterone is indicated in combination with prednisone or prednisolone for the treatment of patients with mCRPC. Hence, patients in both treatment groups will also receive either prednisone or prednisolone 5 mg twice daily with the abiraterone in this study, but throughout this SAP the treatment will be referred to simply as abiraterone.

All patients will attend a screening visit within 28 days before starting study treatment.

Day 1 is defined as the randomisation date; study treatment needs to begin as soon as possible after randomisation, ideally within 24 hours. Randomised patients will continue study treatment until unequivocal radiological progressive disease is assessed by investigator (using Response Evaluation Criteria in Solid tumours version 1.1 [RECIST 1.1] for soft tissue lesions and Prostate Cancer Working Group 3 [PCWG 3] criteria for bone lesions), unacceptable toxicity occurs, or the patient withdraws consent. Following objective disease progression, further treatment options will be at the discretion of the investigator. Patients may be allowed to continue study treatment if the investigator believes, and AZ Study Physician concurs, that the patient could continue to receive clinical benefit, the patient is not experiencing serious toxicity, and there is no available better alternative treatment that could benefit the patient. Crossover from placebo+abiraterone to olaparib+abiraterone is not allowed in this study.

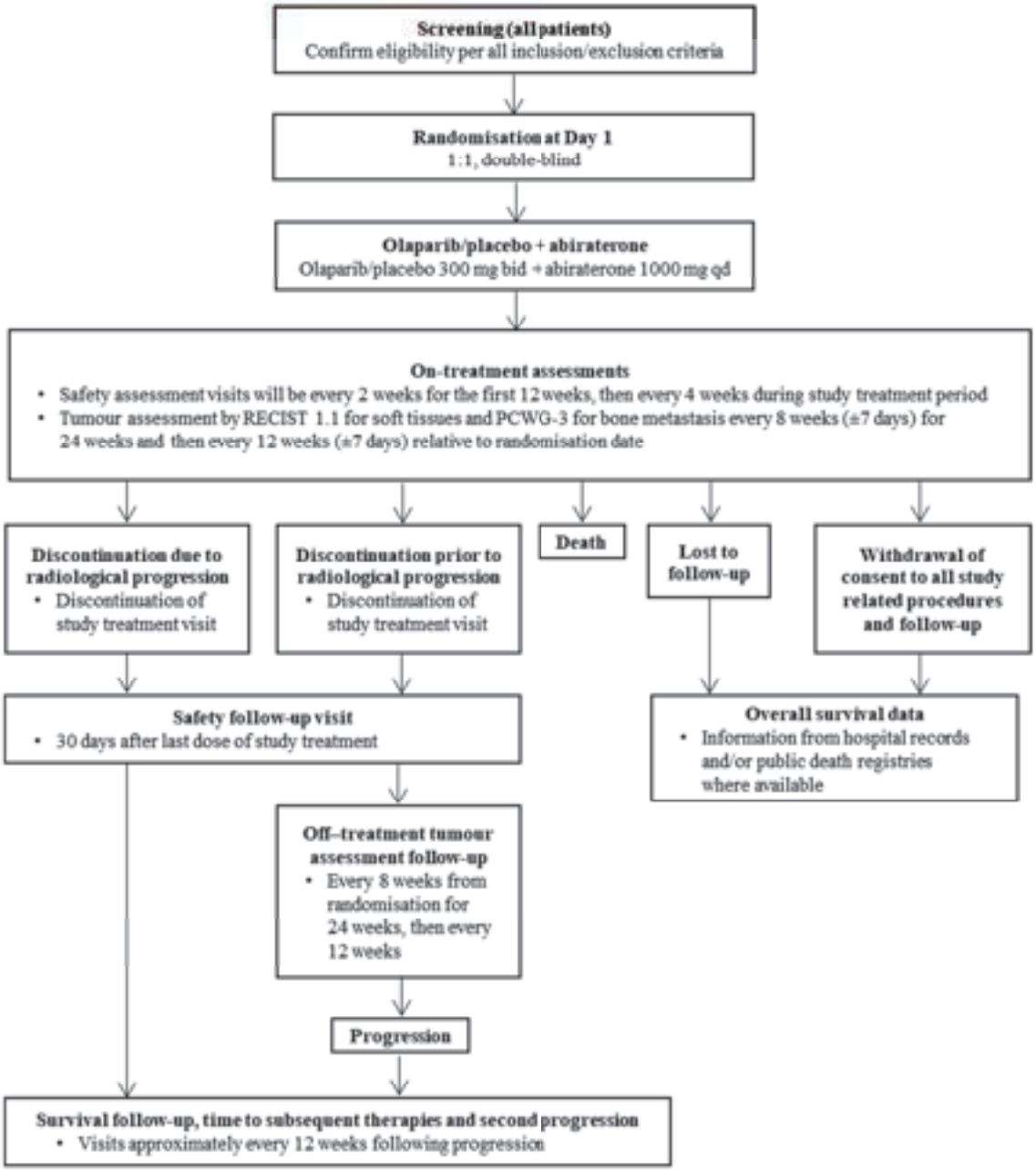
The randomisation scheme will be stratified on the following factors:

- Metastases: bone only vs visceral vs other
- Docetaxel treatment at metastatic hormone-sensitive prostate cancer (mHSPC) stage: yes vs no

The primary endpoint of this double-blind study will be radiological progression-free survival (rPFS) programmatically derived based on investigator recorded assessments using RECIST 1.1 (soft tissue) and PCWG 3 criteria (bone) for all randomised patients. A sensitivity analysis will be conducted using rPFS as assessed for all patients by blinded independent central review (BICR) per RECIST 1.1 (soft tissue) and PCWG 3 criteria (bone).

A study flow chart is illustrated in Figure 1 and the study design is summarised in Figure 1.

Figure 1 Study flow chart



Bid, Twice daily; PCWG-3, Prostate Cancer Working Group 3; qd, Once daily; RECIST, Response Evaluation Criteria in Solid Tumours.

The efficacy of the combination of olaparib plus abiraterone and placebo plus abiraterone will be assessed through the primary and the key secondary endpoints, as described in Section 3.

[redacted]

Primary endpoint
rPFS H_{a0}: HR=1 vs H_{a1}: HR<1

Key secondary endpoint
OS H_{b0}: HR=1 vs H_{b1}: HR<1

[redacted]

[redacted] Patients are, however, permitted to continue to receive study treatment beyond the closure of the database if, in the opinion of the investigator, they are continuing to receive benefit from treatment with study treatment. For patients who do continue to receive treatment beyond the time of this DCO, investigators will continue to report all SAEs to AstraZeneca Patient Safety until 30 days after study treatment is discontinued, in accordance with Clinical Study Protocol (CSP) Section 8.4.1 (Reporting of Serious Adverse Events). If an investigator learns of any SAEs, including death, at any time after a patient has completed the study, and he/she considers there is a reasonable possibility that the event is causally related to the investigational product, the investigator should notify AstraZeneca, Patient Safety. Additionally, as stated in CSP Section 8.3.3 (Follow-up of adverse events [AEs] and SAEs), any SAE or non-serious adverse event that is ongoing at the time of this data cut-off, must be followed up by the investigator for as long as medically indicated.

1.3 Number of patients

Approximately 720 patients were planned to be enrolled into this study. At the time of the protocol amendment, dated January 5, 2021, the enrolment had completed with a total of 796 patients randomised in a 1:1 ratio to treatment with either olaparib and abiraterone or placebo and abiraterone. The study accrual period is about 16 months.

As stated above, the primary endpoint, rPFS, will be formally analysed at DCO1 and DCO2. It is assumed that the true treatment effect was a hazard ratio (HR) of 0.68, corresponding to an assumed increase in median rPFS from 16.5 months (placebo+abiraterone) to 24.3 months (olaparib+abiraterone). The estimated overall dropout rate is 18%. DCO1 is planned to occur

when approximately 379 progression or death events have been accrued in 796 patients (47.6% of patients have an event [maturity], information fraction 83.7%) and would provide approximately 94.1% power to show a statistically significant difference in rPFS. DCO1 is anticipated to occur approximately 31 months after the first patient is randomised in the study.

DCO2 is planned to occur when approximately 453 progression or death events have been accrued (56.9% of patients have an event [maturity], information fraction 100%) and would provide approximately 98.2% power overall to show a statistically significant difference in rPFS. The DCO2 is anticipated to occur approximately 39 months after the first patient is randomised in the study.

[redacted]

Table 5 summarises details of the estimated number of events and power at each DCO with associated, estimated, allowable type I error (alpha) (see Section 4.2.1).

Table 5 Details of planned analyses at each data cut-off

	DCO1	DCO2	[redacted]
Time after first subject randomised	~ 31 months	~ 39 months	[redacted]
rPFS (overall alpha=0.025)			
Power (%)	94.1	98.2	[redacted]
Events	379	453	
Information fraction	83.7	100	
Individual alpha	0.014	0.021	
Critical HR/delta (month)	0.799 / 4.2	0.826 / 3.5	
OS (overall alpha=0.025[a])			
Events	230	295	[redacted]
Information fraction	63.9	81.9	[redacted]
Individual alpha	0.00050	0.01295	[redacted]
Critical HR/delta (month)	0.648 / 19.6	0.772 / 10.7	[redacted]

[a] Overall alpha based upon recycling from primary analysis hypothesis test. Assumes an OS median of 36 months in the control arm.

1-sided alpha is presented. DCO, Data cut-off; HR, Hazard ratio; N/A, Not applicable; OS, Overall survival; rPFS, Radiological progression-free survival.

2 ANALYSIS SETS

2.1 Definition of analysis sets

2.1.1 Full Analysis Set (FAS)

The full analysis set will be used as the primary population for reporting efficacy data and to summarise demographic and patient baseline characteristics. This comprises all patients randomised into the study and will be analysed according to randomised treatment (intention-to-treat [ITT] principle). Any important deviations from randomised treatment will be listed and considered when interpreting the efficacy and safety data.

2.1.2 Evaluable for response (EFR) analysis set

This is a subset of the FAS, who have measurable disease at baseline as per the RECIST 1.1 criteria. Measurable disease will be defined using the investigator assessment for analyses of investigator data, as well as using the BICR assessment data for analyses of BICR assessment.

2.1.3 Safety Analysis Set

The safety analysis set will consist of all randomised patients who received any amount of olaparib, placebo or abiraterone. Safety data will not be formally analysed but summarized using the safety analysis set, according to the treatment received. If a patient receives at least one dose of olaparib study treatment they will be summarized in the olaparib and abiraterone arm for safety summaries. If a patient randomised to olaparib and abiraterone receives placebo and abiraterone (or abiraterone only), they will be summarized in the placebo and abiraterone treatment group for safety summaries.

2.1.4 Pharmacokinetic (PK) Analysis Set

All patients who received at least one dose of randomised study drug and provided at least one post-dose analysable plasma sample for PK analysis will be included in the PK analysis set. Patients with major protocol deviations including changes to the procedures that may impact the quality of the data, or any circumstances that can alter the evaluation of the PK may be excluded from the PK analysis set. These deviations and changes will be specified in a separate protocol deviation specification document. Individual PK concentration or parameter datapoints may be excluded from the PK statistical analyses. The reasons will be agreed with the AZ Clinical Pharmacology Scientist and documented by the PK Scientist.

Table 6 Summary of primary and secondary outcome variables and analysis sets

Outcome variable	Analysis set
Efficacy data	
- Primary efficacy: rPFS	FAS (ITT)
- Secondary efficacy: OS, TFST, TTPP, Time to opiate use, SSRE, PFS2, pain severity, pain interference, HRQoL data, HRR gene mutation status.	FAS (ITT)
[redacted]	[redacted]
- ORR, DCR, DoR and BoR.	EFR
Study Population/Demography Data	
- Demography characteristics (e.g. age, sex etc.)	FAS (ITT)
- Baseline and disease characteristics	FAS (ITT)
- Important deviations	FAS (ITT)
- Medical/surgical history	FAS (ITT)
- Previous anti-cancer therapy	FAS (ITT)
- Concomitant medications/procedures	FAS (ITT)
- Subsequent anti-cancer therapy	FAS (ITT)
PK Data	
- PK data	PK
Safety data	
- Compliance and Exposure	Safety
- Adverse events	Safety
- Laboratory measurements	Safety
- Vital signs	Safety
- ECGs	Safety

BoR, Best objective response; CTC, Circulating tumour cells; DCR, Disease control rate; DoR, Duration of response; ECG, Electrocardiograms; EFR, Evaluable for response; FAS, Full Analysis Set; HRR, Homologous recombination repair; HRQoL, Health-Related Quality of Life; ORR, Objective response rate; OS, Overall survival; PFS2, Time from randomisation to second progression or death; PK, Pharmacokinetics; PSA, Prostate specific antigen; rPFS, Radiological progression free survival; SSRE, Symptomatic skeletal related event; TFST, Time to start of first subsequent anticancer therapy or death; TTPP, Time to pain progression.

2.2 Violations and deviations

The following general categories will be considered important protocol deviations. These will be listed and discussed in the CSR as appropriate:

- Patients who deviate from key inclusion criteria per the CSP (Deviation 1):
  - Histologically or cytologically confirmed prostate adenocarcinoma (inclusion criteria #5).
  - Metastatic status defined as at least 1 documented metastatic lesion on either a bone scan or a CT/MRI scan (inclusion criteria #6)
  - First-line mCRPC (inclusion criteria #7)
- Patients who deviate from key exclusion criteria per the CSP (Deviation 2):
  - Clinically significant cardiovascular disease as evidenced by (exclusion criteria #3)
    - myocardial infarction or arterial thrombotic events (eg, stroke) in the past 6 months,
    - severe or unstable angina, artrial fibrillation or other cardiac arrhythmia requiring therapy, or
    - New York Heart Association Class II-IV heart failure or cardiac ejection fraction measurement of <50% during screening as assessed by echocardiography or multigated acquisition scan
  - Prior revascularisation procedure (significant coronary, carotid or peripheral artery stenosis) (exclusion criteria #5)
  - Prior treatment with PARPi (exclusion criteria #18)
  - Prior treatment with Abiraterone and other CYP17 inhibitors (exclusion criteria #20)
- Patients randomised but (Deviation 3):
  - Did not receive randomised study treatment (olaparib or placebo)
  - Received their randomised study treatment (olaparib or placebo) at an incorrect dose
  - Received an alternative study treatment (olaparib or placebo) to that which they were randomised.
  - Did not receive abiraterone
  - Received other steroid (eg, dexamethasone) than prednisone/prednisolone to support abiraterone treatment
- Received prohibited anti-cancer therapy during study treatment period as per the protocol Table 7 (Deviation 4):
  - Chemotherapy
  - Immunotherapy

  - Radiotherapy (except palliative)
  - Biological therapy
  - Other novel agents
- Met study treatment discontinuation criteria but continued study treatment and potentially had major impact to patients' safety according to clinical judgement (Deviation 5).
- Baseline tumor assessments (Deviation 6):
  - > 42 days before start date of randomised treatment
  - No baseline tumour assessment.
- Persistently missing important protocol required safety assessments (hematology, liver function test, chemistry panel) and potentially having major impact to patient safety (clinical review on a case by case base) (Deviation 7).

A 'deviation bias' sensitivity analysis will be performed on the primary endpoint of rPFS programmatically derived based on investigator recorded assessments excluding patients with deviations that may affect the efficacy (deviations 1, 2, 3, 4 and 6) of the trial therapy if > 10% of patients in either treatment group have IPDs.

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

A listing of all patients affected by a COVID-19 related study disruption by unique subject number identifier and investigational site will be generated along with the description of how the individual's participation was altered.

3 PRIMARY AND SECONDARY VARIABLES

For efficacy analyses, when an event has occurred, every attempt will be made to establish the exact date of the event and enter this into the database. If this is not possible, partial dates will be accepted. If the date of event is not known, then the patient will have an imputed event date as the next day of their last known alive event free date prior to DCO.

For the date variables of historical data (i.e., any data referring to the period prior to the informed consent date), if the year is missing then the value will not be imputed. If the month or day is missing, the value will be imputed: month will be imputed with June; day will be imputed as 15th.

3.1 Derivation of RECIST and PCWG-3 Visit Responses

For all patients, the investigator recorded 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 their best overall response to study treatment. Categorisation of tumour progression of bone lesions will be based on the PCWG-3 criteria. Investigator recorded RECIST 1.1 and PCWG-3 criteria will be used together to assess patient response to treatment.

The baseline assessments of all imaging modalities are to be performed no more than 28 days before the start of randomised treatment and ideally as close as possible to the start of study treatment. Tumour assessments are then performed every 8 weeks (± 7 days) until week 24 and every 12 weeks (± 7 days) thereafter following the start of study treatment until disease progression.

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

From the investigator's review of the imaging scans, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of complete response (CR), partial response (PR), stable disease (SD), no evidence of disease (NED) or progressive disease (PD), using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions and depending on the status of their disease compared with baseline and previous assessments. If a patient has had a tumour assessment that cannot be evaluated then the patient will be assigned a visit response of not evaluable (NE), (unless there is evidence of progression in which case the response will be assigned as PD).

Please refer to Section 3.1.3 for the definitions of CR, PR, SD, NED and PD.

RECIST overall visit responses will be calculated programmatically for the site investigator data (see Section 3.2) and combined with the PCWG-3 bone progression status to assess the efficacy outcomes (i.e. rPFS, ORR etc.).

3.1.1 Target lesions (TLs) – site investigator data

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

maximum of five measurable lesions recorded at baseline as target lesions with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement). 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 NTLs at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

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

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

Table 7 TL Visit Responses (RECIST 1.1)

Visit Responses	Description
Complete response (CR)	Disappearance of all TLs. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.

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

Rounding of TL data

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

Missing TL data

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

- A new lesion is recorded
- A NTL visit response of PD is recorded
- The sum of TLs is sufficiently increased to result in a 20% increase, and an absolute increase of ≥ 5mm, from nadir even assuming the non-recorded TLs have disappeared

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

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

Lymph nodes

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

TL visit responses subsequent to CR

Only CR, PD or NE can follow a CR. If a CR has occurred then the following rules at the subsequent visits must be applied:

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

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure, then this will be indicated as such on the case report form and a value of 5mm will be entered into the database and used in TL calculations. However, a smaller value may be used if the radiologist has not indicated 'too small to measure' on the case report form and has entered a smaller value that can be reliably measured. If a TL response of PD results (at a subsequent visit) then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

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

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

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

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

Scaling (applicable only for irradiated lesions/lesion intervention)

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

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

Example of scaling

Lesion 5 is missing at the follow-up visit; the nadir TL sum including lesions 1-5 was 74 mm.

The sum of lesions 1-4 at the follow-up is 68 mm. The sum of the corresponding lesions at the nadir visit is 62 mm.

Scale up as follows to give an estimated TL sum of 81 mm:

68 x 74 / 62 = 81 mm

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

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of TLs

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

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

3.1.2 Non-target lesions (NTLs) and new lesions – site investigator data.

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

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

Table 8 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).

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

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

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more 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, but should not overtly affect the derivation.

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

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

3.1.3 Overall visit response – site investigator data

Table 9 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 9 Overall visit responses

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT 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[a]	No	**PR**
SD	Non PD[a]	No	**SD**
NA	Non CR/Non PD	No	**SD**
NA	NA	No	**NED**
NE	Non PD[a]	No	**NE**
NA	NE	No	**NE**
PD	Any	Yes or No	**PD**
Any	PD	Yes or No	**PD**
Any	Any	Yes	**PD**

[a] Non PD = CR or Non CR/Non PD or NA or NE.
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 and/or NTLs at baseline), NED = No Evidence of Disease (only relevant when there is no TL and NTL from baseline).

3.1.4 Bone Lesion Progression using PCWG-3

Bone lesions will be assessed by bone scan and will not be part of the RECIST 1.1 malignant soft tissue assessment. If more than one baseline scan is recorded, then measurements from the one that is closest and prior to randomization will be used. Follow-up assessments will be performed every 8 weeks (± 7 days) after randomisation within the first 24 weeks, and then every 12 weeks until objective disease progression assessed by investigator as defined by RECIST 1.1 (soft tissue) or PCWG-3 (bone).

Categorisation of tumour progression of bone lesions will be based on the PCWG 3 criteria. Positive hot spots on the bone scan should be considered significant and unequivocal sites of malignant disease to be recorded as metastatic bone lesions.

Progression on a bone scan is defined as:

At the first visit after baseline:

- If 2 or more new metastatic bone lesions are observed on the bone scan from the first visit after baseline, a confirmatory scan performed at the next schedule visit (and a minimum of 6 weeks later), must show 2 or more additional new metastatic bone lesions (for a total of 4 or more new metastatic bone lesions since the baseline assessment).

  Note - The first bone scan completed after baseline will be considered the '8-week scan' regardless if taken at week 8 or at an unscheduled assessment.

All other visits from the second visit after baseline:

- For patients without progression at the first visit after baseline, the scan from this first visit after baseline now serves as new reference for all subsequent scans, ie, assuming all visits are acquired according to schedule, all bone scans after week 8 are compared to the week 8 scan. If 2 or more new metastatic bone lesions are observed on scans obtained after the first visit after baseline assessment compared to the new reference, a confirmatory scan performed preferably no later than the next scheduled visit and at least 6 weeks later, must show the persistence of, or an increase in, the 2 or more metastatic bone lesions.

The date of radiographic progression is the date of the first scan documenting the 2 new lesions. If the investigator is in doubt as to whether progression has occurred, it is advisable to continue study treatment and reassess the bone lesion status at the next scheduled assessment, or sooner if clinically indicated.

Table 10 provides the definitions for the visit bone progression status for bone lesions.

Table 10 Bone progression status

Bone progression status	
Non Progressive Disease (Non-PD)	No evidence of progression, or appearance of 1 new bone lesion, or non-fulfilment of the progression criteria including new lesions without confirmation of progression
Progressive Disease (PD)	Bone lesions fulfilling the requirements for at least 2 new lesions and confirmation of progression
Not Evaluable (NE)	Only relevant if a follow-up bone scan is not performed

3.1.5 Blinded independent central review (BICR) with RECIST 1.1 and PCWG-3 criteria

A planned BICR of all radiological imaging data will be carried out using RECIST version 1.1 and PCWG-3. All radiological scans for all patients (including those at unscheduled visits, or outside visit windows) will be collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for central analysis. The imaging scans will be reviewed by two independent radiologists using RECIST 1.1 and PCWG-3 and will be adjudicated if required (i.e. two reviewers' review the scans and adjudication is performed by a separate reviewer in case of a disagreement). For each patient, the BICR will define the RECIST 1.1 overall visit response (i.e. the response obtained overall at each visit by assessing TLs, NTLs and new lesions) and a PCWG-3 bone progression status and no programmatic derivation of visit response is necessary. (For patients with TLs at baseline: CR, PR, SD, PD, NE; for patients with NTLs only: CR, SD, PD, NE; for patients with no disease identified at baseline: PD, NED, NE). If a patient has had a tumour assessment that cannot be evaluated then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST and PCWG-3 assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision will be used). If adjudication is performed for the same visit on both the RECIST criteria and PCWG-3 criteria, and the adjudicator agreed on progression on at least one of this two criteria then the visit will be PD. 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/PCWG-3 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.

Results of this independent review will not be communicated to investigators and the management of patients will be based solely upon the results of the RECIST 1.1 and PCWG-3 assessment conducted by the investigator.

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

Further details of the BICR will be documented in the BICR Charter.

3.2 Primary endpoint – Radiological progression free survival (rPFS)

The analysis of the primary endpoint, rPFS, will be programmatically derived based on investigator recorded tumor assessments determined using RECIST 1.1 (soft tissue) and PCWG-3 (bone) criteria. A sensitivity analysis based on BICR assessments will be performed.

rPFS is defined as the time from randomisation until the earlier date of radiological progression or death (by any cause in the absence of progression), regardless of whether the patient withdraws from randomised therapy or receives another anticancer therapy prior to progression (i.e. date of rPFS event or censoring – date of randomization + 1).

Patients who have not progressed (i.e., who have a CR, PR, or SD by RECIST 1.1, non progressive disease by PCWG-3) and not died at the time of analysis will be censored at the earliest date of their last evaluable RECIST assessment (taking the latest target lesion, non-target lesion or new lesion scan date) or bone scan assessment that showed Non-PD (compared to week 8 or compared to baseline at the week 8 visit) (if the RECIST and bone scans are at different visits). If the RECIST and bone scan assessments are performed at the same visit, then the patient will be censored at the latest of the previous RECIST1.1 and bone scan assessments.

However, if the patient progresses or dies immediately after two or more consecutive missed visits, the patient will be censored at the time of the earliest of either the previous RECIST 1.1 assessment (taking the latest target lesion, non-target lesion or new lesion scan date) or previous bone scan assessment prior to the two consecutive missed visits (if RECIST and

bone scan done at different visits). If the RECIST and bone scan assessments are performed at the same visit, then the patient will be censored at the latest of the previous RECIST1.1 and bone scan assessments. If the patient has no evaluable visits or does not have baseline data they will be censored at Day 1 unless they die within 2 visits of baseline (2 visits of baseline equates to Day 120 based on 16 weeks plus 1 week allowing for a late assessment within the visit window), in which case their date of death will be used. Note: NE visit is not considered as missed visit.

Given the scheduled visit assessment scheme (i.e. eight-weekly for the first 24 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change. If the previous assessment (the earliest of the previous RECIST or previous bone scan assessment) is less than study day 106 (i.e. week 15) then two missing visits will equate to 18 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks). If the two missed visits occur over the period when the scheduled frequency of assessments changes from eight-weekly to twelve-weekly this will equate to 22 weeks (i.e. take the average of 8 and 12 weeks which gives 10 weeks and then apply same rationale, hence 2 x 10 weeks + 1 week for an early assessment + 1 week for a late assessment = 22 weeks). The time period for the previous assessment will be from study days 106 to 161 (i.e. week 15 to week 23). From week 23 onwards (when the scheduling changes to twelve-weekly assessments), two missing visits will equate to 26 weeks (i.e. 2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks). See Table 11 for details.

Table 11 Allowable interval from previous radiologic assessment

Date progression first observed	Day of previous radiological assessment *	Allowable interval from previous radiological assessment (2 visits including allowed visit windows)
During 8 weekly assessments	Day 1 to day 105	126 days
During the change from 8 weekly to 12 weekly assessments	Day 106 to day 161	154 days
During 12 weekly assessments	Day 162 onwards	182 days

* Earliest of previous RECIST 1.1 or previous PCWG-3

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

When the Investigator is in doubt as to whether PD has occurred and therefore reassesses the patient at a later date, the date of the initial scan should be declared as the date of progression if the repeat scans confirm progression.

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

- For BICR (RECIST 1.1 and PCWG-3) assessments, 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 the reviewer who read baseline first if there is no adjudication for BICR data.
- For investigator assessments, the date of progression will be determined based on the earliest of the dates of the component that triggered the progression.
- For both BICR and investigator assessments, when censoring a patient for rPFS the patient will be censored at the earliest of the of the previous RECIST 1.1 assessment (taking the latest target lesion, non-target lesion or new lesion scan date) or previous bone scan assessment.

The requirements for determination and confirmation of radiological progression by either bone scan (bone progression) or CT/MRI (soft tissue progression) are summarised in Table 12.

Table 12 Requirements for documentation of progression

Visit Date	Criteria for Bone Progression	Criteria for Soft Tissue Progression
First visit after baseline (expected week 8)	• 2 or more new lesions compared to baseline bone scan. • Requires confirmation at least 6 weeks later with ≥ 2 additional lesions compared to the first scan after baseline	• Progressive disease on CT or MRI by RECIST 1.1 • No confirmation required.
From the 2nd visit after baseline	• 2 or more new lesions compared to the first bone scan after baseline. • Requires confirmation at least 6 weeks later for persistence or increase in number of lesions	• Progressive disease on CT or MRI by RECIST 1.1 • No confirmation required.

CT, computed tomography; MRI, magnetic resonance imaging; RECIST, Response Evaluation Criteria in Solid tumours.

Table 13 provides the definitions how the visit responses for soft tissue (according to RECIST1.1 criteria) and bone progression status (according to PCWG-3 criteria) are combined to give an overall radiological objective visit response.

Table 13 Overall radiological visit response

Overall visit soft tissue response (RECIST 1.1)[a]	Bone progression status (PCWG-3)[b]	Bone lesions at visit Present/Absent	Overall radiological visit response
CR	Non-PD	Absent	CR
CR	Non-PD	Present	PR
CR	NE	Present	PR
PR	Non-PD or NE	Any	PR
SD	Non-PD or NE	Any	SD
NED	Non-PD	Any	Non-PD
NED	NE	Any	NE
NE	Non-PD or NE	Any	NE
PD	Any	Any	PD
Any	PD	Any	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 and NTLs at all visits).
[a] See Section 3.1.3.
[b] See Section 3.1.4.

In order to derive an overall radiological response, the investigator assessments will be merged using windows around the protocolled visit schedule as described in the ADaM specification. The BICR RECIST 1.1 and PCWG-3 assessments will be merged by the BICR visit number.

3.3 Secondary endpoints

3.3.1 Overall survival

Overall survival is defined as the time from date of randomisation to death due to any cause regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy (i.e. date of death or censoring – date of randomisation + 1). Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

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

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

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

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

3.3.2 Time to first subsequent anticancer therapy or death (TFST)

Time to first subsequent anticancer therapy (excluding radiotherapy) is defined as the time from randomisation to the earlier of start date of the first subsequent anti-cancer therapy after discontinuation of randomised treatment or death from any cause (i.e. date of first subsequent cancer therapy/death or censoring – date of randomisation + 1).

When the start date of a subsequent anti-cancer therapy in the Previous Systemic Cancer Therapy module (CAPRX) is partially missing, the missing date will be imputed as the following:

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

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

When the start date of a subsequent anti-cancer therapy in the Post Systemic Cancer Therapy module (CAPRXHC) is partially missing, the missing date will be imputed as the following:

a. For Missing day only – using the maximum (1st of the month, the treatment discontinuation date)

b. For Missing day and Month – using the treatment discontinuation date

Any patient not known to have died at the time of the analysis and not known to have had a further anticancer therapy will be censored at the last known time to have not received subsequent therapy, i.e., the last visit where this was confirmed. If a patient terminated the study for reason other than death before first subsequent therapy, these patients will be censored at the earliest of their last known to be alive and termination dates. Patients not receiving randomised treatment would have TFST calculated as time from date of randomisation to the initial therapy or death.

3.3.3 Time to pain progression (TTPP)

Time to pain progression is defined as time from randomisation to pain progression based on the Brief Pain Inventory-Short Form (BPI-SF) Item 3 "worst pain in 24 hours" and opiate

analgesic use (Analgesic quantification algorithm [AQA] score), i.e. date of pain progression – date of randomisation + 1. See Section 3.5.1 and Section 3.5.2 for details on BPI-SF and AQA score respectively.

Pain progression for asymptomatic patients and symptomatic patients (at baseline) is defined as follows:

For patients who are asymptomatic at baseline (average BPI-SF worst pain [Item 3] score of 0 and not taking opioids):

- A ≥2 point change from baseline in the average (4-7 days) BPI SF worst pain [Item 3] score observed at 2 consecutive visits (with at least 2 weeks between the end of the initial visit and the start of the subsequent visit). The date of pain progression will be the earliest date of the assessments contributing to the average of 7-day assessments for BPI-SF [Item 3].

  or

- Initiation of opioid use for pain;

For patients who are symptomatic at baseline (average BPI SF worst pain [Item 3] score >0 and/or receiving opioids):

- A ≥2 point change from baseline in the average (4-7 days) BPI SF worst pain [Item 3] score observed at 2 consecutive visits (with at least 2 weeks between the end of the initial visit and the start of the subsequent visit) and an average worst pain score ≥4, and no decrease in average opioid use (≥1-point decrease in AQA score from a starting value of 2 or higher). The date of pain progression will be the earliest date of the assessments contributing to the average of 7-day assessments for BPI-SF worst pain [Item 3].

  or

- Increase in opioid use (≥1-point increase, or ≥2-point increase if the starting value is 0) at 2 consecutive follow-up visits (with at least 2 weeks between the end of the initial visit and the start of the subsequent visit). The date of pain severity progression will be the earliest date of the assessments contributing to the average of 7-day assessments.

Information on all analgesics used by patients in pain control will be collected using the analgesic log. For the purposes of pain progression, only information on the actual pain medication collected with the analgesic log will be used. For AQA imputation rules for missing data, see Section 3.5.2.

Any BPI-SF worst pain [Item 3] or analgesic log assessments on or before the date of first treatment will be considered a screening assessment.

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. Study day will be calculated in relation to date of first treatment. For example:

- Day 29, visit window 2 – 42
- Day 57, visit window 43 – 70
- Day 85, visit window 71 – 98

For the Week 4 (Day 29) visit, if there are overlapping screening measurements in the week 4 window on Day 2, 3, 4 etc, resulting in 2 sets of observations in the Week 4 window, then the set of assessments closest to the target day will be used.

Where the average of BPI-SF worst pain [Item 3] score and average AQA score are taken over 7 days for each visit, the 7 day window for both BPI SF worst pain [Item 3] score and AQA score will start from the date of first entry of the BPI-SF worst pain [Item 3] for that visit. For example, if there are medications entered in the analgesic log prior to the first entry of BPI-SF worst pain [Item 3], the data will not be used in the average AQA score. Additionally, if there are medications entered in the analgesic log after the 7 day period, these will not be used in the average AQA score.

To calculate the average BPI-SF worst pain [Item 3] score over 7 days, there must be at least 4 days with the BPI-SF worst pain [Item 3] completed. The denominator for the average BPI SF worst pain [Item 3] over 7 days will be the number of days the BPI-SF worst pain [Item 3] is filled in. The exception to this is baseline where the average will still be calculated if there are less than 4 days filled in, by using the total number of available days.

To calculate the average AQA score, there must be at least 4 out of the 7 days with evaluable data. To count a day as having evaluable data, at least the BPI-SF worst pain [Item 3] or the analgesic log must be filled in. The denominator for the average AQA score will be the number of days either the BPI-SF worst pain [Item 3] or the analgesic log is filled in. The exception to this is baseline where the average will still be calculated if there are less than 4 days with evaluable data, by using the total number of available days.

Pain progression is set to missing at the baseline assessment if BPI-SF worst pain [Item 3] is missing. At subsequent visits, pain progression will be set to missing if there are < 4 days data for BPI-SF worst pain [Item 3] and the average AQA score does not meet the progression

criteria. If average AQA score meets the progression criteria regardless of available BPI-SF worst pain [Item 3] then the visit is set to progression.

For patients who receive a subsequent anti-cancer therapy, data will only be included until the start date of the subsequent anti-cancer therapy. Note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy.

Patients who do not satisfy the pain progression criteria for asymptomatic patients and symptomatic patients (at baseline) will be censored as follows:

- If a patient meets the criteria for pain progression immediately after 2 or more missed visits (visits which showed < 4 days of BPI-SF worst pain [Item 3] assessments and the average AQA score does not meet the progression criteria), then the patient will be censored at the time of the latest evaluable average BPI-SF worst pain [Item 3] assessment (the earliest date of the assessments contributing to the average will be used).
- Patients who have not met the criteria for pain progression at the time of analysis:
  - The censoring date will be the date of the latest evaluable average BPI-SF worst pain [Item 3] assessment (the earliest date of the assessments contributing to the average will be used).
  - Patients with no evaluable baseline or post-baseline data will be censored at Day 1.
- Patients who receive subsequent anti-cancer therapy:
  - The censoring date will be the date of the latest evaluable average BPI-SF worst pain [Item 3] assessment prior to the start date of subsequent anti-cancer therapy (the earliest date of the assessments contributing to the average will be used).
  - Patients with no evaluable baseline or post-baseline data will be censored at Day 1.
- Patients who are randomised but do not receive study treatment will be censored at Day 1.

3.3.4 Time to opiate use

Time to opiate use is defined as the time from date of randomisation to the date of first opiate use for cancer related pain. Patients who have not received opiates during the study or died prior to receiving opiates will be considered censored at the last known on study date prior to DCO of no opiate use. Patients receiving opiates at baseline will not be included in this analysis.

3.3.5 Time to first symptomatic skeletal related event (SSRE)

Time from date of randomisation to date of first symptomatic skeletal–related event as defined by any of the following or a combination:

- Use of radiation therapy to prevent or relieve skeletal symptoms.
- Occurrence of new symptomatic pathological bone fractures (vertebral or non-vertebral). Radiologic documentation is required. A pathological fracture, as determined by investigator, is defined as associated with low or no trauma and deemed to have occurred at a site of bone metastasis.
- Occurrence of spinal cord compression. Radiologic documentation is required.
- Orthopaedic surgical intervention for bone metastasis.

Patients who have not experienced any of the above conditions will be censored at time of last SSRE assessment.

3.3.6 Time to second progression or death (PFS2)

Second progression is based on investigator assessment according to local standard clinical practice and includes radiological, PSA progression and clinical progression. Second progression status is reviewed every 12 weeks following the progression event used for the primary variable PFS (ie, first progression) and the start of the next-line anticancer therapy. Based on two 12-weekly visits plus two allowed 2 week visit windows, a second progression is not evaluable if it was greater than 196 days since the prior assessment.

The time to PFS2 is defined as the time from date of randomisation to date of second progression on next-line (immediately after study treatment) anticancer therapy or death, whichever occurs earlier (i.e. date of PFS2 event or censoring – date of randomisation + 1). If a patient had a first progression (radiological progression) and did not receive any next-line anticancer therapy, the second progression would not be counted as a PFS2 event. This is because the second progression would still be a marker of the effect of the first treatment (i.e. study treatment).

Patients alive and for whom a second progression has not been observed should be censored at the earliest of: date of study termination, date last known alive, DCO or, if a patient has not had a first subsequent therapy, the date last known not to have received a first subsequent therapy (TFST censoring date).

However, if the patient experiences a second progression that is not evaluable, or dies immediately after two or more visits where there was no evaluable PFS2 assessment, the patient will be censored at the time of the later of the first progression date and the latest evaluable second progression assessment.

3.3.7 Pain severity

The BPI-SF (described in Section 3.5.1) pain severity domain consists of 4 items (item #3, item #4, item #5, and item #6) which assess pain at its "worst," "least," "average," and "now"

(current pain) respectively on the 11-point NRS. These 4 items will be averaged (all items must be non-missing) to give a composite pain severity score. The average pain severity subscale score at each visit will be calculated as the average of 7 days starting from the date of the first entry. There must be at least 4 out of the 7 days with a non-missing pain severity subscale score to calculate the average pain severity subscale score.

3.3.8 Pain interference

The BPI-SF (described in Section 3.5.1) pain interference domain includes 7 items: general activity (item #9A), mood (item #9B), walking ability (item #9C), normal work (item #9D), relations with other people (item #9E), sleep (item #9F), and enjoyment of life (item #9G). The pain interference domain is scored as the mean of the 7 interference items. The mean can be used if more than 50% of the total items, or 4 of 7, have been completed on a given administration (Cleeland 2009). The average pain interference subscale score at each visit will be calculated as the average of 7 days starting from the date of the first entry. There must be at least 4 out of the 7 days with a non-missing pain interference subscale score to calculate the average pain interference subscale score.

3.3.9 Functional Assessment of Cancer Therapy- Prostate Cancer (FACT-P)

The following outcome measures will be calculated from the FACT-P questionnaire, the resulting value is the total score for the associated questions or scaled scores:

- Physical well-being subscale (PWB) (Questions GP1 to GP7)
- Social/family well-being subscale (SWB) (Questions GS1 to GS7)
- Emotional well-being subscale (EWB) (Questions GE1 to GE6)
- Functional well-being subscale (FWB) (Questions GF1 to GF7)
- Prostate cancer subscale (PCS) (Questions C2, C6, P1 to P8, BL2 and BL5)
- Total Functional Assessment of Cancer Therapy- General (FACT-G) score, sum of PWB, SWB, EWB and FWB
- Trial Outcome Index (TOI), sum of PWB, FWB and PCS
- Functional Assessment of Prostate Cancer Symptoms Index 6 (FAPSI-6) (Questions P1 to P3, GP1, C2 and GE6)
- Total FACT-P score (sum of scores of all the sub-scales)

Items to be reversed

Each question in the FACT-P questionnaires has a choice of 5 responses, "Not at all", "A little bit", "Somewhat", "Quite a bit" and "Very much". The scores range from 0 ("Not at all") to 4 ("Very much") for positively phrased questions. Negatively phrased questions have a reverse

scoring, from 0 ("Very much") to 4 ("Not at all"). This results in a consistent approach, where higher scores indicate a better quality of life.

Note, questions that are reversed (via subtraction of the response from 4) are: GP1-7, GE1, GE3-6, C2, P1-3, P6-P8 and BL2.

Missing data

As per the functional assessment of chronic illness (FACIT) scoring guidelines (Amit et al 2011),

- More than 80% of questions in a questionnaire must be completed for the questionnaire to have the FACT-P total score evaluable. If 80% or less of questions are completed, the FACT-P total scores will not be calculated. Similarly, FACT-G total score and TOI score require more than 80% of the relevant questions to be completed for the score to be evaluable.
- For each domain (PWB, SWB, EWB, FWB and PCS) if more than 50% of the items were answered (e.g., a minimum of 4 of 7 items, 4 of 6 items, etc), the subscale score will be calculated by multiplying the sum of subscale by the number of items in the subscale, then dividing by the number of items actually answered:
  Subscale score= (sum of item scores x N of items in subscale)/ N of items answered
- If at least 50% of the domain items are missing, that domain will be treated as missing and thus NE. The total score for each variable (FACT-G, FACT-P TOI and FACT-P Total) is then calculated as the sum of the un-weighted prorated scores. If a domain score is NE, any health related quality of life (HRQoL) variable which these domains contribute to is also termed NE. For example, for the FACT-P TOI variable, if PWB is NE at a visit, the FACT-P TOI variable is also NE at this visit. Also, the FACT-P total score cannot be computed if any of the domain scores is NE.

Visit responses

Table 14 details how visit responses will be defined for the FACT-P total, FACT-G total, TOI, FAPSI-6, PCS, PWB, and FWB scores (Cella et al 2009, Webster et al 2003).

A visit response of 'Improved', 'No Change', 'Worsened' or Not evaluable, as defined according to Table 14 will be calculated for each patient for FACT-P Total score, FACT-G Total score, FAPSI-6, TOI, PCS, FWB and PWB subscales.

Table 14 Definition of visit response for FACT-P, FACT-G, FAPSI-6, TOI, PCS FWB, PWB

FACT-P scale	Change from baseline	Visit response
FACT-P-Total	$\geq +10$	Improved
	≤ -10	Worsened

FACT-P scale	Change from baseline	Visit response
	Otherwise (i.e. > -10 and < +10)	No change
	Missing/non-calculable score	Not evaluable
FACT-G-Total	≥ +7	Improved
	≤ -7	Worsened
	Otherwise (i.e. > -7 and < +7)	No change
	Missing/non-calculable score	Not evaluable
FAPSI-6	≥ +3	Improved
	≤ -3	Worsened
	Otherwise (i.e. > -3 and < +3)	No change
	Missing/non-calculable score	Not evaluable
TOI	≥ +9	Improved
	≤ -9	Worsened
	Otherwise (i.e. > -9 and < +9)	No change
	Missing/non-calculable score	Not evaluable
PCS, FWB, PWB	≥ +3	Improved
	≤ -3	Worsened
	Otherwise (i.e. > -3 and < +3)	No change
	Missing/non-calculable score	Not evaluable

FACT-G, Functional Assessment of Cancer Therapy-General; FACT-P, Functional Assessment of Cancer Therapy-Prostate Cancer; FAPSI-6, Functional Assessment of Prostate Cancer Symptoms Index-6; FWB, Functional Well-Being; PCS, Prostate cancer symptoms; PWB, Physical Well-Being; TOI, Trial outcome index.

Note, for some patients it will not be immediately possible to obtain a visit response for a particular subscale, for example:

- Patients with no baseline score for a particular subscale, or no baseline data at all
- Patients whose baseline subscale score is too close to the maximum or minimum possible score to allow an increase or decrease of the specific size to be observed.
  - For patients whose baseline score is greater than the maximum possible score for that subscale minus the score needed to satisfy improvement, the best visit response possible will be "No Change".
  - For patients whose baseline score is less than the threshold needed for worsening (a baseline FACT-P Total score < 10) all post-baseline visit responses will be considered not-calculable.

Best QoL response

The criteria shown in Table 15 will be used to assign a best QoL response based on individual visit responses as defined in Table 14. Patients with no evaluable baseline or no post-baseline PRO assessments with be assigned to 'non-evaluable' for best QoL response.

Table 15 Best QoL Response criteria

Overall score response	Criteria
Non-evaluable	Has no evaluable baseline or no post-baseline PRO assessments.
Improved	Two consecutive visit responses of 'improved'
No change	Does not qualify for overall score response of 'improved'. Two consecutive visit responses of either 'no change', or 'improved' and 'no change'
Worsened	Does not qualify for overall score response of 'improved' or 'no change'. A visit response of 'worsened'
Other	Does not qualify for 1 of the above

Overall improvement rate will be defined as the proportion of patients with a best overall QoL response of "Improved" based on Table 15.

Time to deterioration for FACT-P

Time to deterioration in HRQoL as measured by FACT-P total score will be defined as the interval from the date of randomization until the date of the first sustained clinically meaningful deterioration that is confirmed at a subsequent visit at least 3 weeks apart (except if it was the patient's last available assessment) or death (by any cause in the absence of a clinically meaningful deterioration), regardless of whether the patient discontinues study drug(s) or receives another anticancer therapy prior to the deterioration in FACT-P total score. Death will be included as an event only if it occurs within 2 PRO assessment visits from the last available PRO assessment. For patients who receive a subsequent anti-cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy), data will only be included until the start date of the subsequent anti-cancer therapy. Time to deterioration as measured by FACT-P TOI, FAPSI-6, FACT-G, PCS, PWB and FWB will be derived similarly.

A sustained clinically meaningful deterioration will be defined as a 'worsened' response per definition in Table 14, which must be sustained at the next scheduled visit and there must be no relative improvement between the two visit responses of 'worsened'.

Radiological progression will not be considered as deterioration in symptoms.

Note, under the same principles applied to the primary outcome variable (rPFS), time to deterioration will be derived regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy prior to symptom deterioration. A number of situations will lead to a patient's time to deterioration of HRQoL endpoints being censored. These are:

- If a patient either dies or meets the criteria for deterioration after 2 or more missed assessments, then the patient will be censored at the time of the latest evaluable assessment. These patients will be presented as e.g. "Censored FACT-P total score" in summaries.
- Patients who have not met the criteria for symptom deterioration or died at the time of analysis will be censored at the time of the latest evaluable assessment:
  - The censoring date will be the date of the last assessment that led to evaluable being assigned for FACT-P total score. These patients will be presented as alive and deterioration-free in summaries.
  - Patients with no evaluable baseline or post-baseline data will be censored at Day 1 unless they die within 2 visits of baseline (in which case their date of death will be used). These patients will be presented as censored in summaries.
- Patients who receive subsequent anti-cancer therapy
  - The censoring date will be the date of the last assessment prior to the start date of subsequent anti-cancer therapy that led to evaluable being assigned for FACT-P total score. These patients will be presented as alive and deterioration-free in summaries.
  - Patients with no evaluable baseline or post-baseline data will be censored at Day 1 unless they die within 2 visits of baseline (in which case their date of death will be used). These patients will be presented as censored in summaries. This will include patients with baseline FACT-P total score <10.
- Patients whose baseline subscale score is close to the minimum possible.
  - For patients whose baseline score is less than the threshold needed for worsening (e.g., a baseline FACT-P total score of < 10), time to deterioration will be censored at Day 1 unless they die within 2 visits of baseline. Patients who haven't died will be presented as "Censored FACT-P Total Score" in summaries.

The time to deterioration of HRQoL will be derived based on assessment dates, not visit dates.

PRO compliance

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

- Received form: a form that has been received and has a completion date and at least one individual item completed.
- Expected form: a form that is expected to be completed at a scheduled assessment time e.g. a form from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under FACT-P 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. FACT-P forms are to be completed for 12 weeks after either progression or treatment discontinuation (whichever comes second). However, all data should be used.
- Evaluable form: a form with a completion date and at least one subscale that is non-missing.
- Overall FACT-P compliance rate is defined as the total number of evaluable forms across all time points, divided by total number of forms expected to be received across all time points multiplied by 100.
- Overall patient compliance rate is defined for each randomised treatment group as the total number of patients with both an evaluable baseline and at least one evaluable follow-up form (as defined above), divided by the total number of patients expected to have completed at least a baseline FACT-P form multiplied by 100.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable 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.

3.3.10 HRR gene mutation status

Mutation status will be determined using a tumour tissue test ([redacted]), a ctDNA-based test ([redacted]), and a germline blood test ([redacted]).

Patients will be classified into three subgroups based on HRR gene mutation status:

- HRRm: Any deleterious or suspected deleterious HRR gene mutation detected.
- Non-HRRm: No deleterious or suspected deleterious HRR gene mutation detected.
- Unknown: Patients where mutation testing was not performed or where mutation testing failed.

3.4 [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

3.5 Patient Reported Outcome (PRO) Variables

3.5.1 BPI-SF

The BPI-SF is a validated, 15-item domain-specific instrument designed to assess the severity of pain and the impact/interference of pain on daily functions (Cleeland and Ryan 1994). The BPI-SF will be scored according to the user guide (Cleeland 2009). All BPI-SF pain items including "worst pain" is scored on a 0-10 numeric rating scale (NRS) with 0=No Pain and 10=Worst Pain Imaginable. This instrument consists of 2 domains: pain severity and pain interference.

Pain severity

The pain severity domain consists of 4 items (item #3, item #4, item #5, and item #6) which assess pain at its "worst," "least," "average," and "now" (current pain) respectively on the 11-point NRS. These 4 items may be averaged (all items must be non-missing) as a composite pain severity score or they may be interpreted individually (Dworkin et al 2005, Turk et al 2006, Dworkin et al 2008, Food and Drug Administration 2009). In this study, the "worst pain" (item 3) will be used as a single item in assessing pain progression. A composite pain severity score from all the 4 items will also be evaluated as 'pain severity progression'. A 2- or more point change in the average pain severity or in "worst pain" item is considered clinically meaningful.

Pain interference

The pain interference domain score is a mean of 7 items: general activity (item #9A), mood (item #9B), walking ability (item #9C), normal work (item #9D), relations with other people (item #9E), sleep (item #9F), and enjoyment of life (item #9G), each scored on an 11-point NRS from 0 (Does not interfere) to 10 (Completely interferes). Based on the BPI-SF scoring manual, the following items are not used in scoring pain severity or pain interference domains: items #1, #2, #7 and #8. Item #7 (a free text field) describing pain medication use is captured separately in more detail using the Analgesic Log.

PRO compliance

Summary measures of overall compliance and compliance over time will be derived for the BPI-SF. These will be based upon the compliance derivations described for FACT-P (Section 3.3.9).

3.5.2 Analgesic use scoring

Opiates consumed by patients will be converted into oral morphine equivalents (OMEs) as defined in Chung et al 2014. The AQA developed by Chung et al 2014 will be used to quantify and score analgesic use in the study. The AQA is an 8-point scale that assigns a score as follows:

- 0=No analgesic
- 1=Non-opioid analgesics
- 2=Weak opioids (eg, codeine, tramadol)
- 3=Strong opioids ≤75 mg OME per day
- 4=Strong opioids >75–150 mg OME per day
- 5=Strong opioids >150–300 mg OME per day
- 6=Strong opioids >300–600 mg OME per day
- 7=Strong opioids >600 mg OME per day

The average daily OME will require at least 4 days of data and will be used to assign the AQA score. An increase of 1 point or more in the AQA score from a starting value of 1 or higher OR ≥ 2 points in AQA score from a starting value of 0 is considered a clinically meaningful increase in opiate use. Similarly, a decrease of 1 point or more in the AQA score from a starting value of 2 or higher is considered a clinically meaningful decrease in opiate use.

Missing data

Analgesic or pain medication use allows patients to add new medications as "Other" to the handheld device.

A patient may have taken more than 1 dose, in which case the total dose will be equal to the 'OME amount (single dose)' multiplied by the number of doses. Below 'OME value' refers to the total OME amount for the medication entry.

In the case where there are reconciled or unreconciled "other" pain medication entries in the analgesic log and the OME value is missing, but the medication is clearly identified as a non-opioid (eg, dexamethasone), then the OME value will be set to 0. OME values will not be imputed for reconciled "other" medications which are not clearly identified as non-opioids.

In the case where there are unreconciled "other" pain medication entries which are not clearly identified as non-opioids, OME values will be imputed at two levels while AQA scores will be imputed where OME values cannot be assigned as follows:

Daily completion level:

- If additional pain medications were taken alongside "Other" for a specific day, the highest OME value of the pain medications (based on completed entries) will be selected as the imputed value for each unreconciled "other" entry for the specific day.

7-day completion period level:

- If no additional pain medications were taken alongside "Other" for a specific day, the highest OME value of pain medications (based on completed entries) across the 7 days of assessments will be selected as the imputed value for each unreconciled "other" entry.

AQA score level:

- If no additional pain medications were taken alongside "Other" across the 7 days of assessments, the highest AQA value from all previous visits will be selected as imputed AQA value for the time point. If there are only unreconciled "Other" medications at the baseline assessment, then an average AQA score of 1 will be assumed.

- If additional pain medications taken alongside "Other" over the 7 days of assessments, and they are all non-opioids, then AQA score of 1 will be assigned (i.e. non-opioid analgesics).

If no additional pain medications were taken alongside "Other" over the 7 days of assessments and there are no AQA values from all previous visits, then AQA score of 1 will be assigned (i.e. non-opioid analgesics). This is based on the assumption that opioid analgesics are likely to be captured as concomitant medication and be included in the Analgesic Log for patient completion. Thus, as applicable, unreconciled "other" pain medication will be assigned an AQA score of 1 instead of being considered missing.

3.6 Safety Variables

Throughout the safety section, first dose of study treatment refers to the first dose of either olaparib/placebo or abiraterone. The last dose of study treatment refers to latest discontinuation date of olaparib/placebo and abiraterone.

3.6.1 Exposure

Study drug exposure (days) will be defined as time from first dose of olaparib/placebo, up to and including the, last day that the dose of olaparib/placebo was greater than 0 mg. Exposure to abiraterone will be calculated in the same way.

Exposure (i.e. duration of treatment) will be defined as follows:

Total (or intended) exposure of study treatment

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

Actual exposure of study treatment

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

The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Missed or forgotten doses

Missed and forgotten doses should be recorded on the EX and EX1 modules for olaparib/placebo, and abiraterone respectively as drug interrupted with the reason recorded as "Subject forgot to take dose". These missed or forgotten doses will not be included as dose interruptions in the summary tables but the information will appear in the listing for dosing. However, these missed and forgotten doses will be considered in the derivation of actual exposure.

Safety Follow-up

Total Safety Follow-up = min((last dose date + 30), date of withdrawal of consent, date of death, date of DCO) – first dose date +1.

3.6.2 Dose intensity

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

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the or the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule.

Percentage intended dose (PID) is the percentage of the actual dose delivered relative to the intended dose through to progression. PID will be defined as follows:

- 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.

Intensity of olaparib/placebo and abiraterone acetate will be summarised separately. The intended cumulative dose is defined as 300mg olaparib/placebo twice daily and 1000mg abiraterone acetate once daily.

3.6.3 Adverse events

AEs and SAEs will be collected throughout the study, from date of informed consent until 30 days after the last dose of study treatment. Events will be defined as treatment emergent if they onset, or worsen (by investigator report of a change in intensity) on or after the date of first dose and up to and including 30 days following the date of last dose of study medication. The Medical Dictionary for Regulatory Activities (MedDRA) (using the latest or current MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute (NCI) Common Terminology Criteria for AEs (using the CTCAE version referenced in the Clinical Study Protocol).

Other significant adverse events (OAE)

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

AEs of special interest

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

Adverse events of special interest for olaparib are:

- Important Potential Risks of myelodysplastic syndrome (MDS)/acute myeloid leukaemia (AML)
- New primary malignancy (other than MDS/AML)
- Pneumonitis

Other categories may be added as necessary or existing terms may be merged. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician will review the AEs of interest and identify which higher-level terms and which preferred terms contribute to each AESI. Further reviews may take place prior to database lock (DBL) to ensure any further terms not already included are captured within the categories.

3.6.4 Concomitant medications

Concomitant medications will be classified according to the WHO Drug Dictionary.

Concomitant medications are considered as treatment emergent (taken during study treatment) if they:

- Started on or prior to the date of first dose and either stopped or had a status of 'ongoing' on or after the date of first dose OR
- Started after the date of first dose up to the date of last dose plus 30 days.

3.6.5 Laboratory assessments

Blood samples for determination of clinical chemistry, hematology and coagulation will be taken at each scheduled visit and urine samples to determine urinalysis will be taken at screening and Day 1 visits. The laboratory parameters to be collected are given in Section 8.2.1 of the protocol.

3.6.6 Vital signs

Vital signs, including blood pressure (BP) (mmHg), pulse rate (beats/minute), body temperature (°C) and weight (kg), will be assessed at timelines as specified in the Schedule of

Activities (SoA) (Protocol Section 1.1) and as clinically indicated. Changes in vital signs should be recorded as an AE, if applicable.

3.6.7 Physical examination

Physical examination assessments will be performed at timelines as specified in the SoA (Protocol Section 1.1) and as clinically indicated.

3.6.8 Electrocardiogram

Twelve-lead resting ECGs are required within 7 days of start of study treatment and will be additionally performed every 12 weeks during the study treatment period, study treatment discontinuation visit, 30-days follow-up visit, and if clinically indicated at any other time for all patients.

ECGs will be recorded at 25 mm/sec. 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.

3.7 Pharmacokinetic Variables

Pharmacokinetic sampling will be performed in a subset of patients, planned to include approximately 50 patients per treatment group (i.e., olaparib+abiraterone or placebo+abiraterone), on Study Day 29 at Pre-dose (- 30 min ± 15 min), and post dose at 30 min ± 15 min, 2 ± 0.5 hour, 3 ± 0.5 hour (for abiraterone and Δ4-abiraterone only), 5 ± 0.5 hour and 8 ± 1 hour.

The sample bioanalysis will be performed by analytical test sites on behalf of AstraZeneca, using appropriate validated bioanalytical methods and analytical methods.

The merging of PK concentration data with actual PK sampling times will be performed by IQVIA. The statistical analysis of the PK variables will be performed by Phastar who will also be responsible for the summaries, figures, and data listings. IQVIA and Phastar are CRO organisations.

The plasma concentration-time data will be analysed using non compartmental analysis (NCA) to determine the PK of olaparib, abiraterone, and Δ4-abiraterone (a metabolite of abiraterone) at steady state and to evaluate the effect of olaparib on abiraterone PK.

If deemed appropriate, the plasma concentration-time data may be analysed by nonlinear mixed-effect modelling to determine the PK of olaparib, abiraterone, and Δ4-abiraterone at steady state and to evaluate the effect of olaparib on abiraterone PK. This analysis will be reported separately from the CSR.

3.7.1 Calculation or derivation of pharmacokinetic variables

The NCA will be carried out by the [redacted] on behalf of AstraZeneca R & D. Pharmacokinetic parameters will be derived using non-compartmental methods with Phoenix® WinNonlin® Version 8.1, or higher.

The actual sampling times will be used in the plasma PK parameter calculations. If actual times are missing, nominal times may be used by agreement with the sponsor.

If sufficient concentration data are available for estimation, the following multiple dose PK parameters will be calculated for olaparib, abiraterone and Δ4-abiraterone for each patient:

- $C_{ss,max}$: Maximum plasma concentration at steady state.
- $C_{ss,min}$: Minimum plasma concentration at steady state.
- $t_{ss,max}$: Time to reach maximum plasma concentration at steady state.
- $AUC_{(0-8)}$: Area under the plasma concentration-time curve from time zero to 8 hours post-dose.

Additional PK parameters not listed as outcome parameters in the objective section (Table 2) will be determined if appropriate. These parameters include: the apparent clearance at steady state (CL_{ss}/F), the area under the plasma concentration-time curve at steady state (AUC_{ss}) for olaparib, and the metabolite to parent ratios for $C_{max,ss}$ ($MRC_{max,ss}$) , $C_{min,ss}$ ($MRC_{min,ss}$) and $AUC_{(0-8)}$ ($MRAUC_{(0-8)}$) for Δ4-abiraterone. In addition, the time of the last concentration (t_{last}) will also be determined as a diagnostic parameter and will be listed only.

$C_{max,ss}$, $C_{min,ss}$, $t_{max,ss}$ and t_{last} will be obtained directly from the observed concentration-versus-time data. AUC_{ss}, and $AUC_{(0-8)}$ will be calculated by linear up/log down trapezoidal summation. CL_{ss}/F will be calculated from dose/AUC_{ss}.

Since the olaparib PK concentration-time profile is expected to be at steady state by Day 29 of twice daily dosing, the pre-dose (trough) concentration ($C_{min,ss}$) will also be used, where data allow, as the 12 hour post-dose sample, in order to calculate the olaparib AUC_{ss}.

Should there be any concentrations that are Not Quantifiable (NQ) from the time of pre-dose sampling (t=0) up to the time of the first quantifiable concentration then they will be set to a value of 0 for the NCA. After this point, samples that are below the lower limit of quantification (< LLOQ) will be set to missing for all concentration profiles. If two or more consecutive NQ concentrations are followed by quantifiable concentrations in the terminal portion of the concentration-curve, the profile will be deemed to have terminated and therefore these quantifiable values will be set to missing for the calculation of the PK parameters unless there is a scientific rationale not to do so. If an entire concentration-time profile is NQ, the profile will be excluded from the PK analysis.

4 ANALYSIS METHODS

4.1 General principles

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

- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarised by the number of observations, mean, standard deviation, median, upper and lower quartiles, minimum, and maximum. For log-transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables will be summarised by frequency counts and percentages for each category.
- Unless otherwise stated, percentages will be calculated out of the population total for the corresponding treatment group. Overall totals will be calculated for baseline summaries only.
- For continuous data, the mean and median will be rounded to 1 additional decimal place compared to the original data. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.
- For categorical data, percentages will be rounded to 1 decimal place.
- SAS® version 9.2 (as a minimum) will be used for all analyses.

Results of statistical analyses will be presented using corresponding 95% confidence intervals and p-values for 2-sided tests, where appropriate.

In general, for efficacy endpoints the last non missing measurement prior to randomization will be considered the baseline measurement. However, if an evaluable assessment is only available after randomization but before the first dose of randomised treatment then this assessment will be used as baseline. For safety and PRO endpoints, the last observation before the first dose of study treatment will be considered the baseline measurement unless otherwise specified. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose. If neither time nor a nominal pre-dose indicator are present assessments will be considered pre-dose if such procedures are required by the protocol to be conducted before first dose.

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The percentage change from baseline will be calculated as (post-baseline value - baseline value) / baseline value x 100. Efficacy and Health-Related Quality of Life (HRQoL) data will be summarised and analysed based upon the FAS. Safety and treatment exposure data will be summarised based upon the safety analysis set. Study population and demography data will be summarised based upon the FAS.

Additional post hoc analyses may be conducted to investigate the impact of COVID-19 on study endpoints.

4.2 Analysis methods

Table 16 Formal Statistical Analyses to be Conducted and Pre-planned Sensitivity Analyses

Endpoints Analysed	Notes
rPFS	Primary analysis: Stratified log-rank test (to calculate p-value) Hazard ratio using Cox proportional hazards model (with ties=Efron and the stratification variables as covariates, in line with the primary pooling strategy) KM Plot Key sensitivity analysis[a]: stratified log-rank test assessed for all patients by BICR per RECIST 1.1 and PCWG-3 criteria Additional sensitivity analyses[a]: Evaluation-time bias (using midpoint between time of progression and previous evaluable RECIST assessment): stratified log-rank test Attrition bias (using alternative censoring rules): stratified log-rank test Censoring bias: KM plot of the time to censoring Sensitivity analysis using unequivocal clinical progression in addition to radiological progression: stratified log-rank test Sensitivity analysis for confirmation of bone progression: stratified log-rank test Sensitivity analysis censoring patients with subsequent therapy or discontinuation of study drug: stratified log-rank test 'Deviation bias' sensitivity analysis excluding patients with deviations that may affect the efficacy (deviations 1, 2, 3, 4 and 6) of the trial therapy if > 10% of patients in either treatment group have IPDs Subgroup analyses: Hazard ratio using Cox proportional hazards model (with ties=Efron and the stratification variables as covariates). Forest plot
Overall survival, TFST, TTPP, TTPP (non-opiate users at baseline), Time to opiate use, Time to first symptomatic skeletal related event	Stratified log-rank test (to calculate p-value) Hazard ratio using Cox proportional hazards model (with ties=Efron and the stratification variables as covariates, in line with the primary pooling strategy) KM plot Forest plot
PFS2	Stratified log-rank test (to calculate p-value)

Endpoints Analysed	Notes
	Hazard ratio using Cox proportional hazards model (with ties=Efron and the stratification variables as covariates, in line with the primary pooling strategy) KM plot
FACT-P	Logistic regression, adjusting for metastases category and docetaxel treatment at mHSPC stage, in line with the primary pooling strategy MMRM Time to deterioration in FACT-P (FACT-P Total score, FACT-G total score, TOI, FWB, PWB, PCS and FAPSI 6): Stratified log-rank test (to calculate p-value) Hazard ratio using Cox proportional hazards model (with ties=Efron and the stratification variables as covariates, in line with the primary pooling strategy) Forest plot
BPI-SF[b]	MMRM
HRR gene mutation status	Summary of gene mutation status
ORR (investigator and BICR assessments) - ORR -	Logistic regression adjusted in line with the primary pooling strategy. If there are not at least 5 responses across both treatment groups, then a Fisher's exact test using mid p-values will be used.
DCR (investigator and BICR assessments) - DCR -	Proportion of patients achieving disease control (yes/no) presented with 95% CIs
DoR (investigator and BICR assessments) - DoR -	Summarised descriptively KM plot
BoR (investigator and BICR assessments) - BoR -	Summarised descriptively
CTC conversion rate	Proportion of patients achieving CTC conversion at any time presented with 95% CIs
PSA response	Proportion of patients achieving PSA response presented with 95% CIs
Time to PSA progression	Stratified log-rank test (to calculate p-value) Hazard ratio using Cox proportional hazards model (with ties=Efron and the stratification variables as covariates, in line with the primary pooling strategy)

KM, Kaplan-Meier; MMRM, Mixed model for repeated measures.

[a] See Section 4.2.2.2 for further details

[b] BPI-SF Item 3, BPI-SF pain severity domain, BPI SF pain inference domain

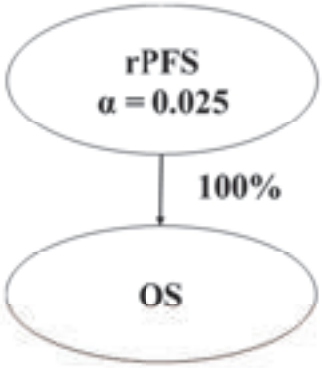
4.2.1 Multiplicity

The statistical hypotheses are described and defined in Section 1.2. The 1-sided alpha of 0.025 is fully allocated to H_{a0} (rPFS). If the result for rPFS is statistically significant, H_{b0} (OS) will be tested in a hierarchical fashion. A multiplicity testing procedure based on the graphical approach in group sequential trials of Maurer and Bretz (Maurer and Bretz 2013), analogous to a simple sequential gatekeeping method, strongly controls the overall family-wise 1-sided error rate of 2.5%.

The rPFS end-point will be tested at DCO1 and DCO2. The OS end-point will be tested at DCO1, DCO2 [redacted] For each endpoint with interim analysis, O'Brien and Fleming spending function (O'Brien and Fleming 1979), calculated based upon actual observed events, will be used to strongly control the overall type 1 error, with the restriction that alpha spend for the OS interim analysis at DCO1 will not exceed 0.0005. (Lan and DeMets 1983).

The multiplicity strategy is illustrated in Figure 2:

Figure 2 Multiplicity strategy maintaining overall type 1 error rate



OS, Overall survival; rPFS, Radiological progression-free survival.

Details of the estimated alpha spending plan are shown in Table 5 (see Section 1.3).

4.2.2 Analysis of the primary efficacy variable (rPFS)

The primary objective of this study is to assess the efficacy of olaparib+abiraterone combination therapy vs placebo+abiraterone in terms of rPFS assessed by the investigator in patients with mCRPC who have received no prior cytotoxic chemotherapy or NHA.

The rPFS end-point will be analysed using a stratified log rank test to calculate a 2-sided p-value, based on the following stratification factors:

- Metastases: bone only vs visceral vs other
- Docetaxel treatment at mHSPC stage: yes vs no

Although it is expected that there will be enough rPFS events in each strata (where strata are defined as categories formed from the six combinations of metastases and docetaxel treatment at mHSPC stage) to allow a meaningful analysis, if a stratum for either treatment arm contains less than 5 events, then a pooling strategy will be employed. If there are less than 5 events in any strata, the factor (Docetaxel treatment at mHSPC stage) will be removed first. If there are still less than 5 events in any strata the factor (Metastases) will also be removed and an unstratified analysis will be conducted. All sensitivity analyses and secondary endpoints will be conducted in accordance with the primary pooling strategy for the analysis of rPFS. If a model does not converge when using the stratification variables from the primary pooling strategy then this will result in collapsing of strata in line with the pooling strategy until the minimum 5 event criterion is achieved. Unstratified analyses will be conducted for any secondary endpoints that still do not conform to the 5-event rule per stratum.

Stratification variables will be defined according to data from the IXRS. If there are any patients who were mis-stratified, as a sensitivity analysis, the primary analysis may be carried out using the baseline data collected in the eCRF.

A hazard ratio (HR) and corresponding 95% confidence interval (CI) will be estimated using a Cox Proportional Hazards Model (with ties=Efron and the stratification variables as covariates) with the CI calculated using a profile likelihood approach.

The HR (olaparib+abiraterone vs. placebo+abiraterone) together with its corresponding 95% CI, and the 2-sided p-value obtained from the stratified log-rank test, will be presented (a HR less than 1 will favour olaparib+abiraterone).

A Kaplan-Meier (KM) plot of rPFS will be presented by treatment group. Summaries of the number and percentage of patients experiencing a progression or death event, and the type of event (RECIST 1.1 or bone progression or death) will be provided along with median rPFS for each treatment arm. For each treatment arm, the rPFS rate and its 95% CI will be summarized

every 6 months for minimum of 24 months until <20 patients in the risk set in either treatment arm.

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 will be fitted to assess the extent to which this represents random variation.

The primary analysis will be based on the investigator assessment of rPFS using all scans regardless of whether they were scheduled or not.

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

4.2.2.1 Subgroup analysis

Subgroup analyses will be conducted for the rPFS endpoint. The purpose of subgroup analyses is to assess consistency of treatment effect across potential or expected prognostic factors.

If there are too few responders or 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 either treatment group per subgroup), the relationship between that subgroup and the endpoint will not be formally analysed. In this case, only descriptive summaries will be provided.

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

- Metastases (bone only, visceral or other)
- Docetaxel treatment at mHSPC stage (yes or no)

Values collected on the eCRF will be used to define subgroups for stratification factors.

Additional subgroups of interest include:

- HRRm status subgroup (HRRm, non-HRRm, unknown) based on a ctDNA-based test ([redacted])
- HRRm status subgroup (HRRm, non-HRRm, unknown) based on a tissue test ([redacted])
- HRRm status subgroup (HRRm, non-HRRm, unknown) based on a germline blood test ([redacted]) (when data are available)
- Eastern Cooperative Oncology Group (ECOG) performance status at baseline (0 or 1)
- Age at randomisation (<65, ≥65)
- Region (Asia, Europe, North and South America)
- Race (White, Black/African-American, Asian, Other)
- Baseline Prostate specific antigen (PSA) (above/below median baseline PSA of the patients across both treatment groups)

KM plots will be produced for each subgroup by treatment group.

In each subgroup, the HRs for radiological progression by investigator assessment (olaparib+abiraterone combination therapy vs. placebo+abiraterone) and associated CIs will be estimated using a Cox proportional hazards model with the Efron method being used for handling ties) that contains the treatment term, factor and treatment-by-factor interaction term. The treatment effect HRs for each treatment comparison along with their confidence intervals will be obtained for each level of the subgroup from this single model. The subgroup HRs and 95% CIs will be presented using a forest plot including the HR and 95% CI from the overall population (using the primary analysis). [redacted]

Consistency of treatment effect between subgroups

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 (stratification factors) and all covariate-by-treatment interaction terms, with one that excludes the interaction terms. This will be assessed at the 2-sided 10% significance level. If the fit of the model is not significantly improved, then it will be concluded that overall the treatment effect is consistent across the stratification factor 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. 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. The p-values reported will represent those from the final model resulting from stepwise backwards selection; the 'selection model'.

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

4.2.2.2 Sensitivity analysis

A sensitivity analysis will be conducted using rPFS as assessed for all patients by BICR per RECIST 1.1 and PCWG-3 criteria.

In addition, to assess the sensitivity of the primary rPFS analysis, the following supportive analyses will be performed:

(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 assessment (RECIST or PCWG-3) will be analysed as described for the primary analysis of rPFS. Note that midpoint values resulting in non-integer values should be rounded down. For patients whose death was treated as a rPFS event, the date of death will be used to derive the rPFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric schedules (Sun and Chen 2010).

(b) Attrition bias

Attrition bias will be assessed by repeating the primary rPFS analysis except that the actual rPFS event times, rather than the censored time, of patients who progressed or died in the absence of progression immediately following 2, or more, missed tumor assessments will be included. In addition, patients who take subsequent therapy prior to progression or death will be censored at their last evaluable assessment prior to taking the subsequent therapy.

(c) Censoring bias

A Kaplan-Meier plot of the time to censoring will be produced where the censoring indicator of the primary rPFS analysis is reversed.

(d) Sensitivity analysis using unequivocal clinical progression in addition to radiological progression

Repeating primary rPFS analysis with the addition of unequivocal progression as an event. Where unequivocal clinical progression is defined as, cancer pain requiring initiation of opioids, need to initiate cytotoxic chemotherapy, radiation therapy or surgical intervention for complications due to tumor progression or deterioration in ECOG performance to ≥ Grade 3 (i.e. baseline ECOG =0 or 1 and a post baseline ECOG >=3).

The initiation of cytotoxic chemotherapy includes cytotoxic chemotherapy and receiving prohibited anti-cancer medication/therapy during study treatment period as per the protocol Table 7.

The initiation of radiation therapy includes use of radiation therapy to relieve or prevent skeletal symptoms and use of current/post palliative radiotherapy for prostate cancer. The event is the earlier date of the rPFS event date and the unequivocal clinical progression date.

(e) Sensitivity analysis for confirmation of bone progression

Repeat primary rPFS analysis with revised confirmation criteria for bone progression where bone progression accompanied by unequivocal clinical progression does not require a confirmatory bone scan. The only difference from the primary analysis is that the second bone progression is not required for a patient's bone progression status when the patient at any time experiences an unequivocal clinical progression. In this analysis, if a patient has a bone progression and an unequivocal clinical progression, this unequivocal clinical progression is used as the confirmation of bone progression.

(f) Sensitivity analysis censoring patients with subsequent therapy or discontinuation of study drug

Repeat primary rPFS analysis censoring patients with subsequent therapy or discontinuation of study drug prior to progression (censor at the last evaluable assessment prior to taking the subsequent therapy or at discontinuation.).

(g) Deviation bias

The important protocol deviations will be reviewed and determined before database lock. If greater than 10% of patients in either treatment group have important protocol deviations that may affect the efficacy, the rPFS analysis will be repeated excluding these patients. Details are in Section 2.2.

4.2.3 Analysis of secondary variables

4.2.3.1 Overall survival

Analysis of OS will be performed using the same method as used for the analysis of rPFS, stratified in accordance with the primary pooling strategy.

A KM plot of OS will be presented by treatment group. Summaries of the number and percentage of deaths and those alive and censored will be provided along with median time to death for each treatment arm. For each treatment arm, the number and percentage of those alive will be summarized every 6 months for a minimum of 48 months until there are <20 patients in the risk sets of either treatment arm.

The subgroup analyses described for rPFS will be repeated for OS. Equivalent subgroup analyses (except the global interaction test) will be conducted comparing OS between treatments as detailed for rPFS in section 4.2.2.1, using the same methodology and model. KM plots will be produced for each subgroup according to treatment group. HR's and 95% CIs will be presented on a forest plot including the HR and 95% CI from the overall population.

4.2.3.2 Time to first subsequent anticancer therapy or death

Time to first subsequent anticancer therapy will be analysed using the same methods as in the analysis of rPFS, stratified in accordance with the pooling strategy.

In addition, medians and a Kaplan-Meier plot of the time to the start of subsequent therapy will be presented by treatment arm and the time between progression and starting subsequent therapy will be summarised per treatment arm. For each treatment arm, the number and percentage of those with no subsequent anticancer therapy will be summarized every 6 months for minimum of 24 months until 20 patients with evaluable data. In patients who received a subsequent anti-cancer therapy, a summary table of first subsequent anti-cancer therapies by treatment arm will be provided, as well as best response to first subsequent anti-cancer therapy by treatment arm.

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

4.2.3.3 Time to pain progression

Time to pain progression will be analysed using the same methods as in the analysis of rPFS, stratified in accordance with the primary pooling strategy.

A KM plot of time to pain progression will be presented by treatment group. Summaries of the number and percentage of patients experiencing pain progression will be provided along with median time to pain progression for each treatment arm. This will be repeated for time to pain progression restricted to patients who are non-opiate users at baseline. For each treatment arm, the number and percentage of those with no pain progression will be summarized every 6 months for minimum of 24 months until there are <20 patients in the risk sets of either treatment arm.

4.2.3.4 Time to opiate use

Time to opiate use will be analysed using the same methods as in the analysis of rPFS, stratified by in accordance with the primary pooling strategy.

A KM plot of time to opiate use will be presented by treatment group. Summaries of the number and percentage of patients using opiate will be provided along with median time to opiate use for each treatment arm. For each treatment arm, the number and percentage of those with no opiate use will be summarized every 6 months until there are <20 patients in the risk sets of either treatment arm.

4.2.3.5 Time to first symptomatic skeletal related event

Time to first symptomatic skeletal–related event will be analysed using the same methods as in the analysis of rPFS, stratified in accordance with the primary pooling strategy.

A KM plot of time to SSRE will be presented by treatment group. Summaries of the number and percentage of patients experiencing SSRE and those who are censored will be provided along with median time to SSRE for each treatment arm. For each treatment arm, the number and percentage of those with SSRE free will be summarized every 6 months until there are <20 patients in the risk sets of either treatment arm.

4.2.3.6 Time to second progression or death (PFS2)

Time from randomisation to second progression on next-line (immediately after study treatment) anticancer therapy will be analysed using the same methods as in the analysis of rPFS, stratified in accordance with the primary pooling strategy.

A KM plot of time to second progression will be presented by treatment group. Summaries of the number and percentage of patients experiencing second progression or death and those who are censored will be provided along with median time to second progression for each treatment arm. For each treatment arm, the number and percentage of those with event free

will be summarized every 6 months until there are <20 patients in the risk sets of either treatment arm.

4.2.3.7 Pain severity

Change from baseline in the BPI-SF pain severity domain will be analysed using a mixed model for repeated measures (MMRM) analysis of all the post-baseline pain severity scores for each visit. The study discontinuation visit and the safety follow-up visit will be excluded from this analysis. Restricted maximum likelihood (REML) estimation will be used. The model will include treatment, visit and treatment by visit interaction as explanatory variables and the baseline pain severity score as a covariate, along with the baseline pain severity score by visit interaction and the stratification variables determined by the primary pooling strategy. Treatment, visit, treatment by visit interaction, baseline pain severity score, baseline pain severity score by visit interaction, and the stratification variables will be fixed effects in the model. The treatment by visit interaction will remain in the model regardless of significance. An unstructured covariance matrix will be used to model the within-patient error and the Kenward-Roger approximation will be used to estimate the degrees of freedom.

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

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

4.2.3.8 Pain interference

Change from baseline in BPI-SF pain interference domain will be analysed using a MMRM as described in Section 4.2.3.7.

4.2.3.9 FACT-P

FACT-P total score, FACT-G total score, TOI, FWB, PWB, PCS, and FAPSI-6, SWB and EWB will be summarised using mean, standard deviation, median and range by treatment group for each visit until there are less than the minimum of 20 or one third of patients with evaluable data. The absolute and change from baseline scores for each time point will be calculated by treatment group. Graphical plots of the mean score, including change from baseline, and associated 95% CI by scheduled visits/time points in the study will be produced.

The proportion of patients with best responses of 'Improved', 'No Change' and "Worsened" will be summarised descriptively as number of patients and corresponding percentages for

each category by treatment group. The proportion of patients with best response of 'Improved' will be compared between treatment groups using logistic regression, adjusting for the stratification factors in accordance with the primary pooling strategy. The results of the analysis will be presented in terms of an odds ratio (with an odds ratio greater than 1 favouring olaparib+abiraterone combination therapy), together with the associated 95% profile likelihood CI and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model).

If there are not at least 5 responses across both treatment groups then a Fisher's exact test using mid p-values will be presented. The mid-p-value modification of the Fisher exact test amounts to subtracting half of the probability of the observed table from Fisher's p-value.

Fisher's exact test mid p-value = two-sided p-value – (table probability ÷ 2).

Time to deterioration in FACT-P (FACT-P Total score, FACT-G total score, TOI, FWB, PWB, PCS and FAPSI 6) will be analysed using the same methodology as in time to pain progression without any adjustments for multiplicity. The HRs and 95% CIs will be presented on a forest plot.

Summaries of the number and percentage of patients experiencing deterioration will be provided along with median time to deterioration for each treatment arm.

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

Supportive analyses

Change from baseline in the FACT-P total score, and scales (FACT-P total score, FACT-G total score, TOI, FWB, PWB, PCS, and FAPSI-6) will be analysed using a MMRM as described in Section 4.2.3.7.

4.2.3.10 HRR gene mutation status

HRR gene mutation status will be summarised descriptively as number of patients and corresponding percentages by treatment group using ctDNA-based test, tumour tissue test, and germline blood test separately.

To investigate the concordance between these three tests, a summary of HRR status (ctDNA-based test) vs. HRR status (tumour tissue test), HRR status (ctDNA-based test) vs. HRR status (germline blood test), HRR status (tumour tissue test) vs. HRR status (germline blood test) will be presented descriptively by number of patients and corresponding percentages.

These analyses will be performed at the time when the data become available.

[redacted]

[redacted]

[redacted]

[redacted]

4.2.5 Concordance between investigator and BICR assessments for rPFS (ITT)

Disagreements between investigator and BICR assessment of RECIST and bone (PCWG-3) progression of each treatment group will be presented for RECIST progression, PCWG-3 progression, and overall progression separately.

The number and percentage of patients in each category listed below will be presented:

- Progression declared by investigator and central review
- Progression declared by investigator but not central review
- Progression declared by central review but not investigator
- No progression by both central review and investigator

The summary will include the early discrepancy rate which is the frequency of investigator review progressions declared before the BICR (≥2 weeks earlier and including progressions declared by investigator but not BICR) as a proportion of all investigator review progressions, and the late discrepancy rate which is the frequency of investigator review progressions declared after the BICR (≥2 weeks later and including progressions declared by BICR but not investigator) as a proportion of all discrepancies (including early and late discrepancies) (Amit et al 2011).

4.2.6 Patient reported outcomes (PROs)

Summaries will include mean, standard deviation, median, minimum, maximum and range by treatment group for each visit until there are less than the minimum of 20 or one third of

patients with evaluable data. The absolute and change from baseline scores for each time point will be calculated by treatment group.

4.2.6.1 BPI-SF

Descriptive statistics will be calculated for BPI-SF Item 3, BPI-SF pain severity domain, BPI-SF pain inference domain at each scheduled visit/time point in the study, for each trial arm and as a total. The absolute and change from baseline scores for each time point will be calculated by treatment group. Graphical plots of the mean score, including change from baseline, and associated 95% CI by scheduled visits/time points in the study will be produced.

Change from baseline in BPI-SF Item 3 will be analysed using a MMRM as described in Section 4.2.3.7.

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

4.2.6.2 Analgesic use scoring

Descriptive statistics will be calculated for AQA score at each scheduled visit/time point in the study, for each trial arm and as a total. Graphical plots of the mean score, including change from baseline, and associated 95% CI by scheduled visits/time points in the study will be produced.

The number and percentage of patients in each AQA score will be summarised at baseline.

Missing data

Tabular summaries will be presented to show the percentage of randomised patients with one or more reconciled or unreconciled "Other" pain medication as well as all imputed OME values by assessment level for each visit for the full analysis set. Percentage of patients with imputed AQA values as applicable will be included in the summary tables.

The sensitivity and robustness of the imputation approaches will be assessed by producing a listing of all patients with one or more imputed OME values. The listing will include:

i) Highest OME value of pre-selected pain medications from Master List

ii) Highest OME value of reconciled "Other" pain medication

iii) Highest imputed OME value of unreconciled "Other" pain medication.

4.2.7 Safety

4.2.7.1 General considerations for safety assessments

Safety analyses will be presented using the Safety Analysis Set unless otherwise stated and will be investigated using descriptive statistics. Safety profiles will be assessed in terms of AEs, vital signs (including BP and pulse rate), laboratory data (clinical chemistry and hematology), and physical examination.

Baseline will generally be the last value obtained prior to the first dose of study medication. If more than one visit are equidistant, the average can be taken as a baseline value. For non-numeric laboratory tests (i.e. some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline as this is the most conservative. When there are multiple assessments on the same day, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline. In the scenario where tests are repeated multiple times with the same values but different normal ranges, one is normal and the other is not normal, the normal one can be taken.

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

- The time windows will be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the time window will be based on the actual date and not the intended date of the visit.

- All unscheduled visit data 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 assessment data (with 4 weeks between scheduled assessments) are:

  - Day 29, visit window 2 – 42
  - Day 57, visit window 43 – 70
  - Day 85, visit window 71 – 98

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).

- Listings should display all values contributing to a time point for a patient.

- For visit based summaries

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

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

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

Where safety data are summarised over time, study day will be calculated in relation to date of first treatment.

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e. below the lower limit of quantification) or > x (i.e. above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings. Additionally, adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

For laboratory data the following applies:

- Numerical summaries will provide the mean, standard deviation, median, minimum, maximum, and lower and upper quartile for visit based tabular summaries.

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

For missing AE start dates, the following will be applied:

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

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

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

For missing AE end dates, the following will be applied:

(a) Missing day - Impute the last day of the month unless month is the same as month of the last dose of study drug then impute last dose date.

(b) Missing day and month - impute 31st December unless year is the same as last dose date then impute last dose date.

(c) Completely missing date – do not impute.

The imputation of dates will be used to decide if an observation is treatment emergent for adverse events or concomitant medications. The imputed dates are not used to calculate durations. Where partial dates occur, listings will contain the date collected in the partial form.

4.2.7.2 Adverse events

All AEs, both in terms of MedDRA preferred term and CTCAE grade, will be listed and summarised descriptively by count (n) and percentage (%) and treatment group. MedDRA dictionary will be used for coding. Any AE occurring before olaparib/placebo/abiraterone (i.e., before Study Day 1) 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'.

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

- All AEs
- All AEs causally related to olaparib/placebo/abiraterone (see sub-types below)
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, causally related to olaparib/placebo (see sub-types below)
- AEs with outcome of death
- AEs with outcome of death causally related to olaparib/placebo/abiraterone
- All SAEs
- All SAEs causally related to olaparib/placebo/abiraterone
- AEs leading to discontinuation of olaparib/placebo/abiraterone

- AEs leading to discontinuation of olaparib/placebo/abiraterone, causally related to olaparib/placebo/abiraterone
- AEs leading to dose reduction of olaparib/placebo/abiraterone
- AEs leading to dose interruption of olaparib/placebo/abiraterone
- Other significant AEs
- Other significant AEs causally related to olaparib/placebo/abiraterone
- AE's for COVID-19 infections

For the corresponding rows of the above list "related to olaparib/placebo/abiraterone" will be broken down into the following sub-types:

- Related to study treatment
- Related to olaparib/placebo
- Related to olaparib/placebo only
- Related to abiraterone only
- Related to prednisone/prednisolone only

An overall summary of the number and percentage of patients in each category listed above will be presented, as will an overall summary of the number of episodes in each category. An overall summary of the number and percentage of patients in each category will also be produced for grouped AE terms of anaemia, neutropenia, thrombocytopenia, nausea, vomiting, fatigue and asthenia, cardiac and thromboembolic.

In addition, a truncated AE table of most common AEs, showing all events that occur in at least 5% of patients in any treatment group 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 (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 (NCI CTCAE version 4.03) 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.

Key patient information tables will be produced for:

- AEs with outcome of death
- All SAEs
- AEs leading to discontinuation of olaparib/placebo/abiraterone
- Other significant AEs
- AE's for COVID-19 infections

AEs which started prior to first dose or > 30 days following date of last dose will be listed only.

Listings of AE data will also be produced. All reported AEs will be included in listings along with the date of onset, date of resolution (if AE is resolved), investigator's assessment of severity and relationship to study drug.

Deaths

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

- Death related to disease under investigation only
- AE with outcome of death only
- AE with outcome of death only (AE start falling after 30-day follow up)
- Number of subjects with death related to disease and AE outcome of death
- Other deaths
- AEs with outcome of death
- Deaths related to disease under investigation and AE with outcome of death
- Deaths unrelated to AE or disease under investigation
- Deaths > 30 days after last treatment dose, related to disease under investigation
- AE with outcome of death with a start date > 30 days after last treatment dose
- Deaths > 30 days after last treatment dose, related to AE or disease under investigation
- Deaths > 30 days after last treatment dose, unrelated to AE or disease under investigation
- Patients with unknown reason for death
- Other deaths (not captured above)

Adverse events of special interest (AESI)

Preferred terms used to identify adverse events of special interest (as defined in Section 3.6.3) will be listed before database lock (DBL) and documented in the Trial Master File. Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the

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

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

- At least one AESI presented by outcome
- At least one AESI causally related to study medication
- At least one AESI leading to discontinuation of study medication

4.2.7.3 Laboratory assessments

Laboratory data (clinical chemistry and haematology) will be summarized. Shift tables will be provided for select tests, where shift from baseline to the worst value within the study will be summarized. Laboratory data outside the reference ranges will be indicated.

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 CTCAE grade will be produced, within each part of the study and overall, 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 study treatment). A scatter plot of alanine aminotransferase (ALT) versus total bilirubin, both expressed as multiples of upper limit of normal range, will be produced. The scatter plot will be repeated for aspartate aminotransferase (AST) versus total bilirubin.

4.2.7.4 Vital signs

Vital signs, including BP (mmHg), body temperature (°C), pulse (beats/minute) and weight (kg), will be summarized at baseline.

4.2.7.5 Exposure

Exposure data will be summarised, the following summaries will be produced:

- Summary of duration of exposure of study treatment, RDI and PID

- Summary of interruptions and reductions of study treatment

These data will also be listed.

4.2.7.6 Electrocardiogram

Overall ECG evaluation and the clinically significance of abnormal ECG finding will be summarised using descriptive statistics at each scheduled assessment time by actual treatment group.

4.2.8 Pharmacokinetic data

PK data from patients excluded from the PK analysis set or any individual data values to be excluded from the statistical analyses will be included in the data listings. Extra measurements (such as unscheduled or repeat assessments) will also be included in the patient listings but will not be included in summary tables. PK listings and individual patient concentration versus time plots will be presented for the randomized analysis set.

Plasma concentrations and derived PK parameters for olaparib, abiraterone and Δ4-abiraterone will be summarised separately by treatment group for the PK analysis set for Study Day 29. Geometric mean and combined individual plasma concentration versus time plots will also be presented for the PK analysis set.

A listing of PK blood sample collection times and all reportable concentrations will be presented for olaparib, abiraterone and Δ4-abiraterone for all patients.

Plasma concentrations at each nominal time point will be summarised separately for each analyte by treatment using the following summary statistics:

- Number of observations (n)
- n > lower limit of quantification (n > LLOQ)
- Geometric mean (gmean), calculated as exp[μ], where μ is the arithmetic mean of the data on a log scale)
- Geometric coefficient of variation (%gCV), calculated as $100 \sqrt{[\exp(s^2)-1]}$, where s is the standard deviation of the data on a log scale)
- Geometric mean + geometric standard deviation (gmean+gSD), calculated as exp[μ+s]
- Geometric mean + geometric standard deviation (gmean-gSD), calculated as exp[μ-s]
- Arithmetic mean (Mean)
- Arithmetic Standard Deviation (SDev)
- Arithmetic coefficient of variation (%CV)
- Median

- Minimum
- Maximum

Reporting of plasma concentrations that are Below Limit of Quantification (BLQ)

Individual olaparib, abiraterone and Δ4-abiraterone concentrations below their LLOQ of the bioanalytical assay will be reported as NQ (Not Quantifiable) in the listings with the LLOQ defined in the footnotes of the relevant TFLs. Individual plasma concentrations that are Not Reportable will be reported as NR and those that are missing will be reported as NS (No Sample) in the listings.

Plasma concentrations that are NQ, NR or NS will be handled as follows for the provision of descriptive statistics:

- Any values reported as NR or NS will be excluded from the summary tables and figures.
- At a time point where less than or equal to 50% of the concentration values are NQ, all NQ values will be substituted with the LLOQ concentration, and all descriptive statistics will be calculated accordingly.
- At a time point where more than half (but not all) of the values are NQ, the gmean, %gCV, gmean+gSD, gmean-gSD, mean, SDev and %CV will be set to Not Calculable (NC). The maximum value will be reported from the individual data, and the minimum and median will be set to NQ.
- If all concentrations are NQ at a time point, the gmean, mean, minimum, median and maximum will be reported as NQ, and the gmean+gSD, gmean-gSD, %gCV, SDev and %CV will be reported as NC.
- The number of values above LLOQ (n > LLOQ) will be reported for each time point together with the total number of collected values.

Three observations > LLOQ are required as a minimum for a plasma concentration to be summarised. Two values are presented as a minimum and maximum with the other summary statistics as NC.

Plasma concentrations that are NQ will be handled as follows for display in figures:

- For gmean concentration-time plots: NQ concentrations will be handled as described for the descriptive statistics. If this handling results in a geometric mean of "NQ", then the value plotted at that time-point will be zero for linear plots and set to missing for semi-logarithmic plots. Any gmean±gSD error bar values that are negative will be truncated at zero on linear concentration-time plots and omitted from semi-logarithmic plots.
- For individual plots and combined individual plots: NQ values prior to the first quantifiable concentration in that profile will be set to zero (linear plots only); after the first quantifiable concentration of the profile any NQ values will be set to missing.

All reportable PK parameters from the NCA will be listed for olaparib, abiraterone and Δ4-abiraterone for patients in the randomised analysis set.

Plasma PK parameters will be summarised separately for each analyte by treatment for the PK analysis set using the following descriptive statistics:

- $C_{max,ss}$, $C_{min,ss}$, $AUC_{(0-8)}$ and AUC_{ss} will present n, gmean, gmean+gSD, gmean-gSD, gCV(%), mean, SDev, CV%, median, min and max.
- $MRC_{max,ss}$, $MRC_{min,ss}$ and $MRAUC_{(0-8)}$ will present n, gmean, gCV(%), mean, SDev, CV%, median, min and max.
- CL_{ss}/F will present n, mean, SDev, CV%, median, min and max.
- $t_{max,ss}$ will present only n, median, min and max.
- t_{last} will be listed only and not summarized.

For the calculation of summary statistics of PK parameters, all NR and NC values will be set to missing. Three reportable values are required as a minimum for a PK parameter to be summarised. Two values are presented as a minimum and maximum with the other summary statistics as NC. For the PK parameters derived directly from the plasma concentration profiles ($C_{max,ss}$ and $C_{min,ss}$) any NQ values will be handled for the descriptive statistics calculations using the same rules as for the plasma concentration data. If one or more values for a given parameter is zero (or imputed with zero), then no geometric statistics will be calculated for that parameter and the results for geometric statistics will be set to "NA", not applicable.

Individual plasma concentrations versus actual elapsed time after dose for Study Day 29 will be plotted on both the linear and semi-logarithmic scale for. The abiraterone and Δ4-abiraterone data will be presented on the same plots and olaparib will be presented on a separate plot.

Combined individual plasma concentration versus actual elapsed times after dose for Study Day 29 will be plotted separately by analyte and treatment on both the linear and semi-logarithmic scale.

Gmean (±gSD) plasma concentration versus nominal sampling time for Study Day 29 will be plotted separately by analyte on both the linear and semi-logarithmic scale with all treatments on the same plot. The semi-logarithmic plot will also be presented without error bars for olaparib only.

Precision and Rounding Rules

PK concentration data will be presented in the listings to the same number of significant digits as the data received from the bioanalytical laboratory and against the same units as received. PK concentration descriptive statistics will all be presented to 4 significant figures, with the exception of the minimum and maximum which will be presented to 3 significant figures, and n and n>LLOQ which will be presented as integers.

For plasma PK parameters, the listings will be presented according to the following rules:

- $C_{max,ss}$, $C_{min,ss}$, AUCss, $AUC_{(0\text{-}8)}$, CL_{ss}/F, $MRAUC_{(0\text{-}8)}$, $MRC_{max,ss}$ and $MRC_{min,ss}$ will be presented to 3 significant figures.
- $t_{max,ss}$ and t_{last} – will be presented to 2 decimal places.

For PK parameter data the descriptive statistics will be presented according to the following rules:

- $C_{max,ss}$, $C_{min,ss}$, AUC_{ss}, $AUC_{(0\text{-}8)}$, CL_{ss}/F, $MRAUC_{(0\text{-}8)}$, $MRC_{max,ss}$ and $MRC_{min,ss}$ – all descriptive statistics will be presented to 4 significant figures with the exception of the minimum and maximum which will be presented to 3 significant figures and n which will be presented as integers.
- $t_{max,ss}$ – all descriptive statistics will be presented to 2 decimal places, with the exception of n which will be presented as an integer.

4.2.9 Concomitant medications

For the purpose of inclusion in prior and/or concomitant mediation or therapy summaries, incomplete medication or radiotherapy start and stop dates will be imputed as detailed in Section 4.2.7.1.

The following summaries will be produced for all patients in the FAS:

- Summary of concomitant medications
- Summary of disallowed medications

All concomitant and other treatment data will be listed.

Missing coding terms should be listed and summarised as "Not coded".

4.2.10 Demographics and baseline characteristics

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

- Patient disposition
- Important protocol deviations

- Inclusion in analysis sets
- Stratification factors
- Demographics (age, age group[< 65, ≥ 65], sex, race and ethnicity)
- Patient characteristics at baseline (weight)
- Patient recruitment by region, country and centre
- Previous disease-related treatment modalities
- Previous chemotherapy prior to this study
- Previous/current/post treatment radiotherapy
- Disease characteristics at baseline (primary tumour location, histology type, gleason score [grade 1, grade 2], TNM classification at baseline time from initial diagnosis in months, time from CRPC to randomization in months, time from mCRPC to randomization in months, prior local therapy with curative intent for prostate cancer, prior enzalutamide, abiraterone or both, patient's with taxane treatment at mCRPC, type of prostate cancer progression [PSA progression, radiological progression, both], ECOG performance status, baseline pain score [BPI-SF Item 3 score: 0-1, 2-3, >3], baseline PSA, haemoglobin, alkaline phosphatase, lactate dehydrogenase, abumin and creatinine)
- Extent of disease at baseline
- Time from most recent disease progression to randomisation
- Disease related medical history (past, past opioid use, and current)
- Relevant surgical history
- Post-discontinuation cancer therapy

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

5 INTERIM ANALYSES

[redacted]

The interim analysis of rPFS at DCO1 will initially be reviewed by an external independent data monitoring committee (IDMC). If and when the interim efficacy boundary is achieved, AstraZeneca will be contacted and a Unblinded Review Committee (URC) may be activated. More details can be found in the Unblinding Communication Plan and in Section 9.5 of the Clinical Study Protocol. The AZ study team will remain blinded during this initial review period, as described in Section 6.3.1 of the Clinical Study Protocol.

The IDMC will also review accumulating study safety data (see Appendix C 5). Committee members will include therapeutic area experts, a cardiologist, and a statistician who are not employed by AstraZeneca or by any participating study group and who do not have any major conflict of interest. An unblinded IDMC review of all myocardial infarction, congestive heart failure, and arterial thrombosis events will be conducted by the expert cardiologist. Following each review, the committee will recommend whether the study should continue unchanged, be terminated, or be modified in any way. Committee membership and responsibilities will be detailed in a committee charter.

The IDMC will separately assess the safety of the olaparib and abiraterone combination therapy in Japanese patients for the initial data review meetings. Further details will be provided in the IDMC charter.

6 CHANGES OF ANALYSIS FROM PROTOCOL

NA

7 REFERENCES

Amit et al 2011
Amit O, Mannino F, Stone Am et al. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis and recommendations from a PhRMA working group. European J Cancer 2011; 47:1772-1778.

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

Cella et al 1994
Cella, D.F. Manual: Functional Assessment of Cancer Therapy (FACT) Scales and Functional Assessment of HIV Infection (FAHI) Scale, Chicago: Rush-Presbyterian-St Luke's Medical Center. 1994.

Cella et al 2009
Cella D, Nichol MB, Eton D, Nelson JB, Mulani P. Estimating clinically meaningful changes for the Functional Assessment of Cancer Therapy-Prostate: results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. Value Health. 2009 Jan-Feb;12(1):124-9.

Cleeland 2009
Cleeland CS. The Brief Pain Inventory User Guide. 2009;1-63.

Chung et al 2014
Chung KC, Barlev A, Braun AH, Qian Y, Zagari M. Assessing analgesic use in patients with advanced cancer: development of a new scale--the Analgesic Quantification Algorithm. Pain Med 2014;15(2):225-32.

Cleeland and Ryan 1994
Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994 March;23(2):129-138.

Dworkin et al 2005
Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9-19.

Dworkin et al 2008
Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008 Feb;9(2):105-21.

Esper et al 1997
Esper P, Mo F, Chodak G, Sinner M, Cella DF, Pienta KJ. Measuring quality of life in men with prostate cancer using the Functional Assessment of Cancer Therapy-Prostate instrument. Urology 1997;50:920-8.

EuroQol Group 2013
EuroQol Group. EQ-5D-5L user guide: basic information on how to use the EQ-5D-5L instrument, version 2.0, October 2013. Available from: URL: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D5L_v2.0_October_2013.pdf. Accessed 21 November 2013.

Food and Drug Administration 2009
Food and Drug Administration: Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims; Availability. Federal Register 2009;74:65132-3.

Gail and Simon 1985.
Testing for qualitative interactions between treatment effects and subject 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.

Maurer and Bretz 2013
Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. Statistics in Biopharmaceutical Research 2013;5(4):311-20.

O'Brien and Fleming 1979
O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35:549-556.

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

Pocock 1977
Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977;64:191-9.

Robins et al 1991
Robins JM, Tsiatis AA. Correcting for noncompliance in randomised trials using rank preserving structural failure time models. Commun Stat Theory 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, ASA. 1993; 24-33.

Sun and Chen 2010
Sun X, Chen C. Comparison of Finkelstein's Method With the Conventional Approach for IntervalCensored Data Analysis. Stat Biopharm Res. 2010; 2:97-108

Turk et al 2006
Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. Pain 2005 Dec 5;125(3):208-15.

Van Hout et al. 2012
Van Hout B, Janssen MF, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012; 15(5):708-15

Webster et al 2003
Webster K, Cella D, Yost E. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Health and Quality of Life Outcomes. 2003;1(1):79.

8 APPENDIX (NOT APPLICABLE)

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d081sc00001-sap-ed-4		
Document Title:	Statistical Analysis Plan Edition 4	
Document ID:	Doc ID-003968970	
Version Label:	5.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
21-May-2021 13:26 UTC	[redacted]	Content Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.