ARCHES: A Multinational, Phase 3, Randomized, Double-blind, Placebo-controlled Efficacy and Safety Study of Enzalutamide Plus Androgen Deprivation Therapy (ADT) Versus Placebo Plus ADT in Patients with Metastatic Hormone Sensitive Prostate Cancer

ISN/Protocol 9785-CL-0335

ClinicalTrials.gov Identifier: NCT02677896

Date of SAP v3.0: 15 Nov 2018

Sponsor: Astellas Pharma Global Development, Inc.

1 Astellas Way

Northbrook, IL 60062

Co-Sponsor: Medivation, Inc. a wholly owned subsidiary of Pfizer Inc.

525 Market Street, 36th Floor

San Francisco, CA 94105

STATISTICAL ANALYSIS PLAN

Version 3.0, dated 15 November 2018

ARCHES: A Multinational, Phase 3, Randomized, Double-blind, Placebo-controlled Efficacy and Safety Study of Enzalutamide Plus Androgen Deprivation Therapy (ADT) Versus Placebo Plus ADT in Patients with Metastatic Hormone Sensitive Prostate Cancer

ISN: 9785-CL-0335
EudraCT number: 2015-003869-28
IND number: 74,563

Astellas Pharma Global Development, Inc
1 Astellas Way
Northbrook, IL 60062

Co-Sponsor:
Medivation, Inc., a wholly owned subsidiary of Pfizer Inc
525 Market Street, 36th Floor
San Francisco, CA 94105

Table of Contents

List of In-text Tables

List of In-text Figures

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
ADT	Androgen deprivation therapy
AE	Adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BPI-SF	Brief pain inventory - Short form
CRF	Case report form
CRPC	Castration-resistant prostate cancer
CT	Computed tomography
CTCAE	Common terminology criteria for adverse events
ECG	Electrocardiogram
ECOG	Eastern cooperative oncology group
eCRF	Electronic case report form
EQ-5D-5L	EuroQol group-5 dimension-5 level instrument
FACT-P	Functional assessment of cancer therapy - prostate
HR	Hazard ratio
ICF	Informed consent form
ICH	International conference on harmonization
ICR	Independent central review
IRT	Interactive response technology
ITT	Intent-to-treat
LHRH	Luteinizing hormone-releasing hormone
mHSPC	Metastatic hormone sensitive prostate cancer
M1	metastatic disease
MRI	Magnetic resonance imaging
NCI	National cancer institute
ORR	Objective response rate
OS	Overall survival
PSADecR	Rate of PSA decline to <2ng/mL
PSA	Prostate-specific antigen
QLQ-PR25	Quality of life prostate-specific questionnaire
QoL	Quality of life
RECIST	Response evaluation criteria in solid tumors
rPD	Radiographic disease progression (i.e., radiographic progression; radiographic progression of disease)
rPFS	Radiographic progression-free survival
SAP	Statistical analysis plan
SOC	System organ class
SMQ	Standardised MedDRA Queries
SSE	Symptomatic skeletal event
TTPP	Time to PSA progression
ULN	Upper limit of normal
WHO	World Health Organization

List of Key Terms

Terms	Definition of terms
Adverse Event	An adverse event is any untoward medical occurrence in a patient administered a study drug or has undergone study procedure and which does not necessarily have a causal relationship with this treatment.
Data analysis cut-off date	A cut-off date will be set so that a minimum of 262 events for the primary variable, radiographic progression-free survival, occurred by that date. All data available for all visits occurring prior to or on the cut-off date will be reported.
Baseline	Observed values/findings that are considered as the value at the starting point.
Enroll	To register or enter into a clinical trial. Note: once a patient has been enrolled, the clinical trial protocol applies to the patient.
Endpoint	A variable that pertains to the trial objectives
Intervention	The drug, therapy or process under investigation in a clinical study that is believed to have an effect on outcomes of interest in a study. (e.g., health-related quality of life, efficacy, safety, pharmacoeconomics).
Investigational period	Period of time where major interests of protocol objectives are observed, and where the test drug or comparative drug is usually given to a patient, and continues until the last assessment after completing administration of the test drug or comparative drug.
Post investigational period	Period of time after the last assessment of the protocol. Follow-up observations for sustained adverse events and/or survival are done in this period.
Randomization	The process of assigning trial patients to treatment or control groups using an element of chance to determine assignments in order to reduce bias.
Screen failure	Potential patient who did not meet 1 or more criteria required for participation in a trial.
Screening	A process of active consideration of potential patients for enrollment in a trial.
Screening period	Period of time before entering the investigational period, usually from the time of starting a patient signing consent until just before the test drug or comparative drug is allocated to a patient (i.e. randomization).
Study period	Period of time from the first site initiation date to the last site completing the study.
Variable	Any quantity that varies; any attribute, phenomenon or event that can have different qualitative or quantitative values.

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary endpoints and other data.

The SAP is finalized and signed prior to database hard lock to ensure lack of bias. If needed, revisions to the approved SAP may be made prior to database hard lock. Revisions will be version controlled.

This statistical analysis is coordinated by the responsible biostatistician of Astellas Pharma Global Development. Any changes from the analyses planned in the SAP will be justified in the Clinical Study Report.

Prior to database hard lock, a final blinded review of data and TLFs meeting will be held to allow a review of the clinical trial data and to verify the data that will be used for analysis set classification. If required, consequences for the statistical analysis will be discussed and documented. A meeting to determine analysis set classifications may also be held prior to database hard lock.

2 FLOW CHART AND VISIT SCHEDULE

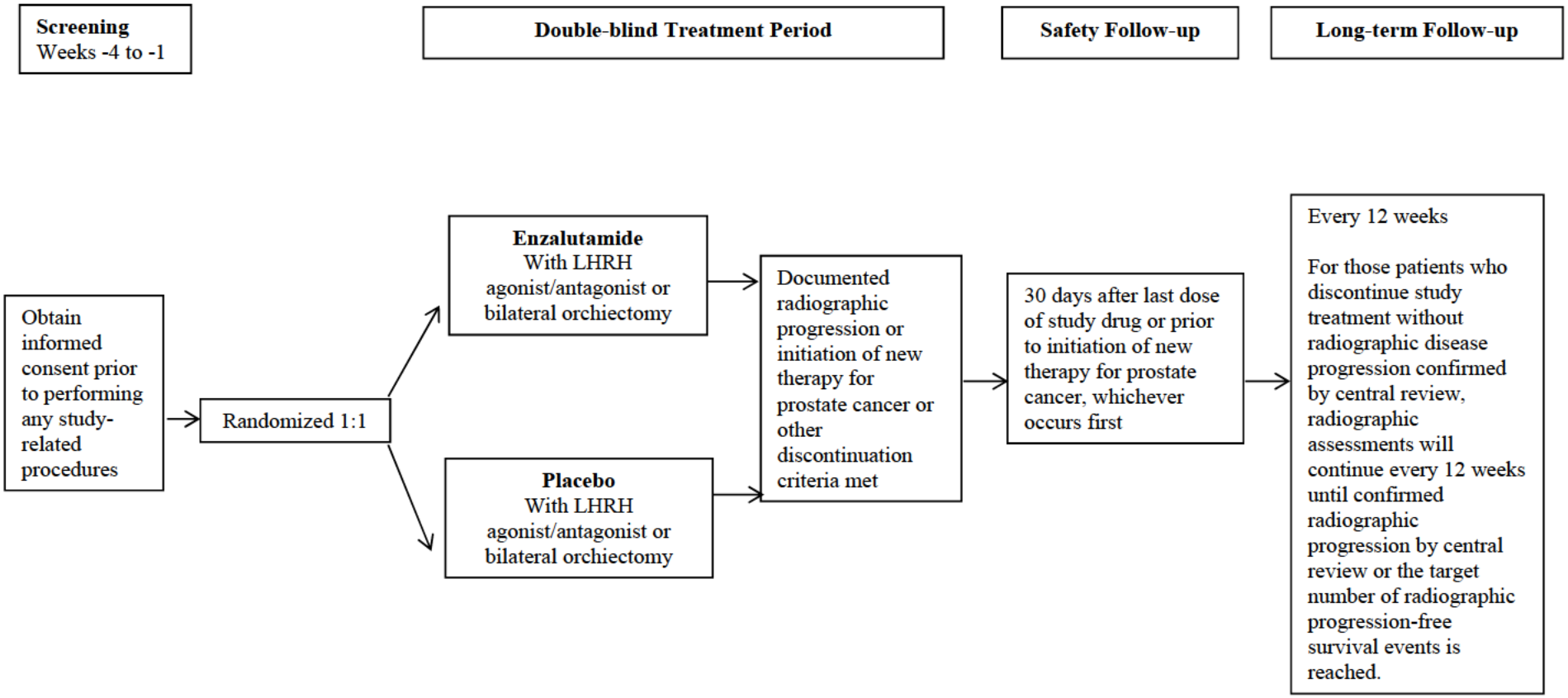


Table 1 Schedule of Assessments

Study Day	**Screening Visit**	**1**	**29**	**85 and Every Subsequent 84 Days**	**Safety Follow-up**	**Unscheduled Visit†**	**Long Term Follow-up‡**
Study Week	**-4 to -1 (28 Days)**	**1**	**5**	**13 and Every Subsequent 12 Weeks**	**30 Days after Last Dose§**		**Every 12 Weeks**
Window (Days)			± 5	± 5	± 7	NA	± 7
Informed Consent	X						
Medical History	X						
Inclusion/Exclusion Criteria	X	X					
Randomization (IRT)		X					
Vital Signs	X	X	X	X	X	X	
Physical Examination including Weight¶	X	X	X	X	X	X	
Height	X						
12-lead Electrocardiogram	X	X			X		
Clinical Labs††	X	X	X	X	X	X	
Prostate-specific antigen	X	X	X	X	X		
Sample for Genotyping Analysis‡‡		X					
[redacted]		X		X			
Testosterone				X			
CT/MRI and Bone Scan§§, ¶¶	X§§			X¶¶			X
Chest X-ray or Chest CT/MRI†††	X			X			X
ECOG Performance Status	X	X	X	X	X	X	
QoL Assessment (QLQ-PR25, EQ-5D-5L, FACT-P, BPI-SF)		X		X	X		X‡
Adverse Events§§§	X	X	X	X	X	X	
Previous and Concomitant Medications	X	X	X	X	X	X	
Study Drug Dispensing		X	X	X			
Study Drug Treatment		X	X	X			

CT: computed tomography; ECOG: Eastern Cooperative Oncology Group; EQ-5D-5L: EuroQol Group-5 Dimension-5 Level Instrument; IRT: Interactive Response Technology; MRI: magnetic resonance imaging; NA: not applicable; FACT-P: Functional Assessment of Cancer Therapy-Prostate; QLQ-PR25: Quality of Life Prostate-specific Questionnaire; QoL: quality of life; BPI-SF: Brief Pain Inventory-Short Form

Footnotes continued on next page

† Unscheduled visits may be performed at any time during the study whenever necessary to assess for or follow-up on adverse events at the patient's request or if deemed necessary by the Investigator. Procedures and assessments are to be performed as clinically indicated. Testosterone testing through central laboratory at unscheduled visit requires prior approval by sponsor.

‡ After study drug discontinuation, all patients MUST undergo long term follow-up. Long term follow-up assessments will include monitoring for survival status, new antineoplastic therapies for prostate cancer, and symptomatic skeletal events. Follow-up may be conducted by telephone interview. Patients will continue to be scanned every 12 weeks until radiographic progression is confirmed by independent review or the number of radiographic progression-free survival events is reached. For patients continuing with radiographic assessments, if seen in clinic, QoL assessment will also be completed until the initiation of new antineoplastic therapy for prostate cancer or the number of progression events is reached. Additional follow-up contacts may be requested. Patients will be followed for overall survival until death, lost to follow up, overall survival final analysis or termination of the study by the sponsor.

§ Or prior to initiation of new antineoplastic therapy for prostate cancer, whichever occurs first.

¶ A brief physical examination is required at each visit, with the exception of the screening visit during which a complete physical examination will be completed.

†† Laboratory assessments include serum chemistries and hematology.

‡‡ Genotyping samples will only be collected from patients who agree to provide genotyping samples as documented by signing a separate genotyping informed consent form(ICF).

∫∫ [redacted]

§§ The abdominal-pelvic CT scan or MRI, bone scan, chest x-ray or chest CT must occur within 6 weeks of day 1; otherwise the screening visit assessment must be repeated. Radiographic assessments performed prior to informed consent, as part of the routine care, may be used as the baseline assessment if performed within 6 weeks of day 1 and if digital format images are available for submission to the sponsor-designated independent central review facility.

¶¶ The window for all radiographic (CT/MRI) assessments is ± 7 days. For patients who discontinue study treatment without radiographic progression, confirmed by central review, patients will continue to be scanned every 12 weeks until radiographic progression is confirmed by independent review or the number of radiographic progression-free survival events is reached.

††† Chest CT is required at all imaging time points if screening chest x-ray demonstrates metastatic chest disease.

§§§ Adverse events will be collected from the time the patient signs the ICF until the end of the safety reporting period (or until screen failure). The safety reporting period ends at the time of the safety follow-up visit, 30 days after last dose of study drug or initiation of new antineoplastic therapy for prostate cancer.

3 STUDY OBJECTIVE(S) AND DESIGN

3.1 Study Objective(s)

The objective of this phase 3 study is to evaluate the efficacy and safety of enzalutamide plus Androgen Deprivation Therapy (ADT) versus placebo plus ADT in patients with metastatic hormone sensitive prostate cancer (mHSPC).

3.1.1 Primary Objective

- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by radiographic progression-free survival (rPFS) based on Independent Central Review (ICR)

3.1.2 Secondary Objectives

- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by overall survival (OS)
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by time to first Symptomatic Skeletal Event (SSE)
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by time to castration resistance
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by Quality of Life (QoL) (as measured by QoL Prostate-specific Questionnaire [QLQ-PR25] / Functional Assessment of Cancer Therapy-Prostate [FACT-P] and EuroQol Group-5 Dimension-5 Level Instrument [EQ-5D-5L]), in particular by:
    - time to deterioration in urinary symptoms using a modified urinary symptoms scale from QLQ-PR25
    - time to deterioration in QoL using the FACT-P global score
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by time to start of new antineoplastic therapy
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by time to PSA progression
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by PSA undetectable rate (< 0.2 ng/mL)
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by objective response rate (ORR)
- To determine the benefit of enzalutamide plus ADT as compared to placebo plus ADT as assessed by worsening of pain (using the Brief Pain Inventory-Short Form [BPI-SF])

3.1.3 Safety Objectives

- To determine the safety of enzalutamide plus ADT as compared to placebo plus ADT

3.1.4 Exploratory Objective ([redacted])

- [redacted]

3.2 Study Design

This is a multinational phase 3, randomized, double-blind, placebo-controlled efficacy and safety study of enzalutamide plus ADT versus placebo plus ADT in patients with mHSPC.

Approximately 1100 subjects will be randomized centrally 1:1, and the randomization will be stratified by volume of disease (low versus high) and prior docetaxel therapy for prostate cancer (no prior docetaxel, 1-5 cycles, 6 cycles). High volume of disease is defined as metastases involving the viscera or, in the absence of visceral lesions, there must be 4 or more bone lesions, at least 1 of which must be in a bony structure beyond the vertebral column and pelvic bone. Prior docetaxel therapy is defined as 1 or more cycles of docetaxel but no more than 6 cycles.

In this study, patients received 4 capsules of enzalutamide (40 mg each) or placebo per day orally. As long as the patient is tolerating the study drug, the treatment should be continued until radiographic disease progression (rPD) is documented as outlined in the Table 2 below or starting an investigational agent or new therapy for treatment of prostate cancer. It is recommended that patients remain on study drug until radiographic progression is confirmed by ICR. Study films (Computed tomography [CT]/ Magnetic resonance imaging [MRI] and bone scan) are read on site and also by ICR. Patients who discontinue study drug without radiographic progression will continue to follow the radiographic assessment schedule until radiographic progression event is confirmed by ICR or until the target number of progression events is reached as assessed by ICR.

The following assessments of prostate cancer status will be collected during the course of the study: PSA, soft tissue disease on CT scan or on MRI, bone disease on radionuclide bone scans, survival status, EQ-5D-5L, QLQ-PR25, FACT-P for QoL and BPI-SF for pain symptom assessment. Throughout the study, safety and tolerability were assessed by the recording of adverse events (AE), vital signs, physical examinations, 12-lead electrocardiograms (ECG), and safety laboratory evaluations.

An Independent Data Monitoring Committee (also referred as Data Safety Monitoring Board in the protocol) will monitor the unblinded safety data on an ongoing basis.

Patients will have a safety follow-up visit 30 days after their last dose of study drug or prior to initiation of new antineoplastic therapy for prostate cancer, whichever occurs first. All patients are to be followed for survival until death, loss to follow-up, withdrawal of consent, or study termination by the sponsor. All patients will be followed until the final OS analysis.

The sponsor will monitor study enrollment for proportion of subjects enrolled with a history of prior docetaxel treatment, and may either change the sample size, or cap the number of

subjects who received prior docetaxel to ensure that the primary endpoint is not driven either by the subjects who received prior docetaxel, or by the subjects who did not receive it.

3.3 Randomization

Subjects will be entered into the Interactive Response Technology system (IRT) at screening and assigned a subject number. Treatment will then be randomly assigned in a 1:1 ratio to enzalutamide 160 mg/day or placebo in the IRT system.

Subjects will be stratified by prior docetaxel use (none, 1-5 cycles, and 6 cycles) and disease volume (low versus high). High-volume disease is defined as metastases involving the viscera or, in the absence of visceral lesions; there must be 4 or more bone lesions, at least 1 of which must be in a bony structure beyond the vertebral column and pelvic bone.

4 SAMPLE SIZE

Approximately 1100 subjects (550 subjects per treatment arm) will be randomized in the study. The final analysis of rPFS will be conducted when a minimum of 262 progression events have occurred, based on the following considerations:

- A target hazard ratio (HR) is 0.67. The expected median rPFS for the ADT arm is 20 months as measured from the date of randomization. Under the assumption of an exponential distribution, a target HR of 0.67 corresponds to approximately 50% increase in median rPFS for the enzalutamide plus ADT arm relative to the placebo plus ADT arm (approximately 30 versus 20 months).
- 262 rPFS events (radiographic progression at any time or death from any cause within 24 weeks after study drug discontinuation, whichever occurs first) provides 90% power to detect the target HR based on a 2-sided log-rank test and significance level of 0.05.

Additionally, the study is powered for OS. Specifically, 342 death events will be required to provide 80% power to detect a target HR of 0.73 with a target difference in Kaplan-Meier estimated median of approximately 15 months (40 months for placebo versus 55 months for enzalutamide) at the 4% significance level under the assumption of an exponential distribution. This significance level was chosen to apply a parallel testing strategy between OS and some other secondary endpoints (with allocated type I error rate of 1%) as described in Section 7.4.2.

5 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the following analysis sets will be used for the analyses.

Detailed criteria for analysis sets will be laid out in Classification Specifications (CS) and the allocation of subjects to analysis sets will be determined prior to database hard lock.

5.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population is defined as all subjects who were randomized in this study. The ITT population will be analyzed by treatment arm as randomized, (i.e., treatment arm by randomization assignment) regardless of whether or not study drug was administered. The ITT population will be used to conduct efficacy analyses, unless otherwise specified.

5.2 Safety Population

The safety (SAF) population is defined as all randomized subjects who received at least 1 dose of study drug. The safety population will be used to conduct safety analyses by treatment arm as treated (i.e., based on the actual study drug the subject mostly received, rather than the study drug to which the subject was randomized to).

6 ANALYSIS VARIABLES

6.1 Efficacy Endpoints

6.1.1 Primary Efficacy Endpoint(s)

6.1.1.1 Primary analysis

The primary endpoint is radiographic progression-free survival (rPFS).

Table 2 Protocol-specified Documentation for Radiographic Evidence of Disease Progression

Date Progression Detected (Visit)†	Criteria for Progression	Criteria for Confirmation of Progression (Requirement and Timing)	Criteria for Documentation of Disease Progression on Confirmatory Scan
Week 13	Bone lesions: ≥ 2 new lesions compared to **baseline** bone scan	Timing: ≥ 6 weeks after progression identified or at week 25 visit	≥ 2 new bone lesions on bone scan compared to week 13 scan (≥ 4 new lesions compared to baseline bone scan)
	Soft tissue lesions: progressive disease on CT or MRI by RECIST v1.1	No confirmatory scan required for soft tissue disease progression	Not applicable
Week 25 or Later	Bone lesions: ≥ 2 new lesions on bone scan compared to best response on treatment (i.e. smallest number bone lesions on bone scan during treatment period)	No confirmatory scan required	Not applicable
	Soft tissue lesions: progressive disease on CT or MRI by RECIST v1.1	No confirmatory scan required for soft tissue disease progression	Not applicable

Progression detected by bone scan at an unscheduled visit prior to week 25 will require the same criteria for documentation of disease progression as week 13 with a confirmatory scan at least 6 weeks later or at the next scheduled scan.

In rPFS, an rPFS event is defined as objective evidence of rPD as assessed by ICR or death, as follows:

- Death from any cause within 24 weeks from study drug discontinuation.
- rPD at any time defined by Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 for soft tissue disease or the appearance of 2 or more new bone lesions on bone scan. The documentation and confirmation required for the determination of rPD are listed in Table 2. The date of rPD is the date of the first objective evidence of rPD is documented. Unconfirmed disease progression on bone scan at week 13 will not be considered as an event.

In patients with an rPFS event, rPFS will be calculated as the time interval from the date of randomization to the date of first objective evidence of rPD at any time or death from any cause within 24 weeks from study drug discontinuation, whichever occurs first.

In patients with no rPFS event, rPFS will be censored on the date of last radiographic assessment prior to the data analysis cut-off date. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as “Not Evaluable”, the radiographic progression free survival will be censored on the date of randomization. No other criteria for censoring are used in the primary analysis.

6.1.1.2 Sensitivity Analyses of Primary Endpoint

Appendix 1 - Summary of Sensitivity Analyses of Primary Endpoint summarizes the different rPFS definitions used in sensitivity analyses of the primary endpoint.

Unless otherwise stated, rPD by ICR is defined as described in Section 6.1.1.1

Sensitivity analysis 1 - rPFS_1 - Impact of study drug discontinuation as an additional event

rPFS_1 events are defined as rPD by ICR or death from any cause within 24 weeks from study drug discontinuation, whichever occurs first, or study drug discontinuation in the absence of rPD or death.

In patients with a rPFS_1 event, rPFS_1 will be calculated as the time interval from the date of randomization to the first date of rPD or death, whichever occurs first. In patients with no documented rPD, and death is not recorded, and whom discontinued treatment, rPFS_1 will be calculated as the time interval from the date of randomization to the date of study drug discontinuation.

In patients with no rPFS_1 event, rPFS_1 will be censored on the date of last radiographic assessment prior to the data analysis cut-off date. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as “Not Evaluable”, rPFS_1 will be censored on the date of randomization.

Sensitivity analysis 2 - rPFS_2 - Impact of new antineoplastic therapy and occurrence of a SSE as additional events

rPFS_2 events are defined as rPD by ICR, the occurrence of a SSE (definition in Section 6.1.2.2), the initiation of a new antineoplastic therapy (definition in Section 6.1.2.5), or deaths from any cause within 24 weeks from study drug discontinuation.

In patients with a rPFS_2 event, rPFS_2 is calculated as the time interval from the date of randomization to the date of first rPFS_2 event.

In patients with no rPFS_2 event, rPFS_2 is censored on the last disease assessment date prior to the data cut-off date. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as "Not Evaluable", rPFS_2 will be censored on the date of randomization.

Sensitivity analysis 3 - rPFS_3 - Impact of all deaths (with no time limit) as events

rPFS_3 events are defined as rPD by ICR, or deaths from any cause (at any time).

In patients with a rPFS_3 event, rPFS_3 is calculated as the time interval from the date of randomization to the first date of rPD or death, whichever occurs first.

In patients with no rPFS_3 event, rPFS_3 will be censored on the date of last radiographic assessment prior to the data analysis cut-off date. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as "Not Evaluable", rPFS_3 will be censored on the date of randomization.

Sensitivity analysis 4 - rPFS_4 - Impact of rPD documented between per protocol visits

rPFS_4 events are defined as rPD by ICR, or death from any cause within 24 weeks from study drug discontinuation.

In patients with a rPFS_4 event, rPFS_4 will be calculated as the time interval from the date of randomization to the first date of rPD or death (from any cause within 24 weeks from study drug discontinuation), whichever occurs first. If the rPD date was not documented as per scheduled protocol visit date (+/- 7 days per protocol allowed window), rPFS_4 will be calculated as the time interval from the date of randomization to the date of the first per protocol scheduled visit date after the rPD date, or to the date of death, whichever occurs first.

In patients with no rPFS_4 event, rPFS_4 will be censored on the date of last radiographic assessment prior to the data analysis cut-off date. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as "Not Evaluable", rPFS_4 will be censored on the date of randomization.

Sensitivity analysis 5 - rPFS_5 - 'Missing' data impact - Last scan not documented as NE

rPFS_5 events are defined as rPD by ICR, or death from any cause within 24 weeks from study drug discontinuation.

In patients with a rPFS_5 event, rPFS_5 will be calculated as the time interval from the date of randomization to the first date of rPD or death, whichever occurs first.

In patients with no rPFS_5 event, rPFS_5 will be censored on the date of the last radiographic assessment which is not documented as 'Not Evaluable' prior to the data cut-off date. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as 'Not Evaluable', rPFS_5 will be censored on the date of randomization.

Sensitivity analysis 6 - rPFS_6 - Missing' data impact - Absence of 2 consecutive scans

rPFS_6 events are defined as rPD by ICR which is documented before any period of 2 missing consecutive scheduled visit scans, or death from any cause within 24 weeks from study drug discontinuation.

In patients with a rPFS_6 event, rPFS_6 will be calculated as the time interval from the date of randomization to the first date of rPD or death, whichever occurs first.

In patients with no rPFS_6 event, rPFS_6 will be censored on the date of the last radiographic assessment prior to the data cut-off date and prior to any periods with 2 missing consecutive scheduled visit scans (if applicable). In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as "Not Evaluable", rPFS_6 will be censored on the date of randomization.

Sensitivity analysis 7 - rPFS_7 - Censoring rPD on competing risks**:** new antineoplastic therapy and occurrence of a SSE

rPFS_7 events are defined as rPD by ICR prior to the start of a new antineoplastic therapy and prior to the occurrence of a SSE, or death from any cause within 24 weeks from study drug discontinuation.

In patients with a rPFS_7 event, rPFS_7 will be calculated as the time interval from the date of randomization to the first date of rPD (i.e. documented prior to the start of a new antineoplastic therapy and prior to the occurrence of a SSE) or death, whichever occurs first.

In patients with no rPFS_7 event, rPFS_7 will be censored on the date of the last radiographic assessment prior to the data cut-off date, and is prior to the start of a new antineoplastic therapy and prior to the occurrence of a SSE, where applicable. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as "Not Evaluable", rPFS_7 will be censored on the date of randomization.

Sensitivity analysis 8 - rPFS_8 - 'Missing' data impact and censoring rPD on competing risks: new antineoplastic therapy, occurrence of a SSE, and study drug discontinuation in M1 patients based on ICR assessments

In M1 patients, whom are identified from the baseline assessments made by ICR, rPFS_8 events are defined as death from any cause within 24 weeks from study drug discontinuation, or rPD by ICR which is documented prior to any period of 2 missing consecutive scheduled visit scans, prior to the start of a new antineoplastic therapy, prior to the occurrence of a SSE and not after the date of study drug discontinuation.

In M1 patients with rPFS_8 event, rPFS_8 will be calculated as the time interval from the date of randomization to the first date of rPD (i.e., documented prior to any period of 2 missing consecutive scheduled visit scans, prior to the start of a new antineoplastic therapy, prior to the occurrence of a SSE and not after the date of study drug discontinuation) or death, whichever occurs first.

In M1 patients with no rPFS_8 event, rPFS_8 will be censored on the date of the last radiographic assessment other than "Not Evaluable" and is prior to the data cut-off date, prior to any periods with 2 missing consecutive scheduled visit scans, prior to the start of a new antineoplastic therapy, prior to the occurrence of a SSE, and not after the date of study drug discontinuation. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments and patients with all post-baseline radiographic assessments documented as "Not Evaluable", rPFS_8 will be censored on the date of randomization.

Sensitivity analysis 9 – rPFS_9 – rPFS in M1 patients

rPFS_9 is defined as rPFS restricted to M1 patients, whom are identified from the baseline assessments made by ICR.

Sensitivity analysis 10 - rPFS_10 - Impact of rPD documented by investigators

rPFS by investigators, rPFS_10, is similarly defined to the primary definition of rPFS, but using the radiographic assessments documented by investigators rather than by ICR.

Sensitivity analysis 11 – rPFS_11 – Impact of rPD according to PCWG2 criteria and documented by investigators

The rPFS according to PCWG2 criteria, rPFS_11 event is defined as objective evidence of rPD as assessed by investigators or death, as follows:

- Death from any cause within 24 weeks from study drug discontinuation.
- rPD by PCWG2 criteria. More specifically, the documentation and confirmation required for the determination of rPD on bone lesions are described in Table 2 (Section 6.1.1.1), except that the appearance of ≥2 new bone lesions is to be compared to the week 13 bone scan for week 25 or later visits.

Sensitivity analysis 12 – rPFS_12 – Impact of rPD according to PCWG2 criteria and documented by ICR

The rPFS according to PCWG2 criteria, rPFS_12 event is defined as objective evidence of rPD as assessed by ICR or death, as follows:

- Death from any cause within 24 weeks from study drug discontinuation.
- rPD by PCWG2 criteria. More specifically, the documentation and confirmation required for the determination of rPD on bone lesions are described in Table 2 (Section 6.1.1.1), except that the appearance of ≥2 new bone lesions is to be compared to the week 13 bone scan for week 25 or later visits.

6.1.2 Secondary Efficacy Endpoints

For variables derived, unless otherwise specified, only results taken before the data analysis cut-off date will be considered and are referred to in this section.

6.1.2.1 Overall Survival

OS is defined as the time from randomization to death from any cause. All events of death will be included.

For patients who are alive at the time of the cut-off date, OS time will be censored on the last date the patient is known to be alive or the cut-off date, whichever occurs first. The date the patient is last known alive by the cut-off date will be derived as follows:

- for patients on treatment by the cut-off date (i.e., patients who did not discontinue from treatment), the date last known alive will be the cut-off date.
- for patients who withdraw consent by the cut-off date, the date last known alive is the date of consent withdrawal.
- for patients lost to follow up, the date last known alive depends on the period the patient was lost to follow-up. It will be as follows:

Period lost to follow-up	**The date last known alive**
Since randomization (i.e., without further post-randomization visit)	The date of randomization
During treatment period	The last assessment/visit date or the date of the last dose of study drug, whichever occurs later
Safety follow-up period	The date of the last dose of study drug or the last treatment visit date collected on the End of Treatment Case Report Form (CRF) page, whichever occurs last.
Long-term follow-up period	The last visit/contact date (collected either on the 30 Day Follow-Up Status CRF page, or the Patient Status - Survival CRF page or the Long-Term Follow-Up Status CRF page).

- for other patients (i.e., not on treatment, not withdrawn and not lost to follow-up) who are alive by the cut-off date, the date last known alive is the date of last visit (collected either on the Patient Status - Survival CRF page or the Long-Term Follow-Up Status CRF page) or the date of randomization, whichever occurs last.

6.1.2.2 Time to First Symptomatic Skeletal Event

SSE is defined as radiation to bone, surgery to bone, clinically apparent pathological bone fracture and spinal cord compression.

In patients with SSE, the time to first SSE is defined as the time from randomization to the occurrence of the first SSE prior to the data analysis cut-off date.

In patients with no SSE by the time of the data analysis cut-off date, the time to first SSE will be censored on the last visit date or the date of randomization, whichever occurs last.

6.1.2.3 Time to Castration Resistance

A castration resistance event is defined as the occurrence of rPD by ICR (as defined in Section 6.1.1.1), PSA progression (as defined in Section 6.1.2.6), or SSE, whichever occurs first with castrate levels of testosterone (< 50 ng/dL). As testosterone was not reported at baseline, testosterone is considered <50 ng/dL up to the first post baseline measurement. The latest testosterone value measured prior to or at the date of radiographic disease progression by ICR, PSA progression or SSE, is used to determine if this event is a castration resistance event.

In patients with castration resistance event, time to castration resistance is defined as the time from randomization to the first castration-resistant event.

In patients with no documented castration resistance event, the time to castration resistance will be censored on the latest date from: the date of last radiographic assessment, the last PSA sample taken prior to the start of any new antineoplastic therapy and prior to 2 or more consecutive missed PSA assessments (if applicable), and the last visit date performed. In those patients, patients with no baseline radiographic assessment, patients with no post baseline radiographic assessments, patients with all post-baseline radiographic assessments documented as “Not Evaluable”, patients with no baseline PSA, and in patients with no post-baseline PSA results, the time to castration resistance will be censored on the date of randomization.

6.1.2.4 Quality of Life

Additional Quality of Life analyses with further details about questionnaires and associated derived variables can be found in a separate SAP specific to patient reported outcomes.

QLQ-PR25

The EORTC QLQ-PR25 is a 25-item module designed to assess QoL in prostate cancer patients.

Each item score ranges from 1 (not at all) to 4 (very much).

It includes multi-item scales and single-item scales assessing urinary symptoms, bowel symptoms, and hormonal treatment-related symptoms, use of incontinence aids, and sexual activity and sexual functioning, as follows:

- Urinary symptoms (8 items: Q31 – Q37, Q39)

- Incontinence aids (Q38)
- Bowel symptoms/function (4 items: Q40 – Q43)
- Hormonal treatment-related symptoms (6 items: Q44 – Q49)
- Sexual activity (2 items: Q50, Q51)
- Sexual functioning (4 items: Q52 – Q55)

In addition, a modified scale of urinary symptoms will be derived from a selected subset of symptoms from the 'urinary symptoms' scale to define a secondary efficacy endpoint related to QLQ-PR25:

- Modified urinary symptoms (3 items: Q31 – Q33)

Higher scores represent higher functioning for the two sexual domains but, conversely, higher scores represent more symptoms (i.e., worse QoL) for the symptom scales.

In order to score any scale, first the raw score (RS) is computed by averaging the raw values of the individual items that contribute to the scale and then the RS is linearly transformed using the range of item raw values so that scores range from 0 to 100. Specifically, the scores of these scales and the individual items will be calculated based on {(RS-1)/range}×100. However, for Q53 to Q55, before scoring the raw value of an individual item will be subtracted from 5 (i.e., 5 – raw value). For handling missing items, if half or more questions within scale are answered then a score will be calculated for that scale. Otherwise the patient score for that scale will be missing.

Time to Deterioration in Urinary Symptoms

The QLQ PR-25 secondary efficacy endpoint will be the time to deterioration in urinary symptoms.

A deterioration in urinary symptoms is defined as an increase in the modified urinary symptoms scale score by >= 50% of the standard deviation observed in the modified urinary symptoms scale score at baseline (i.e., 1/2SD at baseline over the pooled arms, no rounding).

In patients with a deterioration in urinary symptoms, the time to deterioration in urinary symptoms is defined as the time interval between randomization and the first deterioration in urinary symptoms at any post baseline visit.

In patients without a deterioration in urinary symptoms, the time to deterioration in urinary symptoms will be censored on the date of the last urinary symptom score is calculable. Patients with no baseline urinary symptoms score and patients with no post baseline urinary symptoms score, the time to deterioration in urinary symptoms will be censored on the date of randomization.

In addition, the time to confirmed deterioration in urinary symptoms will be defined as the time interval to the first deterioration in urinary symptoms which is confirmed by a second consecutive assessment of the deterioration.

Additional sensitivity analyses around the threshold and censoring rules will be described in the separate SAP specific to patient reported outcomes.

FACT-P

The FACT-P questionnaire is a multi-dimensional, self-reported QoL instrument specifically designed for use with prostate cancer patients.

It is composed of 27 core items which assess patient function in four domains and 12 prostate cancer-related items, as follows:

- Physical well-being (PWB): 7 items;
- Social/family well-being (SWB): 7 items;
- Emotional well-being (EWB): 6 items;
- Functional well-being (FWB): 7 items;
- Prostate cancer subscale (PCS): 12 items.

The Table 3 lists each FACT-P item and its appropriate scoring. Each item is rated on a 0 to 4 Likert-type scale as: 0=not at all; 1= a little bit; 2= somewhat; 3= quite a bit; 4= very much. For some items a response of "4= very much" is better than a response of "3= quite a bit" (e.g., "I get support from my friends"), while for other items a response of "4= very much" is worse than a response of "3= quite a bit" (e.g. "I have pain").

Before calculating the subscale and global scores, the items for which "4" is worse than "3" must be reversed, by subtracting the response from 4. The reversals are performed in the following:

- PWB: reverse all items (GP1 - GP7);
- SWB: do not reverse any items;
- EWB: reverse items GE1 and GE3 - GE6;
- FWB: do not reverse any items;
- PCS: reverse items C2, P1 - P3, P6 - P8 and BL2.

After reversing proper items, for all FACT-P scales, the higher the score the better the QoL. Each subscale score is the sum of the scores for the items in the subscale. If there are missing items, subscale scores can be prorated, as long as more than 50% of the items are answered in any given subscale (e.g., a minimum of 4 of 7 items, 4 of 6 items, etc). The score is prorated as follows:

> Prorated subscale score = (sum of item scores) * (number items in the subscale)/(number of items answered).

The FACT-P total score is the sum of all 5 subscale scores. The total score will be calculated only if the overall item response rate is greater than 80% (i.e., a minimum of 32 of 39 items currently scored in the FACT-P have been answered), and no subscale scores are missing.

Table 3 Scoring of FACT-P items

		Scoring				
	Item number	**Not at all**	**A little bit**	**Somewhat**	**Quite a bite**	**Very much**
PWB	GP1	4	3	2	1	0
	GP2	4	3	2	1	0
	GP3	4	3	2	1	0
	GP4	4	3	2	1	0
	GP5	4	3	2	1	0
	GP6	4	3	2	1	0
	GP7	4	3	2	1	0
SWB	GS1	0	1	2	3	4
	GS2	0	1	2	3	4
	GS3	0	1	2	3	4
	GS4	0	1	2	3	4
	GS5	0	1	2	3	4
	GS6	0	1	2	3	4
	GS7	0	1	2	3	4
EWB	GE1	4	3	2	1	0
	GE2	0	1	2	3	4
	GE3	4	3	2	1	0
	GE4	4	3	2	1	0
	GE5	4	3	2	1	0
	GE6	4	3	2	1	0
FWB	GF1	0	1	2	3	4
	GF2	0	1	2	3	4
	GF3	0	1	2	3	4
	GF4	0	1	2	3	4
	GF5	0	1	2	3	4
	GF6	0	1	2	3	4
	GF7	0	1	2	3	4
PCS	C2	4	3	2	1	0
	C6	0	1	2	3	4
	P1	4	3	2	1	0
	P2	4	3	2	1	0
	P3	4	3	2	1	0
	P4	0	1	2	3	4
	P5	0	1	2	3	4
	P6	4	3	2	1	0
	P7	4	3	2	1	0
	BL2	4	3	2	1	0
	P8	4	3	2	1	0
	BL5	0	1	2	3	4

Time to Deterioration of QoL

A deterioration of QoL is defined as a decrease of at least 10-point in the FACT-P total score from baseline.

In patients with QoL deterioration, the time to deterioration of QoL is defined as the time interval from the date of randomization to the first date a decline from baseline of 10 points or more in the FACT-P total score is recorded.

In patients without FACT-P deterioration, the time to deterioration of QoL will be censored on the date of the last FACT-P total score is calculable. Patients with no baseline FACT-P total score and patients with no post baseline FACT-P total score, time to deterioration of QoL will be censored on the date of randomization.

EQ-5D-5L

The EQ-5D-5L is an international standardized nondisease specific (i.e., generic) instrument for describing and valuing health status.

The EQ-5D-5L has 5 domains: Mobility, Self-Care, Usual Activities, Pain/Discomfort and Anxiety/Depression. Each domain has 5 response levels (coded as 1 = no problems, 2 = slight problems, 3 = moderate problems, 4 = severe problems, 5 = extreme problems). There should be only one response for each domain.

In addition, EQ-5D-5L has a Visual Analogue Scale (VAS) that elicits a self-rating by the respondent of his/her health status from the worst health status (0) to the best health status (100).

6.1.2.5 Time to Initiation of a New Antineoplastic Therapy

The initiation of a new antineoplastic therapy is based on the information collected in the (new) prostate cancer therapy CRF pages about all antineoplastic therapies, including cytotoxic and hormone therapies, initiated for prostate cancer subsequent to the study drug.

In patients with a new antineoplastic therapy initiated for prostate cancer, time to initiation of a new antineoplastic therapy (i.e., time to new antineoplastic therapy) is defined as the time interval from randomization to the date of first dose administration of the first antineoplastic therapy.

In patients with no new antineoplastic therapy initiated for prostate cancer, time to start of new antineoplastic therapy will be censored on the last visit date or the date of randomization, whichever occurs last.

6.1.2.6 Time to PSA Progression

Only results from PSA samples taken before the initiation of any new antineoplastic therapy after the start of study drug will be considered and are referred to in this section.

A PSA progression is defined as a $\geq$ 25% increase and an absolute increase of $\geq$ 2 ng/mL above the nadir (i.e., lowest PSA value observed postbaseline or at baseline), which is confirmed by a second consecutive value at least 3 weeks later.

The date of PSA progression is the first date the PSA progression is observed. In patients with PSA progression, time to PSA progression (TTPP) will be calculated as the time from randomization to the date of first observation of PSA progression.

In patients with no PSA progression, TTPP will be censored on the date of the last PSA sample taken. Patients with PSA progression after 2 or more consecutive missed PSA assessments (i.e., time interval >6 months=182 days between 2 consecutive PSA samples) will be censored on the date of last PSA assessment prior to the first missed assessment.

In patients with no baseline PSA and patients with no post-baseline PSA results, TTPP will be censored on the date of randomization.

6.1.2.7 PSA Undetectable Level

Only results from PSA samples taken before the start of any new antineoplastic therapy will be considered and are referred to in this section.

The undetectable level of PSA is defined as a level < 0.2 ng/mL.

For patients with a detectable level of PSA at baseline, a dichotomous variable ('Y'/'N') is derived to assign 'Y' to patients with any post-baseline PSA sample results < 0.2 ng/mL. otherwise 'N' is assigned. For patients with undetectable level of PSA at baseline, this variable will not be calculated.

6.1.2.8 Objective Response Rate

The objective response rate is based on the response assessments made on soft tissue lesions according to RECIST version 1.1 (on CT/MRI). ORR is evaluated both by ICR and by investigators' assessments. The ORR by ICR is the primary variable for this endpoint.

The RECIST response assessments on target and non-target soft tissue lesions as well as the RECIST overall (timepoint) response are defined as one of the following categories:

- CR = complete response
- PR = partial response (not an option for non-target lesions)
- SD = stable disease (not an option for non-target lesions)
- PD = progressive disease
- Non-CR/non-PD= not complete response and not progressive disease (not an option for target lesions)
- NE = not evaluated/ not all evaluated/ not evaluable
- NA = not applicable (assessment in case of no lesion of the kind at baseline)

RECIST Response by ICR

The RECIST response assessments on target and non-target soft tissue lesions as well as the RECIST overall (timepoint) response are provided by the ICR at every study visit.

For patients with no target and no non-target soft tissue lesion at study entry, the RECIST overall time point response is set to 'NA', unless an unequivocal new soft tissue lesion is identified (in which case the time point response assessment is PD) or unless imaging is not evaluable (in which case the time point response assessment is NE).

Details on the processes for RECIST overall time point response assessments by ICR are provided in the ICR Charter. Only the assessments from the ICR radiologist selected by the adjudicator are used for analysis.

Best RECIST Response by ICR

The best RECIST response by ICR in the soft tissue disease corresponds to the best of the RECIST overall time point response assessment reported by ICR at any time during the treatment period, or up to the start of a new antineoplastic therapy after the study treatment.

The best RECIST response assessment by ICR is derived according to the following decreasing order, starting with the possible best overall timepoint response: CR, PR, SD, Non-CR/Non-PD, PD, NA, NE. For patients still on treatment after or by the cut-off date, the best RECIST response can be derived as the best of the RECIST overall time point response assessments recorded by ICR for the time on study up to the data cut-off date. Patients with no post-baseline assessment at any visit are reported in a ‘‘Not evaluated’ category.

Patients with CR or PR as best RECIST response by ICR will be considered as responders in ORR by ICR, others will be considered as not responders.

RECIST Response by Investigators

The RECIST response assessments on target and non-target soft tissue lesions are provided by investigators at every study visit. At each of these time points, the RECIST overall time point response is then derived according to RECIST for all possible combinations of tumor response assessments made by investigators in target and non-target lesions in soft tissues with or without the appearance of new unequivocal lesions as described in Table 4

Table 4 RECIST Overall Time Point Response derivation for all Combinations of Tumor Responses in Target and Non-target Lesions with or without Appearance of new Soft tissue Lesions as assessed by Investigators

Target Lesions*	Non-Target Lesions*	Unequivocal New Lesions^	RECIST Overall Response^
CR	CR	No	CR
CR	NA	No	CR
NA	CR	No	CR
CR	Non-CR/Non-PD	No	PR
CR	NE	No	PR
PR	Not PD	No	PR
SD	Not PD	No	SD
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD
NA	Non-CR/Non-PD	No	Non-CR/Non-PD
NA	NA	No	NA [$]
NA	NE or Not Done	No	NE
NE or Scan not Done	Not PD	No	NE

* The category “Any” includes all possible categories (incl. NE or NA or Not Done). The category “Not PD” includes: CR, Non-CR/Non-PD, NA, NE and Not Done.

^ Missing assessments during study visits will not be imputed. Missing assessments for “Unequivocal new lesions” will be reported as “Missing” and will lead to “NE” for the derived RECIST Overall Response.

[$] The RECIST overall time point response will only be derived to “Not applicable” for patients with no target and no non-target soft tissue lesions at study entry in absence of new unequivocal lesion.

Best RECIST Response by Investigators

The best RECIST response by investigators is similarly derived as the best RECIST response by ICR but based on investigators assessments.

6.1.2.9 BPI-SF

The BPI-SF allows subjects to rate the severity of their pain and the degree to which their pain interferes with common dimensions of feeling and function. The BPI used in this study is the short form.

It is a self-administered questionnaire designed to capture pain severity and pain interference, as follows:

- Pain severity: worst, least, average and current pain (Question 3 to 6) rated on 0–10 scale with 0 = "no pain" and 10 = "pain as bad as you can imagine"
- Pain interference: general activity, mood, work, walking ability, relations, sleep and enjoyment of life (item A to G of Question 9), rated on 0–10 scales (with 0 = "no interference" and 10 = "interferes completely").

Composite scores of pain severity and the pain interference will be calculated by averaging their items. To calculate the composite score of severity, all four items should be completed; otherwise, the score will be treated as a missing value. For the interference score, if there are missing items, the score will be prorated as follows:

> Prorated score = (sum of item scores)/(number of items answered), as long as more than 50% of the items are answered in Question 9 (i.e., a minimum of 4 of 7 items).

Each score ranges from 0 to 10 with higher scores representing a higher level of pain or interference.

Time to Pain Progression

Pain progression event is defined as an increase of ≥ 30% from baseline in the average BPI-SF pain severity score.

In patients with pain progression event, time to pain progression is defined as time from randomization to the first pain progression event.

In patients with no pain progression event, time to pain progression will be censored on the last visit date where BPI-SF was collected. In patients with baseline score missing or no post-baseline score, the time to pain progression will be censored on the date of randomization.

6.1.3 Other Efficacy Variables

6.1.3.1 Combined Response (Soft tissue Lesions and Bone Lesions)

At each timepoint, the combined (i.e., overall) timepoint response assessment is based on the combination of the RECIST overall timepoint response assessment based on RECIST

version 1.1 for soft tissue lesion on CT/MRI, and the response assessment for bone lesions on bone scans.

Table 5 describes the possible combinations of tumor response to report the overall timepoint response assessment.

Table 5 Combined Time Point Response Provided for any Combinations of Tumor Responses in Soft tissue Lesions and Bone Lesions

RECIST 1.1	Bone Scan		The Combined (i.e., Overall) Time Point Response
CR	NED *		CR
CR	NA		CR
NA	NED*		CR
CR	NON-PD		PR
CR	PDu		PR
PR	NED*	Not NE = & not PD	PR
PR	NON-PD		PR
PR	PDu		PR
PR	NA		PR
SD	NED*	Not NE = & not PD	SD
SD	NON-PD		SD
SD	PDu		SD
SD	NA		SD
PD	Any		PD
Any	PD		PD
NON-CR/NON-PD	NED*	Not NE = & not PD	NON-CR/NON-PD
NON-CR/NON-PD	NON-PD		NON-CR/NON-PD
NON-CR/NON-PD	PDu		NON-CR/NON-PD
NON-CR/NON-PD	NA		NON-CR/NON-PD
NA	NON-PD		NON-CR/NON-PD
NA	PDu		PDu
Any but PD	NE		NE
NE	Any but PD		NE
NA	Not applicable		NA

PDu: unconfirmed Progression of Disease

NED *: ICR reported "No Evidence of Disease" in patients with bone lesion(s) at baseline; Investigators reported CR for those cases.

Combined Response by ICR (Soft tissue Lesions and Bone Lesions)

The combined (i.e., overall) time point response is provided by ICR for each study visit. These are based on the assessments provided by the ICR radiologist and ICR bone scan reader which are selected by the ICR adjudicators.

Best Combined Response by ICR (Soft tissue Lesions and Bone Lesions)

The best combined (i.e., overall) response is the best of the overall time point response assessment reported at any time during the study. It is provided by ICR.

The best of the overall time point response assessments is according to the following decreasing order, starting with the possible best response: CR, PR, SD, Non-CR/Non-PD, unconfirmed PD, PD, NA, NE. For patients still on treatment after or by the cut-off date, the

best overall response is the best of the overall time point response assessments recorded for the time on study up to the data cut-off date. Patients with no post-baseline assessment at any visit are reported in the 'Not evaluated' category.

Combined Response by Investigators (Soft tissue Lesions and Bone Lesions)

The combined (i.e., overall) time point response is provided by investigators for each study visit.

Best Combined Response by Investigators (Soft tissue Lesions and Bone Lesions)

The best combined (i.e., overall) response is derived from the overall time point response assessments provided by investigators. It is the best of the overall time point response assessment reported at any time during or at the end of the treatment period, or up to the start of a new antineoplastic therapy after the last dose of study treatment.

It is similarly derived as for the best overall response by ICR but based on investigators assessments.

6.1.3.2 PSA reduction

Only results from PSA samples taken before the start of any new antineoplastic therapy will be considered and are referred to in this section.

The largest PSA decrease from baseline will be calculated as the percentage change from baseline to the smallest PSA value after baseline. For subjects with no decrease from baseline in PSA, the smallest increase from baseline in PSA will be used. For subjects with no post baseline PSA value, the largest decrease from baseline in PSA will be set to missing.

The PSA decline of at least 50% from baseline, will be defined as binary variable for achieving this criterion based on the lowest PSA value observed post-baseline. For subjects with no post baseline PSA value, this variable will be set to missing.

The PSA decline of at least 90% from baseline, will similarly be defined as binary variable for achieving this criterion.

6.2 Safety Variables

Safety will be assessed by evaluation of the following variables:

- Treatment-emergent AEs (TEAEs; frequency, severity, seriousness, and relationship to study drug).
- Clinical laboratory variables (hematology, biochemistry, PSA and testosterone)
- Vital signs (systolic and diastolic blood pressure and pulse rate) and weight
- 12-lead ECG

Treatment-Emergent Adverse Event (TEAE)

A TEAE is defined as an AE that occurs or worsen at any time during the treatment emergent period. The treatment emergent period is defined as the time interval from the first study drug intake up to 30 days after the date of the last dose of study drug, study discontinuation or the start of new antineoplastic therapy, whichever occurs first.

If AE start date is the same date as the study start, then it will be considered as a TEAE if the box *'Onset after first dose of study'* is ticked in the Electronic case report form(eCRF). If a patient experiences an event both during the pre-investigational period and during the investigational period, the event will be considered as TEAE only if it has worsened in severity (i.e., it is reported with a new start date).

AE with both a missing start and stop dates, and AEs with a missing start date but has a known stop date which is on or after the first dose of study drug will be considered treatment-emergent.

AEs will be coded to System Organ Class (SOC) and preferred term using MedDRA v21.0 (or later version) and severity graded using National Cancer Institute's Common Terminology Criteria version 4.03 (NCI-CTCAE v4.03) for AEs.

Drug-Related TEAE

A drug related TEAE is defined as any TEAE with possible or probable relationship to study drug as assessed by the investigator in the eCRF or with missing assessment of the causal relationship.

AEs of Special Interest

AEs of special interest are defined in Table 6

Table 6 Selection Criteria for AEs of Special Interest

Event of special interest	Selection based on MedDRA v21.0
Convulsion	Narrow SMQ of 'Convulsions'
Hypertension	Narrow SMQ 'Hypertension'
Neutrophil count decreased	Preferred terms of 'Neutrophil count decreased', 'Neutropenia', 'Agranulocytosis',' Granulocyte count decreased', 'Granulocytopenia', 'Febrile neutropenia', 'Neutrophil percentage decreased', 'Band neutrophil count decreased', and 'Band neutrophil percentage decreased'
Cognitive/memory impairment	All preferred terms under the MedDRA High Level Group Term: 'Mental impairment disorders'
Ischemic Heart Disease	Narrow SMQs of 'Myocardial Infarction' and 'Other Ischaemic Heart Disease'
Other selected cardiovascular events	Narrow SMQs of 'Haemorrhagic central nervous system vascular conditions,' 'Ischaemic central nervous system vascular conditions' and 'Cardiac failure'
Posterior reversible encephalopathy syndrome	Preferred term 'Posterior reversible encephalopathy syndrome'
Fatigue	Preferred terms of 'Fatigue', 'Asthenia'
Fall	Preferred term 'Fall'
Fractures	All preferred terms under the MedDRA High Level Group Terms: 'Fractures'
Loss of consciousness	Preferred terms of 'Loss of consciousness', 'Syncope',' Presyncope'
Thrombocytopenia	Preferred terms of 'Thrombocytopenia', 'Platelet count decreased'
Musculoskeletal events	Preferred terms of 'Back pain', 'Arthralgia', 'Myalgia', 'Musculoskeletal pain', 'Pain in extremity', 'Musculoskeletal stiffness', 'Muscular weakness', 'Muscle spasms'
Severe cutaneous adverse reactions (SCAR)	Narrow SMQ of 'Severe cutaneous adverse reactions'
Table continued on next page	

Event of special interest	Selection based on MedDRA v21.0
Angioedema	Narrow SMQ of 'Angioedema'
Rash	All preferred terms including term "Rash"
Second primary malignancies	Narrow SMQs of 'Malignant or unspecified tumours' customized to exclude preferred terms of 'Congenital fibrosarcoma', 'Congenital malignant neoplasm', 'Congenital retinoblastoma', ', 'Metastases to…', 'Metastasis', 'Metastatic neoplasm', 'Prostate cancer…', 'Carcinoid tumour of the prostate', and 'Neoplasm prostate' AND (inclusive of) Narrow SMQ of 'Myelodysplastic syndrome' AND (inclusive of) All preferred terms under High Level Term of 'Myeloproliferative disorders (excl leukaemias)' Note: Non-melanoma skin cancers are excluded (preferred terms of 'Basal cell carcinoma', 'Basosquamous carcinoma', 'Basosquamous carcinoma of skin', 'Keratoacanthoma', 'Skin cancer', 'Skin cancer metastatic', 'Squamous cell carcinoma', 'Squamous cell carcinoma of skin', 'Lip squamous cell carcinoma') Note: Those selected SPM cases will be adjudicated by medical review, which will confirm the evidence of a second primary malignancy

6.3 Exploratory Endpoint ([redacted])

[redacted]

6.4 Other Variables

Previous and Concomitant Medication

Previous medications are defined as non-prostate cancer related medications taken within 28 days prior to the screening visit and up to the first dose of study, and all prior prostate cancer related medications.

Concomitant medication is defined as medication with at least one dose taken between the date of first dose (inclusive) and up to 30 day safety follow-up visit after study drug ended.

A medication can be both flagged as previous and concomitant.

Dose Reduction and Interruption

During the study, patients who experience a NCI-CTCAE (version 4.03) grade 3 or higher AE (except liver function test AE) that is attributed to the study drug and cannot be ameliorated by the use of adequate medical intervention and/or dose reduction, may interrupt study drug for 1 week or until the toxicity grade improves to grade 2 or lower in severity. Study drug may be restarted at the original dose (160 mg/day) or a reduced dose (120 mg or 80 mg/day) in consultation with the Medical Monitor. After dose reduction, based on patient tolerance, study drug may be increased to a maximum dose of 160 mg/day per investigator discretion.

Enzalutamide must be interrupted during the evaluation of symptoms suspicious of PRES (headache, lethargy, confusion, blindness and other visual and neurological disturbances, with or without associated hypertension).

Restarting treatment at a reduced dose or after treatment interruption for > 2 weeks must be discussed with the Medical Monitor.

The total number of dose reduction and the total number of interruptions will be calculated.

Duration of Exposure

The length of time on treatment will be calculated in days and in months.

For patients who discontinued treatment prior to the data analysis cut-off date

Duration of exposure (months) = [(date last dose of study - date of first dose) + 1]/ 30.4375

For patients who did not discontinue treatment prior to the data analysis cut-off date

Duration of exposure (months) = [(data cut-off date - date of first dose) + 1]/ 30.4375

Average Daily Dose

The average daily dose is based on the actual dose taken while taking into account dose reduction and dose interruption periods as recorded in the dosing page of the eCRF. The average daily dose is;

$$\frac{\text{The cumulative dose}}{\text{Duration of exposure (in days)}}$$

where the cumulative dose is defined as the sum of all daily dose actually taken.

Percent Overall Compliance (compared to the theoretical full dose of 160 mg/day)

Percent overall compliance is based on the drug accountability data as recorded in the IVRS system.

Percent overall compliance is defined as the total number of capsules taken divided by the total number of capsules that should have been taken:

$$\frac{\text{[Total number of capsules consumed]}}{\text{[('Date last study drug returned' - 'Date first dose') + 1] x 4}} \times 100$$

The total number of capsules consumed will be calculated based on the number of capsules dispensed at all study visits minus the number of capsules indicated as returned, from the kits returned. The kits dispensed at the last visit for patients still on treatment by the data analysis cut-off date and for which the number of capsules returned is unknown will not be considered.

Refer to Appendix 2 - Adjustments in Calculation of Dose Compliance Details the calculations for dose compliance.

7 STATISTICAL METHODOLOGY

7.1 General Considerations

For continuous variables, descriptive statistics will include the number of patients (n), mean, standard deviation, median, minimum and maximum. When needed, the use of other percentiles (e.g., 10%, 25%, 75% and 90%) will be mentioned in the relevant section.

Frequencies and percentages will be displayed for categorical data. Unless otherwise specified, percentages by categories will be based on the number of patients with no missing data, i.e. will add up to 100%.

All summary tables and figures will be presented by treatment arm and overall, unless stated otherwise. All listings will be produced by site and patient id numbers in ascending order.

Disposition, demographics, other baseline characteristics and efficacy data will be summarized based on the ITT population, unless stated otherwise. Safety analysis summary and other summaries based on SAF are presented by actual treatment received, unless stated otherwise.

All statistical comparisons will be made using two sided tests. All null hypotheses will be of no treatment difference. Multiplicity adjustments are specified in Section 7.4.2.1

All data processing, summarization, and analyses will be performed using SAS® Version 9.1.3 or higher on Unix. Specifications for table, figures, and data listing formats can be found in the TLF specifications for this study.

For the definition of subgroups of interest, please refer to Section 7.9

A data analysis cut-off date for the database will be used. All data from visits or assessments done prior to the cut-off date will be reported.

7.2 Study Population

7.2.1 Disposition of Subjects

The following subject data will be presented:

- Number of patients with informed consent form (ICF), discontinued before randomization, randomized, based on patients with ICF (not presented by treatment);
- Number and percentage of patients randomized in each analysis set, patients who took study drug , patients who did not take study drug
- Number and percentage of patients by study visit, based on the SAF
- Number and percentage of patients with scan performed by study visit and unscheduled visit
- Number and percentage of subjects who discontinued from the treatment by primary reason for discontinuation
- Number and percentage of subjects who discontinued from the 30 day follow-up by primary reason for discontinuation

- Number and percentage of subjects who discontinued from the long term follow-up by primary reason for discontinuation
- Number of subjects with protocol deviations
- Number of subjects per protocol version

Screen failures information, inclusion/exclusion from analysis set, treatment disposition, 30 day follow-up disposition, and long term follow-up disposition, randomization information and dates of first and last evaluations will be listed. The protocol deviations, as well as the description of protocol deviation criteria will also be listed.

7.2.2 Protocol Deviations

Protocol deviations as defined in the study protocol (Section 8.1.6 Protocol Deviations) will be assessed for all randomized subjects.

The protocol deviation criteria will be uniquely identified in the summary table and listing. The unique identifiers will be as follows:

PD1 - Entered into the study even though they did not satisfy entry criteria,

PD2 - Developed withdrawal criteria during the study and was not withdrawn,

PD3 - Received wrong treatment or incorrect dose,

PD4 - Received excluded concomitant treatment.

Subjects deviating from a criterion more than once will be counted once for the corresponding criterion. Any subjects who have more than one protocol deviation will be counted once in the overall summary.

The number and percentage of subjects meeting any criteria will be summarized for each criterion and in total, by treatment group, as well as by study site. Inclusion and exclusion criteria will be summarized for each criterion by treatment group and overall.

A data listing will be provided by site and subject.

7.2.3 Demographic and Other Baseline Characteristics

Descriptive statistics for age and height at study entry will be presented along with frequency tabulations for age group (< 65, 65- <75 and >= 75; and EudraCT age groups), ethnicity, race, Geographic region, at study entry. The weight, body mass index (BMI), ECOG status, PSA, and ALP at baseline will also be presented by descriptive statistics. The 10%, 25%, 75% and 90% percentiles will be provided for PSA, and ALPat baseline.

Number and percentage of subjects randomized in each country and site will also be summarized.

Medical history is coded in the Medical Dictionary for Regulatory Activities (MedDRA) version 18.0, and will be summarized by SOC and preferred terms, as well as by preferred terms alone, by treatment group and overall on the ITT population.

Prostate cancer history will be summarized by presenting the number and percentage of subjects for tumor and lymph node stages, Gleason scores (7 or less versus 8 or more) at initial diagnosis, volume of disease (Low versus High), incidence and location of metastases (incl. the number of bone lesions per category: 1, 2-4, 5-9, 10-19, 20 or more, and TNC; if a range which does not fall in one of these categories is reported, it will be reported in the category that includes the lower limit of the reported range) as well as per summary category (Bone only versus Soft tissue only versus Both bone and soft tissue), and previous therapies: prior docetaxel therapy use (none, 0-5 cycles, 6 cycles) and ADT prior use or orchiectomy (none, ≤3 months, >3 months; considering the ATC 4th level 'gonadotropin releasing hormone analogues' or Preferred WHO name 'degarelix'). Descriptive statistics will also be used to summarize the duration of disease, which is the duration between the date of randomization and the date of initial diagnosis (expressed in months).

Prior radiation, prior procedures will also be summarized.

Non-prostate cancer related medical history including data up to the start of study drug will be coded using MedDRA. The summary table (number and percentage of subjects) will be presented alphabetically by SOC and decreasing order of frequency of preferred terms within each SOC.

The family history of cancer and underlying conditions and malignancy risk factors will be listed.

7.2.4 Previous, Concomitant Medications and New Antineoplastic Therapy

Previous and concomitant medications, including prostate cancer drug therapies, will be coded with World Health Organization Drug Reference List (WHO-DD), and will be summarized by presenting the number and percentage of subjects by therapeutic subgroup (ATC 2nd level) and chemical subgroup (ATC 4th level) and preferred WHO name. It will be ordered alphabetically by ATC subgroup and decreasing order of frequency of preferred WHO name within each ATC class. Subjects taking the same medication multiple times will be counted once per medication and period.

Concomitant medications are those medications or therapies with at least one dose taken between the date of first dose (inclusive) and the date of last dose (inclusive) of study drug and up to 30 day safety follow-up visit. All concomitant medications will be summarized.

New antineoplastic therapy for prostate cancer are therapies with at least one dose taken after the last dose of study drug, including cytotoxic and hormone therapies. They are recorded as 'antineoplastic medication' in the eCRF (either prostate cancer concomitant eCRF page, with 'antineoplastic medication' ticked; or on the new prostate cancer drug therapy eCRF page with 'antineoplastic medication' ticked).

All previous prostate cancer related medications or therapies reported in the corresponding eCRF will be summarized. All concomitant prostate cancer related medications or therapies reported in the corresponding eCRF will be summarized separately. All (new) antineoplastic prostate cancer therapies (started on or after first dose of study drug date) will also be

summarized separately, along with a summary table on the first new antineoplastic prostate cancer after study drug ended.

Non-prostate cancer related previous medications which will be summarized, are those medications or therapies with at least one dose taken within four weeks prior to the first dose of study drug.

All previous and concomitant non-prostate cancer related medication which can be classified into several chemical and/or therapeutic subgroups is presented in all chemical and therapeutic subgroups.

All previous and concomitant medications recorded in the eCRF will be listed, as well as all new prostate cancer therapies.

7.3 Study Drugs

7.3.1 Exposure

The following information on drug exposure will be presented for each treatment group for the SAF:

- The duration of exposure (number of months) to study medication will be summarized using descriptive statistics, including 10%, 25%, 75% and 90% percentiles.
- Number and percentage of patients on study drug at 6 months and at year 1, 2, 3, and 4 (that is with duration of exposure superior or equal to day 182, 365, 730, 1095, and 1461).
- Number and percent of patient with dose reductions or interruptions, the reasons for these, as well as the number of these per patient
- Descriptive statistics for the average daily dose of the drug patient was exposed to

7.3.2 Treatment Compliance

Percent overall compliance with the dosing schedule will be examined on drug accountability data for patients in the SAF for whom, at least a kit was returned and the first and last days of treatment (or the patient still on treatment by the cut-off date) are known.

Percent overall compliance will be summarized as follows:

- Descriptive statistics will be presented by treatment group and overall.
- Percent compliance will be categorized according to the following categories:
    - less than or equal to 70%
    - greater than 70%, less than or equal to 90%
    - greater than 90%, less than or equal to 110%
    - greater than 110%
      Unknown.

7.4 Analysis of Efficacy

Primary and secondary efficacy analysis will be conducted on the ITT population unless otherwise specified.

The categories '1-5 cycles' and '6 cycles' used at randomization for the stratification factor 'prior docetaxel use' are regrouped in the stratified analyses because of the small number of randomized patients with 1 to 5 cycles of docetaxel as prior medication. This stratification factor therefore becomes prior docetaxel use (yes versus no) in the stratified analyses.

7.4.1 Analysis of Primary Endpoint(s)

The primary efficacy variable is rPFS.

The analysis will be conducted when at least 262 rPFS events (as defined in Section 6.1.1) have occurred.

The effect of Enzalutamide+ADT compared to placebo+ADT will be tested using a stratified log-rank test at the level of significance of 0.05 (2-sided). Stratification factors are the factors used at randomization, prior docetaxel use (yes versus no) and disease volume (low versus high).

The null and alternative hypotheses to be tested are as follows:

The null hypothesis: rPFS for Placebo+ADT and Enzalutamide+ADT are not different
The alternative hypothesis: rPFS for Placebo+ADT and Enzalutamide+ADT are different

The following SAS code will be used to compute the Kaplan-Meier estimates and curves, and the stratified log-rank test:

```
PROC LIFETEST DATA=INPUT
                    ATRISK
                    PLOTS=SURVIVAL(CB)
                    OUTSURV=SURVPL
                    ALPHA=0.05
                    ALPHAQT=0.05 METHOD=KM;
   TIME AVAL*CNSR(1);
   STRATA STRATUM1 STRATUM2/ GROUP=TREATMENT;
RUN;
```

where INPUT is the input dataset
AVAL is the time to the event variable,
CNSR is 0 (patients with events) or 1 (patients without no event, i.e. patient censored)
STRATUM are the stratification variables (volume of disease and prior docetaxel use)
TREATMENT is the treatment variable

Kaplan-Meier methods will be used to estimate the distribution of rPFS events by treatment group. The median rPFS will be estimated using the corresponding 50th percentile of Kaplan-Meier estimates. A two-sided 95% confidence interval will be provided for this estimate by use of the Brookmeyer and Crowley method. The 25^{th} percentile and the 75^{th} percentile of rPFS will also be provided. A Kaplan-Meier plot by treatment group will be presented. The estimates of the event free rate on a 3-monthly basis up to 1 year and every 6 months thereafter will be summarized by treatment group, as long as at least 10 patients are at risk.

The benefit of Enzalutamide+ADT compared to placebo+ADT will be summarized by a single HR with its 95% CI based on a Cox regression model stratified for the prior docetaxel

use and disease volume. The null and alternative hypotheses regarding rPFS can be rephrased in terms of the HR, $\lambda_{ArmA} / \lambda_{ArmB}$, where λ_{ArmA} represents the hazard of rPFS for Enzalutamide+ADT and λ_{ArmB} represents the hazard of rPFS for placebo+ADT. A HR of < 1 indicates that the rPFS is prolonged for patients randomized to Enzalutamide+ADT compared with patients randomized to placebo+ADT. The null and alternative hypotheses, respectively, can be written as follows:

$$H_0: \frac{\lambda_{ArmA}}{\lambda_{ArmB}} = 1 \qquad H_1: \frac{\lambda_{ArmA}}{\lambda_{ArmB}} \neq 1$$

The estimated HR of Enzalutamide+ADT to placebo+ADT, $\lambda_{ArmA} / \lambda_{ArmB}$, and its 95% confidence interval will be provided.

SAS PROC PHREG will be used for the analysis with the "DISCRETE" option for tie breaker as follows:

```
PROC PHREG DATA=INPUT;
   CLASS TRTP(REF='PLACEBO + ADT') STRATUM1 STRATUM2;
   MODEL AVAL*CNSR(1)= TREATMENT / RL TIES=DISCRETE;
   STRATA STRATUM1 STRATUM2;
RUN;
```

where INPUT is the input dataset
AVAL is the time to the event variable,
CNRS is 1 (patients with no events) or 0 (patients with events)
STRATUM are the stratification variables (volume of disease and prior docetaxel use)
TREATMENT is the treatment variable

If the estimate of the HR $\lambda_{ArmA} / \lambda_{ArmB} < 1$ and the results from the log-rank test lead to the rejection of H_0 in favor of H_A, then it will be concluded that Enzalutamide+ADT prolongs rPFS compared to placebo+ADT.

Sensitivity analyses

The efficacy sensitivity analyses for rPFS as defined in Section 6.1.1.2 will be conducted on the ITT population using the same analysis methods as described above. No adjustment will be made for the multiple comparisons in these sensitivity analyses.

A forest plot displaying the HR for treatment comparison and 95% confidence interval will be presented for the different rPFS sensitivity analyses. The HR will be estimated by use of Cox proportional hazards models stratified for the prior docetaxel use and disease volume and treatment as covariate, as in the primary analysis.

SAS PROC PHREG will be used for these sensitivity analyses:

```
PROC PHREG DATA= INPUT;
   CLASS TRTP(REF='PLACEBO + ADT') STRATUM1 STRATUM2;
   MODEL AVAL*CNSR(1)= TREATMENT / RL TIES=DISCRETE;
   STRATA STRATUM1 STRATUM2;
RUN;
```

Subgroup analyses

Subgroup analyses of rPFS will similarly be performed to determine whether the treatment effect is concordant among subgroups. To avoid possible issue related to small number of events, subgroup analyses will not be adjusted for the stratification factors used at randomization. The 25th and the 75th percentiles of rPFS and the reason for censoring will not be summarized in tables for subgroup analyses. Subgroups are defined in Section 7.9

A forest plot displaying the HR for treatment comparison and 95% confidence interval will be presented by subgroup. The HR will be estimated by use of Cox proportional hazards models with treatment as covariate.

SAS PROC PHREG will be used for these subgroup analyses:

```
PROC PHREG DATA=INPUT;
   CLASS TRTP(REF='PLACEBO + ADT');
   MODEL AVAL*CNSR(1)= TREATMENT / RL TIES=DISCRETE;
  BY SUBGROUP;
RUN;
```

7.4.2 Analysis of Secondary Endpoints

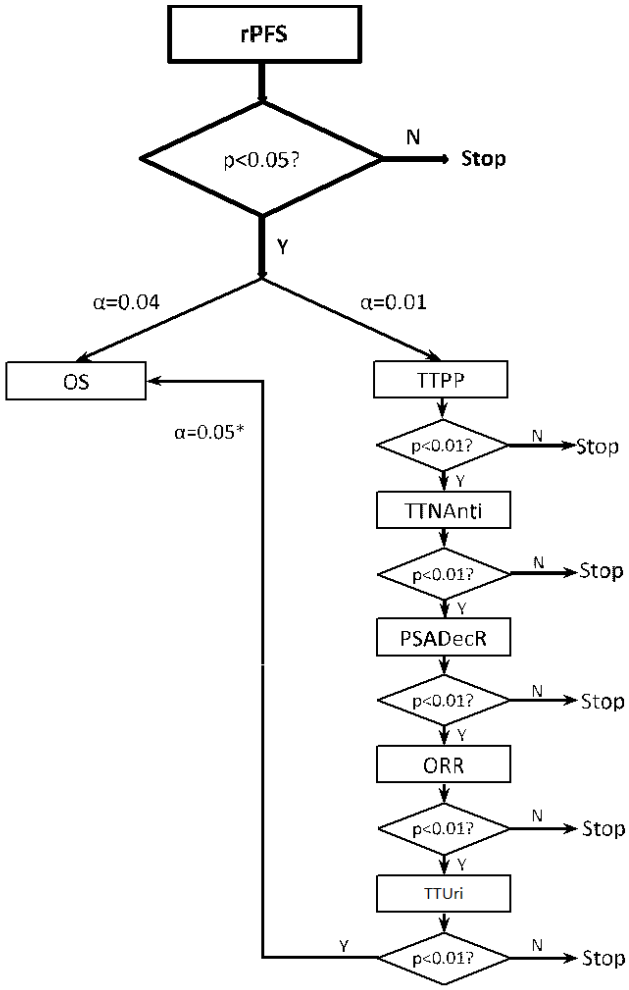
7.4.2.1 Multiplicity Adjustment

All secondary endpoint analyses will be performed at the time of the rPFS final analysis (i.e., when at least 262 rPFS events have occurred).

If the primary endpoint statistical analysis test, conducted at the level of significance of 0.05 (2-sided), is statistically significant then selected secondary endpoints will be tested utilizing a method to preserve the family-wise type I error rate at 5% (2-sided) as described below.

The following six key secondary endpoints will be tested: OS, TTPP, time to initiation of a new antineoplastic therapy, the rate of PSA decline to <0.2ng/mL (PSADecR), ORR, and the time to deterioration in urinary symptoms from the QLQ-PR25. To maintain the family-wise 2-sided type I error rate at 0.05, a parallel testing strategy between OS (with allocated type I error rate 0.04) and the other five endpoints (with allocated type I error rate 0.01) will be performed, as summarized in Figure 1

Figure 1 Testing Strategy for the Primary and Six Selected Secondary Endpoints



rPFS: radiographic progression-free survival; OS: overall survival; TTPP: time to PSA progression; TTNAnti: time to initiation of new antineoplastic therapy; PSADecR: rate of PSA decline to <0.2 ng/mL; ORR: objective response rate; TTUri: the time to deterioration in urinary symptoms from the QLQ-PR25
*OS will be tested at 0.05 only, if all other 5 secondary endpoints analyses are statistically significant at 0.01.

7.4.2.2 Overall Survival

OS is a key secondary endpoint. It will be analyzed using the same analysis methods as for rPFS. At the time of the OS final analysis, subgroup analyses will be conducted using the same analysis methods as for subgroup analyses of rPFS.

One interim analysis and a final analysis are planned for OS. The interim analysis of OS will be performed at the time of the rPFS final analysis (i.e., when at least 262 rPFS events have occurred). The exact significance level for this analysis will be calculated using the O'Brien-Fleming alpha spending function (Lan and DeMets 1983). It will be used to determine the stopping boundaries based on the number of events observed at the interim look to control the overall 2-sided alpha at 0.05 or at 0.04 (as described in Figure 1).

```
PROC SEQDESIGN ALTREF=0.314710 BOUNDARYSCALE=PVALUE ;
   OS_TWOSIDEDOBRIENFLEMING: DESIGN INFO=CUM(XXXX 100) NSTAGES=2
   METHOD=ERRFUNCOBF ALT=TWOSIDED STOP=REJECT ALPHA=0.04 BETA=0.2;
SAMPLESIZE MODEL=TWOSAMPLESURVIVAL;
RUN;
```

If this interim analysis of OS is statistically significant, it will be reported as the final analysis and no subsequent analysis will be performed. At the time of the interim analysis of OS, the death events will be summarized by treatment group and presented relative to control arm.

If this interim analysis of OS is not statistically significant, the final analysis of OS is planned when approximately 342 deaths are observed to ensure an adequate number of events for the final evaluation of OS. At the time of the planned final analysis of OS, no additional analyses of other efficacy endpoints will be conducted.

The primary cause of death will be summarized and listed.

The median follow-up time on study will be calculated as the 50th percentile of Kaplan-Meier estimates from the OS time analysis when reverting the censoring (flag).

Sensitivity analysis

Based on the LATITUDE results, it can be expected that the OS curves for both treatments are closed to each other (i.e., HR~1) over the first 6 months. The proportion of the early events (occurring within the first 6 months from randomization) to later events is much higher at the time of the interim analysis compared to what it will be in the final analysis. A piecewise analysis of OS with 2 periods, before and after 6 months, allows to better estimate the expected treatment effect size in the long run, as it provides an estimate of the HR for the period after 6 months from randomization. This sensitivity analysis at the time of interim OS analysis will not be stratified to avoid possible issues with small number of events in some of the strata.

SAS PROC PHREG will be used for this sensitivity analysis:

```
PROC PHREG DATA = INPUT;
  MODEL AVAL*CNSR(1) = INT1 INT2/ TIES=DISCRETE;
  Int1=0; Int2=0;
  If 0<Aval<=6 and TRTP='Enzalutamide + ADT' then int1=1;
  If 6<Aval and TRTP='Enzalutamide + ADT' then int2=1;
RUN;
```

7.4.2.3 Time to PSA Progression

TTPP is a key secondary endpoint. It will be analyzed at the level of significance of 0.01 (2-sided) using the same analysis methods as for rPFS.

7.4.2.4 Time to Initiation of a New Antineoplastic Therapy

Time to initiation of a new antineoplastic therapy is a key secondary endpoint. It will be analyzed at the level of significance of 0.01 (2-sided) using the same analysis methods as for rPFS.

7.4.2.5 PSA Undetectable Rate

The PSA undetectable rate (also referred as the rate of PSA decline to <0.2 ng/mL) is a key secondary endpoint analyzed at the level of significance of 0.01 (2-sided). It is defined as the percentage of patients with detectable (≥ 0.2 ng/mL) PSA at baseline, which becomes undetectable (< 0.2 ng/mL) during study drug.

In patients of the ITT population with detectable PSA at baseline, the PSA undetectable rates in enzalutamide+ADT and placebo+ADT will be compared by use of the stratified Cochran-Mantel-Haenszel score test.

The SAS code used to implement the Cochran-Mantel-Haenszel score test will be:

```
PROC FREQ DATA=INPUT;
     TABLES STRATUM*TREATMENT*RESPONSE / CMH;
RUN;
```

where INPUT is the input dataset

RESPONSE is the analysis variable with the PSA undetectable response,

STRATUM is the stratification variables (volume of disease and previous docetaxel use)

TREATMENT is the treatment variable

7.4.2.6 Objective Response Rate and RECIST Assessments

The ORR is calculated as the percentage of patients of the ITT population with measurable disease (presence of target lesion at baseline) at baseline who achieved a complete or partial response in their soft tissue disease using the RECIST version 1.1 criteria; that is with CR or PR as best RECIST response.

ORR by ICR

The ORR based on assessments by ICR is a key secondary endpoint. The stratified Cochran-Mantel-Haenszel score test will be used at the level of significance of 0.01 (2-sided) to compare ORR in enzalutamide+ADT versus placebo+ADT.

The SAS code used to implement the Cochran-Mantel-Haenszel score test is similar as described in Section 7.4.2.5

The ORR and associated 95 %CI will be presented in each treatment arm. The 2-sided 95% CI will be calculated using the Clopper-Pearson method based on exact binomial.

On patients of the ITT population with measurable disease at baseline, descriptive statistics will be provided for the best RECIST response categories based on assessment by ICR.

The target lesions description, non-target lesions description, new lesions description, as assessed by ICR, will be listed by visit. The target lesion response, non-target lesion

response, unequivocal new lesions presence along with the RECIST time point response as assessed by ICR and the best RECIST response by ICR will also be listed.

ORR by Investigators

The ORR based on investigators assessments (non-key secondary endpoint) and their RECIST assessments for target lesions description, non-target lesions description, and new lesions will similarly be summarized and listed.

ICR *vs* Investigators

The number and percentage of CR, PR and non-responders as assessed by investigators compared to by the ICR, as well as the number and percentage of concordant and discordant cases of responders will be summarized descriptively.

The number and percentage of rPD and non-rPD as assessed by investigators compared to by the ICR, as well as the number and percentage of concordant and discordant cases of rPD will be summarized descriptively.

7.4.2.7 Quality of Life

Further details about Quality of Life questionnaires, associated derived variables, as well as additional analyses on Quality of Life endpoints can be found in a separate SAP specific to patient reported outcomes.

QLQ-PR25

The QLQ-PR25 questionnaire will be summarized by descriptive statistics as well as the scores in the derived scales.

Time to Deterioration in Urinary Symptoms based on QLQ-PR25

Time to deterioration in urinary symptoms is a key secondary endpoint. It will be analyzed at the level of significance of 0.01 (2-sided) using the same analysis methods as for rPFS.

Time to confirmed deterioration in urinary symptoms will only be summarized descriptively, using the descriptive analysis methods as for rPFS.

FACT-P

The 5 FACT-P subscales (Physical, Social/Family, Emotional, Functional, Prostate Cancer Subscale) and the total FACT-P will be summarized descriptively by visit, by presenting the mean, standard deviation, minimum, maximum and median of the results and change from baseline.

Time to Deterioration of QoL based on FACT-P

Time to deterioration of QoL will be analyzed using the same analysis methods as for rPFS.

EQ-5D-5L

A listing of all items scores will be provided along with a listing of the EQ-5D-5L will be provided as well.

7.4.2.8 Time to First SSE

Time to first SSE (i.e., time to SSE) will be analyzed using the same analysis methods as for rPFS.

7.4.2.9 Time to Castration Resistance

Time to first castration resistance will be analyzed using the same analysis methods as for rPFS.

7.4.2.10 BPI-SF

BPI-SF average score will be summarized descriptively by visit, by presenting the mean, standard deviation, minimum, maximum and median of the results and change from baseline. A listing of all items scores will be provided.

Time to Pain Progression

Time to pain progression will be analyzed using the same analysis methods as for rPFS.

7.4.3 Analysis of Other Efficacy Variables

7.4.3.1 PSA reduction

The number and percentage of subjects with at least 50% decrease in PSA from baseline to the lowest PSA value observed post-baseline will be summarized. Similarly, the number and percent of subjects with PSA decrease of at least 90% from baseline will also be summarized. A 2-sided 95% CI for those rates will be reported by treatment arm using the Clopper-Pearson method based on exact binomial.

```
PROC FREQ DATA=INPUT; distribution
TABLES RESPONSE / BINOMIAL(EXACT) ALPHA=.05;
By TREATMENT;
RUN;
```

The largest decrease from baseline in PSA will be summarised. These variables will be listed.

7.4.3.2 Combined Response (Soft tissue Lesions and Bone Lesions)

On the ITT population, descriptive statistics will be provided, based on both investigators and ICR assessments, for:

- the best combined (i.e., overall) response categories
- the percentage of patients who achieved a complete or partial response as best overall response
- the RECIST time point response, along with the bone lesion response, and the combined (i.e., overall) time point response by visit

The bone lesion response and the overall time point response as assessed by ICR and by investigators will be listed by visit. The best overall response will also be listed for both ICR and investigators assessments.

7.5 Analysis of Safety

All analysis of safety will be presented by treatment group for SAF, unless specified otherwise. All AEs will be listed.

7.5.1 Adverse Events

The coding dictionary for this study will be MedDRA v21.0 or later. Treatment-emergent AEs will be coded to SOC and preferred terms using MedDRA and graded using NCI-CTCAE v4.03.

Treatment-emergent AEs will be tabulated alphabetically by SOC and by preferred terms within SOC.

An overview table will include the following details per treatment group and overall:

- Number of TEAEs,
- Number and percentage of patients with TEAEs,
- Number of NCI-CTC grades 3 and 4 TEAEs,
- Number and percentage of patients with NCI-CTC grades 3 and 4 TEAEs,
- Number of drug related TEAEs,
- Number and percentage of patients with causally drug related TEAEs,
- Number of serious TEAEs,
- Number and percentage of patients with serious TEAEs,
- Number of serious drug related TEAEs,
- Number and percentage of patients with serious drug related TEAEs,
- Number of TEAEs leading to death,
- Number and percentage of patients with TEAEs leading to death,
- Number of drug related TEAEs leading to death,
- Number and percentage of patients with drug related TEAEs leading to death,
- Number of TEAEs leading to withdrawal of treatment,
- Number and percentage of patients with TEAEs leading to withdrawal of treatment,
- Number of TEAEs leading to dose reduction,
- Number and percentage of patients with TEAEs leading to leading to dose reduction,
- Number of TEAEs leading to dose interruption,
- Number and percentage of patients with TEAEs leading to leading to dose interruption,
- Number of drug related TEAEs leading to withdrawal of treatment,
- Number and percentage of patients with drug related TEAEs leading to withdrawal of treatment.

An overview table of the TEAE of special interest will describe by the number and percentage of patients with TEAEs of special interest per treatment group and overall.

The number and percentage of patients with TEAEs, as classified by SOC and preferred terms will be summarized for each treatment group and overall. Summaries will be provided for:

- TEAEs
- NCI-CTC grade 3 or higher TEAEs
- drug related TEAEs,
- serious TEAEs,
- drug related serious TEAEs,
- TEAEs leading to withdrawal of treatment,
- TEAEs leading to dose reduction,
- TEAEs leading to dose interruption,
- drug related TEAEs leading to withdrawal of treatment,
- TEAEs, excluding serious AEs, that equal to or exceed a threshold of 5.0% in any treatment group,
- TEAEs leading to death,
- drug related TEAEs leading to death,

The number and percentage of patients with TEAEs, classified by preferred terms by decreasing frequency within the enzalutamide group will also be summarized.

The number and percentage of patients with TEAEs that equal to or exceed a threshold of 5.0% in any treatment group, classified by preferred terms by decreasing frequency within the enzalutamide group, on a 6 monthly basis (0- <2 months; 2-<6 months; 6-<12 months; 12 months or more). The percentage of patients are calculated based on the number of patients still on treatment at the timepoint corresponding to the lower limit of the time intervals.

To adjust for the treatment duration, the number of TEAEs per 100 patients-years will be summarized, as classified by SOC and PT, per treatment group and overall. The number of TEAEs per 100 patients-years is calculated as the number of events *100 / (sum of the treatment emergent period duration of all patients treated in the corresponding treatment group, in years).

The number and percentage of patients with TEAEs, as classified by SOC and preferred terms will also be summarized by maximum severity (reported according to NCI- CTCAE version 4.03).

In the patient count, if a patient has multiple TEAEs with the same SOC or PT, but with different severity, then the patient will be counted only once with the worst severity. However, if any of the severity values are missing then the patient will be counted only once with missing severity. In the AE count, the AEs will be presented in each category they were classified to. Summaries will be provided for:

- TEAEs by NCI- CTCAE
- drug related TEAEs by NCI- CTCAE,
- serious TEAEs by NCI- CTCAE

7.5.2 Clinical Laboratory Evaluation

Laboratory assessment will be done for the following parameters:

Hematology	Biochemistry	Other
Red blood cell count	Albumin	Testosterone
White blood cell count	Alkaline phosphatase (ALP)	PSA
White blood cell differential	Alanine aminotransferase (ALT)	
Hemoglobin	Aspartate aminotransferase (AST)	
Hematocrit	Blood urea nitrogen	
Platelet count	Calcium	
	Creatinine	
	Glucose	
	Phosphorus	
	Potassium	
	Sodium	
	Total bilirubin	
	Total protein	

Quantitative clinical laboratory variables (hematology, serum chemistry and testosterone), will be summarized using mean, standard deviation, minimum, maximum and median for each treatment group at each visit. Additionally, a within-patient change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way.

Based on the NCI-CTCAE grade of laboratory data, clinical laboratory evaluations will be summarized by grade and by visit, for their comparison to the upper limit. Shift analysis tables on the enzalutamide group will present the shift from baseline to each visit and to the highest grade among the post-baseline visits, by grade. The number and percentage of patients with an increase in grade will be summarized by visit.

Each laboratory result will also be classified as low (L), normal (N), or high (H) at each visit according to the laboratory supplied reference ranges.

Laboratory data will be displayed in listings, along with their NCI-CTCAE grade.

7.5.2.1 Liver Enzymes and Total Bilirubin

The following potentially clinically significant criteria for liver tests - defined as ALP, ALT, total bilirubin, AST, their combination are defined. The patient's highest value during the investigational period will be used.

Parameter	Criteria (*Upper limit of normal[ULN]*)
ALT or AST	> 3xULN
	> 5xULN
	> 8xULN
Total Bilirubin	> 2xULN

Parameter	Criteria (*Upper limit of normal[ULN]*)
ALT and/or AST AND Total Bilirubin(*)	(ALT and/or AST > 3xULN) and total bilirubin > 2xULN
ALT and/or AST AND Total Bilirubin (*) AND Alkaline phosphatase	(ALT and/or AST > 3xULN) and total bilirubin > 2xULN and Alk phos <2x ULN

(*) Combination of values measured within same sample

The number and percentage of patients with potentially clinically significant values in liver enzyme and total bilirubin tests during the investigational period will be presented by treatment group and overall.

7.5.3 Vital Signs

The baseline visit is the last measurement taken prior to initial study drug administration.

Vital signs (systolic blood pressure, diastolic blood pressure, and pulse rate) will be summarized using mean, standard deviation, minimum, maximum and median by treatment group at each time point. Additionally, a within-patient change will be calculated per visit as the post-baseline measurement minus the baseline measurement and summarized by treatment group at each time point. Finally based on the patient's highest value during the treatment period, a summary will present the number and percentage of patients with blood pressure elevation (systolic: ≥ 140 mmHg, ≥ 180 mmHg; diastolic: ≥ 90 mmHg, ≥ 105 mmHg), with increase from baseline (systolic: ≥ 10 mmHg, ≥ 20 mmHg; diastolic: ≥ 5 mmHg, ≥ 15 mmHg) or combination criteria (systolic: ≥ 140 mmHg & ≥ 20 mmHg increase from baseline, ≥ 180 mmHg & ≥ 20 mmHg increase from baseline; diastolic: ≥ 90 mmHg & ≥ 15 mmHg increase from baseline, ≥ 105 mmHg & ≥ 15 mmHg increase from baseline) or any of these criteria.

All vital signs results will be provided in a listing.

7.5.4 Electrocardiograms

12 Lead ECG results and change from baseline will be summarized for each visit by treatment group using mean, standard deviation, minimum, maximum and median. Baseline is defined as the last available measurement prior to the first dose.

Number and percent of patients with normal, not clinically significant abnormal and clinically significant abnormal results for the 12 lead ECG will be tabulated by treatment arm and visit.

QTc will be calculated using Fridericia Formula, QTcF= QT interval / $RR^{0.33}$.

All ECG results will be provided in a listing.

Any abnormal findings/conditions identified during the physical examination are reported in the medical history form or AE form. As a consequence, no separate physical examination listing can be produced.

7.5.5 Pregnancy

Not applicable.

7.5.6 Other Safety-Related Observations

The number and percentage of patients falling in each category of the performance status ECOG score will be summarized per visit and will be listed.

7.6 Analysis of Exploratory Endpoint

[redacted]

7.7 Analysis of Pharmaco-kinetic

Not applicable

7.8 Analysis of Pharmaco-dynamic

Not applicable.

7.9 Subgroups of Interest

Subgroup analyses of rPFS will be conducted to assess the consistency of the treatment effect across the following subgroups of interest:

- Age category (less than 65 years old versus 65 years old or more);
- Geographic region (Europe, North America, Rest of the World);
- ECOG Performance Status (0 versus 1) at baseline;
- Gleason score (less than 8 versus 8 or more) at initial diagnosis;
- Disease location (bone only, versus soft tissue only, versus both bone and soft tissue) at baseline;
- Baseline PSA value (at or below overall median versus above overall median);
- Volume of disease at baseline (low versus high)
- Prior docetaxel use (yes versus no)
- Previous use of ADT or orchiectomy (yes versus no)

7.10 Other Analyses

Not applicable.

7.11 Interim Analysis (and Early Discontinuation of the Clinical Study)

No interim analysis is planned for rPFS. One interim analysis of OS will be performed at the time of the rPFS final analysis. If this interim analysis of OS is statistically significant, it will be reported as the final analysis and no subsequent OS analysis will be performed.

7.12 Handling of Missing Data, Outliers, Visit Windows, and Other Information

The baseline measurement is the last measurement taken prior to initial study drug administration. Both date and time of drug administration and measurement should be

considered to identify the baseline value. If the time is not available, then date only will be used and it will be considered that assessments on day 1 are done pre-dose.

Change from baseline is defined as (post baseline value - baseline value).

To calculate time interval duration, a month is 30.4375 days and a year 365.25 days. Duration expressed in years or months are rounded up to 1 significant digit. The duration between 2 dates d1 and d2 is (d2-d1+1) in days (with d1 before d2).

Treatment day and study day will be calculated in reference to the date of the first dose of study drug. Treatment Day 1 corresponds to the date the patient received the first dose of study drug. For assessments conducted on or after the date of the first dose of study drug, treatment day will be calculated as (assessment date - date of first dose of study drug) + 1. There will be no Treatment Day 0.

Unless otherwise specified, the date of study drug discontinuation refers to the study drug last dose date.

Time to event endpoints will be based on the actual date of event rather than visit date. The date of randomization will always be considered as the start date for the time interval.

For laboratory results collected as < or > a numeric value, 0.0000000001 will be subtracted or added, respectively, to the value unless otherwise specified.

Percentages will be calculated based on the number of patients with non-missing data as the denominator unless otherwise specified.

7.12.1 Missing Data

As a general principle, no imputation of missing data will be done. Exceptions are the start and stop dates of AEs, previous and concomitant medications, the date of initial diagnosis (to estimate the relative study day to calculate cancer duration), dates of cancer treatment (e.g. previous procedure, previous radiotherapy, etc…), the last dose date and the date of death.

The imputed dates will be used to allocate the relative study day, and in addition to determine whether an AE is/is not treatment emergent.

Cases where the onset date of an AE is (partially) missing, will be addressed during the data review meeting in order to determine whether the AE must be considered treatment emergent or not.

Imputation on missing non-prostate cancer related medication dates (to categorize them as previous medications or concomitant medications, or post treatment) and AE dates (to categorize them as TEAE or not) will be done as follows:

- Incomplete Start Day from start date and the corresponding end date is complete: use the later of (first day of the month, first dosing day if first dosing month); but if later than the end date, then impute the start day as the day of the end date.
- Incomplete Start Day from start date and incomplete End Day from end date: use the later of (first day of the month, first dosing day if first dosing month).

- Incomplete End Day from end date: use the earliest of (last day of the month, day of the 30-day follow-up visit if it is the month of the 30-day follow-up visit);
- Incomplete Month or Year: no imputation.

Imputation on missing date of initial diagnosis (cancer duration) and prior cancer treatment, including prostate cancer drug medication/therapies, (e.g., start date, stop date, or date of procedure) will be done as follows:

- Incomplete Day: use the 15th day of the month, if month/year is before first dosing or after last dosing (-for start date imputation- but if later than the end date, then impute the start day as the day of the end date; -for end date imputation- but if earlier than the start date, then impute the end day as the day of the start date).
- Incomplete Month: use 1st of July if the Year is before Year of first dosing, otherwise missing.
- Incomplete Year: no imputation, the derived variable is considered to be missing.

If missing for subjects who started treatment, the last dose date of treatment will be imputed as follows:

- Incomplete Day only: use the earliest of (last day of the month, end of treatment [form] day -if on the same month and year-, day of the 30-day follow-up visit-if on the same month and year-);
- If fully missing or Incomplete Month or/and Year: the date will be imputed by the earliest of (end of treatment [form] date, date of the 30-day follow-up visit)

If partially missing, the date of death will be imputed as follows:

- Incomplete Day: use the earliest of (last day of the month, end of study [form] day)
- Incomplete Month or Year: no imputation

Imputation methods will not be used to determine other endpoints.

Listings will always show the original date information without imputation, and derived parameters requiring imputation (e.g., TEAE indicator, start day, end day, study day) will be flagged.

7.12.2 Outliers

All values will be included in the analyses.

7.12.3 Visit Windows

Visit windows are allowed for certain visits per the schedule of assessments. Patient data will not be excluded from analyses due to the patient's failure to comply with the visit schedule.

For summary tables reporting results by visit, analyses of efficacy and safety variables will be performed according to the analysis visit windows described in the following Table 7

In the case of multiple observations in the same analysis visit window, the observation which is closest to the target date will be used. If the observations have the same distance to the target visit day, the latest one will be used (using date, and time if available). Should there be

two assessments documented at the same time due to the repetition of analysis of the same sample, the one reported as scheduled will be used.

Table 7 Analysis Visit Windows

Visit Day Interval	Scheduled Visit	Analysis Visit
Up to Day 1	Week 1 (Day 1)	Baseline
Day 2 - Day 57	Week 5 (Day 29)	Week 5
Day 58 - Day 127	Week 13 (Day 85)	Week 13
For the next visits during the treatment period: Week X = Week (i*12+13) (i=1,2,3,…) Week X target day = i*12*7 + 85 (i=1,2,3,…) Week X (Wk X target day - 41 days ; Wk X target day + 42 days)		
However for the (last) visit during the treatment period, i.e. the last dose day of study drug is recorded, the upper boundary of the analysis visit window is 5 days after the last dose day, e.g.: Week X = Week (i*12+13) (i=0,1,2,3,…) Week X target day = i*12*7 + 85 (i=0,1,2,3,…) Week X (Wk X target day - 41 days ; last dose day + 5 days)		
(last dose day + 6 ; last dose day + 42)	last dose day + 30 days	30 Day Safety F-up *(after last dose or prior to new therapy*)
Patients who end treatment before the data analysis cut-off date and have long-term (LT) follow-up visits, have visits every 12 weeks after last dose day: LT F-up j = j*12 weeks visit after last dose day (j=1,2,3,…) LT F-up j target day = j*12*7 + last dose day (j=1,2,3,…) LT F-up j (j*12*7 + last dose day - 41 ; j*12*7 + last dose day + 42)		

F-up: follow-up

8 DOCUMENT REVISION HISTORY

Version	Date	Changes	Comment/rationale for change
1.0	06-Jul-2016	NA	Document finalized
2.0	30-Jul-2018	Description of testing strategy in Section 7.4.2.1, details on interim analysis of OS in Section 7.9, addition of Section 3.1.4 and Section 7.6 in relation to [redacted]	Changes implemented in accordance with protocol amendments
2.0	30-Jul-2018	Revise sensitivity analyses of the primary variable and categories of the stratification factors for stratified efficacy analyses	Update per latest available baseline data
2.0	30-Jul-2018	Update of section on safety, including additional summary of adverse events	Update per latest available safety data
2.0	30-Jul-2018	Details and clarification in various sections of the statistical analysis plan text	Clarification/details for programming
3.0	15-Nov-2018	Additional sensitivity analysis of rPFS on M1 patients	Further characterize rPFS results

Version	Date	Changes	Comment/rationale for change
3.0	15-Nov-2018	Addition of time to deterioration in urinary symptoms as a key secondary efficacy endpoint	Further evaluate the benefit of enzalutamide plus ADT as compared to placebo plus ADT on QoL
3.0	15-Nov-2018	Add variables to characterize PSA reduction and 95% CI for ORR and PSA reduction	further characterize efficacy results
3.0	15-Nov-2018	Add QTc formula	Adjust QT, further characterize safety results
3.0	15-Nov-2018	Add rules on imputation for partially missing last dose date or date of death	Clarification/details for programming
3.0	15-Nov-2018	Details and clarification in various sections of the statistical analysis plan text	Clarification/details for programming

9 REFERENCES

ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, May 2007, US Department of Health and Human Services, FDA

European Journal of Cancer. New Response Evaluation Criteria in Solid Tumours: Revised RECIST guideline, 2009 (www.ejconline.com; version 1.1;228-247)

Journal of Clinical Oncology. Design and End Points of Clinical Trials for Patients with Progressive Prostate Cancer and Castrate Levels of Testosterone: Recommendations of the Prostate Cancer Trials Working Group, March 2008. (www.jco.org; vol 26:1148 1159)

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 35:549-56, 1979.

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

Applied Survival Analysis, Regression Modeling of Time to Event Data, David W. Hosmer, Jr. and Stanley Lemeshow, 1999, John Wiley & Sons, Inc.

Ryan CJ, Morris M, Molina A, Piulats JR, De Souza P, Li J, et al. Association of radiographic progression-free survival (rPFS) adapted from prostate cancer working group 2 (PCWG2) consensus criteria (APCWG2) with overall survival (OS) in patients (PTS) with metastatic castration-resistant prostate cancer (mCRPC): results from COU AA-302. Annals of Oncol. 2012;23;suppl 9:ix294-ix318.

Fizazi K et al. Abiraterone plus Prednisone in Metastatic, Castration-Sensitive Prostate Cancer. N England Journal of Medicine. 2017; 377:352-36

10 APPENDICES

10.1 Appendix 1 - Summary of Sensitivity Analyses of Primary Endpoint

		Uninformative censoring	Treatment Discontination as event impact	SSE and New therapy as event impact	All deaths as event impact	rPD documented per prot. schedule impact	Not Evaluable/ Evaluated scan impact	No scan for more than 6 months impact	Censoring on competing risk impact	Missing data & censoring on competing risks (incl EOT) impact in M1 patients	rPFS in M1 patients impact	rPD assessed by INV	rPD assessed per PCGW2 by INV	rPD assessed per PCGW2 impact
		rPFS(ICR)_Primary	**rPFS (ICR)_1**	**rPFS (ICR)_2**	**rPFS (ICR)_3**	**rPFS (ICR)_4**	**rPFS (ICR)_5**	**rPFS (ICR)_6**	**rPFS (ICR)_7**	**rPFS (ICR)_8**	**rPFS (ICR)_9**	**rPFS (INV)_10**	**rPFS (INV)_11**	**rPFS (ICR)_12**
EVENT	rPD (ICR)	•	•	•	•		•	• (*)	• (**)	• (***)	•			•
	rPD (INV)											•	•	
	SSE			•										
	New therapy			•										
	Death (within 24w after EOT)	•	•	•		•	•	•	•	•	•	•	•	•
	End of treatment in absence of rPD/death		•											
	Death (at any time)				•									
	rPD (ICR) not documented at visit day +/- 7 days (as per protocol)					• rPD event at the next W12 schedule visit day per protocol								
	(*) i.e. if first rPD date is post a period of 2 consecutive missing scans THEN rPD is not event + censoring date rule													
	(**) i.e. IF first date of (SSE present or new therapy started) < first of rPD date THEN rPD is not event + censoring date rule													
	(***) i.e. IF first date of (SSE present or new therapy started or EOT) < first of rPD date; or rPD is post to a period of 2 consecutive missing scans, THEN rPD are not event + censoring date rule													

		Uninformative censoring	Treatment Discontination as event impact	SSE and New therapy as event impact	All deaths as event impact	rPD documented per prot. schedule impact	Not Evaluable/ Evaluated scan impact	No scan for more than 6 months impact	Censoring on competing risk impact	Missing data & censoring on competing risks (incl EOT) impact in M1 patients	rPFS in M1 patients impact	rPD assessed by INV	rPD assessed per PCGW2 by INV	rPD assessed per PCGW2 impact
		rPFS(ICR)_Primary	**rPFS (ICR)_1**	**rPFS (ICR)_2**	**rPFS (ICR)_3**	**rPFS (ICR)_4**	**rPFS (ICR)_5**	**rPFS (ICR)_6**	**rPFS (ICR)_7**	**rPFS (ICR)_8**	**rPFS (ICR)_9**	**rPFS (INV)_10**	**rPFS (INV)_11**	**rPFS (ICR)_12**
rPFS and NO EVENT	Censoring date Censoring rules on rPFS	on last imaging date	on last imaging date	on last disease assessment date in db	on last imaging date	on last imaging date	on last imaging date	on last imaging date	on last imaging date	on last imaging date	on last imaging date	on last imaging date (by INV)	on last imaging date (by INV)	on last imaging date
	if no post baseline assess (i.e. trt not started)	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date
	if all post baseline overall response assess = NE	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date
	If no baseline assess and in absence of rPD	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date	Rando date
	If last scan(s) overall timepoint asses. are NE for reviewer selected by adjudicator (if adjudicated), or for both reviewers otherwise	-	-	-	-	-	use the last "non-NE" post baseline scan date	-	-	use the last "non-NE" post baseline scan date	-		-	-
	if Scan/rPD date > end of treatment date	-	-	-	-	-	-	-	-	on last scan date before EOT	-		-	-
	if Scan/rPD date > SSE date	-	-	-	-	-	-	-	on last scan date before SSE	on last scan date before SSE	-		-	-
	if Scan/rPD date> New therapeutic treat (1 pt has 2 weeks before rPD)	-	-	-	-	-	-	-	on last scan date before New Therapy	on last scan date before New Therapy	-		-	-
	if 2 or more consecutive missed imaging (i.e no scan for 2 consecutive visits)	-	-	-	-	-	-	on last scan date before missing scans	-	on last scan date before missing scans	-		-	-

10.2 Appendix 2 - Adjustments in Calculation of Dose Compliance Details

The steps described below are used for the calculation of the ***cumulative dose*** based on the drug accountability data. The calculation of the ***cumulative dose*** based on the dose records in eCRF is not covered in this appendix.

By protocol, the dispensed kits should be returned at the next scheduled visit, but that may not always happen in reality and for some patients the kits may be returned later or not all returned. The protocol study visits are scheduled at unequal time intervals. Patients who are ongoing treatment by the time of data cut-off may have not returned all the dispensed kits and some patients may return all or some of the treatment kits after the end of treatment date.

The following general rules apply to the calculation of percent dose compliance:

1. Percent dose compliance should not be calculated for kits not returned, or for which the number of capsules returned are unknown, or returned on the same day or earlier than dispensed
2. Percent dose compliance calculations should be based on treatment kits returned
3. Every kit dispensed should have a total of X capsules, where X=number of capsules that should be taken per day*number of days until next refill. In this study, this number should be 124 (4 capsules per day*31 days) at week 1 day 1 visit, 248 at week 5 visit and 372 at week 13 visit and all treatment period visits thereafter.

The aim is to calculate a running sum of the drug dose at each visit where kit/s is/are dispensed, while adjusting the calculation for time interval between visits, time returned and the time of return in regards to end of treatment date. The cumulative dose is then used to calculate the percent overall compliance using the formula specified in Section 6.3

I. For patients with kits returned prior to the end of treatment

1. <u>At week 1 day 1 visit</u>

- Patients receive one kit only (1 bottle of 124 capsules)
- It is possible that patients forget to return their kit at the next visit and bring it at a later visit than planned
- 124 capsules is maximum 31 days of treatment which is more than the 28 days planned between visits
- the last day of any period will be added to the first kit taken, therefore 1 is added

➔ So, the theoretical cumulative number of capsules used between day 1 and week 5 for a kit returned prior to the end of treatment is:

4*((date of return - start date) +1 or 31 at most)

2. At week 5 visit on day 29

- Patients receive two kits
- It is possible patients forget to return their kits at the next visit and bring it at a later visit than planned or even after treatment ended
- The treatment can also end during the use of the first or the second kit (but it will be assumed that the drug will then be returned after end of treatment for at least one kit or on that day at the earliest)
- 124 capsules is maximum for 31 days of treatment
- 248 capsules is maximum for 62 days of treatment which is more than the 2*28=56 days planned between visits

➔ So, ***per kit***, for kits returned prior to the end of treatment, the theoretical number of capsules used between week 5 and week 13, to add to the cumulative number, is:
4*((date of return - start date)/2 or 31 at most)

3. At week 13 visit and every visit thereafter

- patients receive three kits
- It is possible patients forget to return their kits at the next visit, as requested by protocol, and bring them at a later visit or even after treatment ended
- It is possible patients returned only part of the dispensed kits (i.e., 1 or 2 of the 3 kits) at the next visit but the other one(s) at a later visit than planned - or even after treatment ended-
- The treatment can also end during the use of the first , second or third kit (but it will be assumed that the drug will then be returned after end of treatment for at least one kit or on that day at the earliest)
- 124 capsules is maximum 31 days of treatment
- 248 capsules is maximum 62 days of treatment
- 372 capsules is maximum 93 days of treatment which is more than the 3*28=84 days planned between visits

➔ So, ***per kit***, for kits returned prior to the end of treatment, the theoretical number of capsules between 2 of the 12-weekly visits, to add to the cumulative number, is:
4*((date of return - date of dispense)/3 or 31 at most)

Note: In the calculations in the three steps above, a correction is introduced for the late returns, by capping off the interval between dispensed and returned time points at 31 days. However, after each visit, if all dispensed kits up to that point are returned and the return time is prior to end of treatment, then a more accurate theoretical number of capsules can be calculated based on the dispensed date or the maximum total number of days between current and last visit. E.g.,

➔ At week 13 visit, the ***cumulative*** theoretical number of capsules is:
4* (date of dispense at week 13 - date of trt start+1)
+ 4* ((date of return - date of dispense) or 62 at most)

➔ At week 25 visit or later the cumulative theoretical number of capsules is:
4* (date of dispense at week 25 - date of trt start+1)
+ 4* min ((date of return - date of dispense) or 93 at most)

II. For patients with kits returned on or after end of treatment

The following points should be considered when going through the steps outlined below:

- It is possible that some patients return their kits (some or all of them) after end of treatment (i.e., kits were dispensed before end of treatment and then returned after end of treatment)
- In theory for patients who end treatment there should always be at least one kit returned on or after the end of treatment date
- In these cases the date the kit/s was/were returned should not be considered and the end of treatment date should be used instead
- Such cases would be the last ones to consider in the calculation of the cumulative dose

1. <u>At week 1 day 1 visit</u>

- Patients receive one kit only (1 bottle of 124 capsules)
- It is possible patients forget to return their kits at the next visit and instead bring them after EOT despite that they might have continued with treatment for a much longer time
- 124 capsules is maximum 31 days of treatment which is more than the 28 days planned with next visit
- The last day of any period will be added to the first kit taken => +1 is added

➔ So, the cumulative theoretical number of capsules between day 1 and week 5 for kit returned after the end of treatment, is:
4* ((date of end of treatment - start date) +1 or 31 at most)

2. <u>At week 5 visit on day 29</u>

- Patients receive 2 kits
- It is possible patients forget to return their kits at the next visit and bring it after treatment ended
- It is possible that patients returned one of the kits at the next visit, but the other one after treatment ended
- 124 capsules is maximum 31 days of treatment
- 248 capsules is maximum 62 days of treatment which is more than the 2*28=56 days planned between visits

➔ A) If only one of the 2 kits was returned and it was done after the end of treatment, the theoretical cumulative number of capsules between week 5 and week 13 is:
4*((date of end of treatment - dispense date) or 31 at most)

➔ B) If both kits were returned after the end of treatment, the theoretical number of capsules between week 5 and week 13, to add to the cumulative number up to the previous visit, is:
4*((date of end of treatment - dispense date) or 62 at most)

3. <u>During 2 consecutive 12-weekly visit</u>

- Patients receive 3 kits
- It is possible patients returned one of the kits at the next visit, but the other ones after treatment ended
- It is possible patients forget to return their kits at the next visit and instead bring all of them after end of treatment
- 124 capsules is maximum 31 days of treatment
- 248 capsules is maximum 62 days of treatment
- 372 capsules is maximum 93 days of treatment which is more than the 3*28=84 days planned between visits

➔ A) If only one of the 3 kits was returned after the end of treatment, the theoretical cumulative number of capsules between those 2 consecutive 12-weekly visits is:
4*((date of end of treatment - dispense date) or 31 at most)

➔ B) If two kits were returned after the end of treatment, the theoretical number of capsules between those 2 consecutive 12-weekly visits, to add to the cumulative number up to the previous visit, is:
4*((date of end of treatment - dispense date) or 62 at most)

➔ C) If three kits were returned after the end of treatment, the theoretical number of capsules between those 2 consecutive 12-weekly visits, to add to the cumulative number up to the previous visit, is:
4*((date of end of treatment - dispense date) or 93 at most)

10.3 Appendix 3 - Key Contributors and Approvers

List of Key Contributors and Approvers

Key Contributors

The following contributed to or reviewed this SAP as relevant to their indicated discipline or role.

Primary author (s)

[redacted]

Main Contributors and Reviewers

[redacted]

Author and Approver Signatories

(E-signatures are attached at end of document)

[redacted] Astellas Pharma Europe, was the study statistician for this study.

[redacted], Astellas Pharma Global Development, was the Global Statistical Lead and peer reviewer of this Statistical Analysis Plan

This Statistical Analysis Plan was approved by:

[redacted]
Astellas Pharma Global Development

ELECTRONIC SIGNATURE PAGE

Document Type : Statistical Analysis Plan

Document Control Number : MGC1600724

Amendment Number : N/A

International Study Number : 9785-CL-0335

Departmental Study Number : N/A

Actual Version Number : Version 3.0

Document Version : 4.0

Nonclinical Initial SD Approved Date (UTC) : N/A

Date (UTC)	Signed by	Sign Off Meaning
11/15/2018 21:17:11	[redacted]	Scientific Lead Approval
Full Name / Legal Name	[redacted]	
11/15/2018 21:31:06	[redacted]	Authorized Document Officer Approval
Full Name / Legal Name	[redacted]	
11/15/2018 21:52:04	[redacted]	Study Physician Approval
Full Name / Legal Name	[redacted]	
Full Name / Legal Name		
Full Name / Legal Name		
Full Name / Legal Name		
Full Name / Legal Name		
Full Name / Legal Name		
Full Name / Legal Name		
Full Name / Legal Name		

*UTC: Coordinated Universal Time