Document Type:	Statistical Analysis Plan
Official Title:	A multinational, randomised, double-blind, placebo-controlled, Phase III efficacy and safety study of darolutamide (ODM-201) in men with high-risk non-metastatic castration-resistant prostate cancer
NCT Number:	NCT02200614
Document Date:	12 March 2021

Title page

A multinational, randomized, double-blind, placebo-controlled, Phase III efficacy and safety study of darolutamide (ODM-201) in men with high-risk non-metastatic castration-resistant prostate cancer

Bayer study drug	BAY 1841788 / darolutamide		
Clinical study phase:	III	**Date:**	12 MAR 2021
Study No.:	17712	**Version:**	1.0
Author:	PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Title page **1**

Table of Contents **2**

Table of Tables **2**

List of Abbreviations **3**

1. Introduction **3**

2. Study Objectives **3**

3. Study Design **4**

4. General Statistical Considerations **4**

5. Analysis Sets **4**

6. Statistical Methodology **4**

6.1 Population characteristics 4
6.1.1 Disposition of patients 4
6.1.2 Demographic and Baseline Characteristics 4
6.1.3 Medical history 5
6.1.4 Prior and concomitant medications 5
6.2 Efficacy 5
6.3 Pharmacokinetics/pharmacodynamics 5
6.4 Biomarker evaluation 5
6.5 Safety 5
6.5.1 Extent of exposure 6
6.5.2 Adverse events 6
6.5.3 Deaths and Serious Adverse events 13
6.5.4 Clinical laboratory data 13
6.5.5 12-Lead ECG, QTc 15
6.5.6 Other safety measures 15
6.5.7 Physical examinations 16

7. Document history and changes in the planned statistical analysis: **16**

8. References **16**

9. Appendices **16**

Table of Tables

Table 6–1: MedDRA search criteria for special topics 8
Table 6–2: TEAEs grouped by data driven approach: 10

List of Abbreviations

AE	Adverse Event
BMI	Body Mass Index
CI	Confidence Interval
CRF	Case Report Form
ECG	Electrocardiogram
FAS	Full Analysis Set
HLGT	High level group term
HLT	High level term
MedDRA	Medical Dictionary for Regulatory Activities
MLG	MedDRA Labelling Grouping
n	Number of non-missing values
N/A	Not Applicable
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
nmCRPC	Non-metastatic castration-resistant prostate cancer
PT	Preferred Term
QTc	Corrected QT
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SOC	System organ class
TEAE	Treatment-Emergent AE
TESAE	Treatment-Emergent SAE
WHO-DD	World Health Organization Drug Dictionary

1. Introduction

Study 17712 (ARAMIS) is a multinational, randomized, double-blind, placebo-controlled, phase III efficacy and safety study of darolutamide in men with high-risk non-metastatic castration-resistant prostate cancer (nmCRPC).

This supplement to the statistical analysis plan (SAP) describes the analyses and data presentations for final clinical study report addendum when all patients have discontinued the trial.

This SAP Supplemental 7 version 1.0 is based on integrated protocol version 5.0 (amendment 4) dated 06 JUL 2019, on Supplemental 4 SAP version 1.0 dated 04 DEC 2019 and SAP v4.2 dated 20 SEP 2018.

2. Study Objectives

Refer to main SAP v4.2 dated 20 SEP 2018.

3. Study Design

Refer to main SAP v4.2 dated 20 SEP 2018.

4. General Statistical Considerations

Refer to main SAP v4.2 dated 20 SEP 2018.

5. Analysis Sets

Refer to main SAP v4.2 dated 20 SEP 2018.

6. Statistical Methodology

The statistical analyses will be descriptive. Summaries will be provided for both treatment groups, darolutamide and placebo.

For disposition tables, concomitant medications, concurrent procedures, study-drug exposure tables and adverse events tables, the patients who switched from placebo to darolutamide during the open-label period will be displayed. These following treatment groups will be shown: darolutamide total (double-blind and open-label periods), darolutamide double-blind period, placebo double-blind period, placebo-darolutamide cross-over patients.

6.1 Population characteristics

6.1.1 Disposition of patients

A summary table will be presented for the number of patients enrolled and the number and percentage of patients in each of the defined populations. Re-screened patients will only be counted once.

The reasons for patients excluded from each of the patient populations will also be tabulated. The reasons for discontinuation of study treatment will be tabulated.

6.1.2 Demographic and Baseline Characteristics

No demographics and baseline characteristics analyses will be performed in this addendum.

6.1.3 Medical history

No medical history analysis will be performed.

6.1.4 Prior and concomitant medications

The following concomitant medications and procedures tables will be created:

- Concurrent diagnostic and therapeutic procedures: frequency of patients by procedure
- Concomitant medications: frequency of patients for each drug category
- Concurrent diagnostic and therapeutic procedures related to adverse event (AE): frequency of patients by procedure

The dictionary used for coding medications is the WHO Drug Dictionary.

Concomitant medications, subsequent treatment and concurrent procedures will be displayed for darolutamide total (double-blind and open-label periods), darolutamide double-blind period, placebo double-blind period, placebo-darolutamide cross-over patients.

If the start date of the concomitant medications or concurrent procedures is missing, the respective medications or procedures will be considered under the double-blind period.

In addition to the full analysis set (FAS) population, concomitant medication will also be summarized for the safety analysis set (SAF) population.

6.2 Efficacy

No efficacy analyses will be performed, only follow-up time will be presented.

The duration of follow-up, defined as time from randomization to last contact or death at the database cut-off date, will be summarised using descriptive statistics (number of observations, mean, standard deviation, minimum, median and maximum).

6.3 Pharmacokinetics/pharmacodynamics

No pharmacokinetic analyses will be performed.

6.4 Biomarker evaluation

No biomarker analyses will be performed.

6.5 Safety

All data including the open-label period will be summarized.

For patients randomized to darolutamide, safety events will be displayed separately for:

- the double-blind period

- the double-blind and open-label periods (total)

For patients randomized to placebo, safety events will be displayed separately for:

- the double-blind period
- the open-label period after cross-over to darolutamide treatment

The double-blind period of the study only will be considered for risk difference, risk ratio, interval-specific incidence and prevalence.

If the start date of the AE is missing then the event will be considered occurring during the double-blind period.

No formal statistical tests will be done for the safety endpoints. All analyses for safety will be performed in the SAF population.

6.5.1 Extent of exposure

Extent of exposure will be summarized for the SAF by treatment group, using descriptive statistics. Patients who switched from placebo to darolutamide during the open-label period will be described.

Duration of study treatment will be calculated in days and presented in months as the date of the last dose of any study treatment – date of the first dose of any study treatment + 1.

Dose modifications will be summarized.

6.5.2 Adverse events

All adverse events (AE) whether considered drug-related or not, will be reported on the case report form (CRF) with diagnosis, start/stop dates, dates of any grade change, action taken, whether treatment was discontinued, any corrective measures taken, and outcome. For all events, the relationship to treatment and the severity of the event will be determined by the Investigator. AEs will be classified and coded using the National Cancer Institute (NCI) Common Terminology Criteria Adverse Event (CTCAE), version 4.03.

The treatment period for this study, for purposes of safety analyses, extends from the initiation of study treatment until 30 days after the last administration of study treatment.

Each change in AE grade was entered as a separate record with no automatic link to the original AE record, which may lead to variations of the verbatim for the same AE. This makes a correct grouping of AE grades, which actually belong together, not always possible. An AE is considered as treatment-emergent (TEAE) if there is an AE record which starts during treatment or within the post-treatment time window of 30 days.

For further definitions of the terms AE, SAE, seriousness, intensity, causal relationship with treatment, causal relationship to protocol-required procedures, action taken, and outcome; see Protocol Section 6.6.1.

Descriptive summary tables (frequency and percentage of patients, not of events) will be presented by treatment group and MedDRA version 24.0, or the most recent version for the following:

- Pre-treatment AEs
- TEAEs
- TEAEs with grade 3, 4, or 5
- TEAEs occurring in at least 1% of patients
- TEAEs occurring in at least 5% of patients
- TEAEs leading to study drug withdrawal
- TEAEs leading to dose reduction
- TEAEs leading to drug interruption
- TEAEs leading to dose reduction and/or drug interruption
- Drug-related TEAEs
- Drug-related TEAEs with grades 3, 4, or 5
- Drug-related TEAEs occurring in at least 5% of patients in any treatment group
- Drug-related TEAEs leading to study drug withdrawal
- Drug-related TEAEs leading to dose reduction
- Drug-related TEAEs leading to drug interruption
- Drug-related TEAEs leading to dose reduction and/or drug interruption
- Interval specific and cumulative event rates for TEAEs (for TEAEs with at least a 5% total incidence rate (any grade)).
- Post-treatment non-treatment-emergent AEs

Listings of non-treatment-emergent AEs and treatment-emergent AEs not coded per NCI-CTCAE dictionary will be created.

No subgroup analyses will be performed.

Grouped AE terms defined as special topics

Following TEAEs are considered as special topics, Table 6–1:

Table 6–1: MedDRA search criteria for special topics

Grouped term	MedDRA search criteria
Bone fractures	HLT: Fractures and dislocations NEC (without PTs: Joint dislocation, Joint dislocation pathological) HLT: Limb fractures and dislocations (without PT: Radial head dislocation) HLT: Pelvic fractures and dislocations HLT: Skull fractures, facial bone fractures and dislocations HLT: Spinal fractures and dislocations (without PT: Dislocation of vertebra) HLT: Thoracic cage fractures and dislocations (without PT: Dislocation of sternum))
Fall	PT: Fall PT: Accident
Seizure	MLG: Seizures
Fatigue	MLG: Decreased general strength and energy PT: Lethargy; PT Chronic fatigue syndrome; PT Malaise
Weight decreased	MLG: Weight decreased

MLG: MedDRA Labeling Grouping; PT: Preferred Term; HLT: High Level Term

Following tables will be created for the TEAE of special topics as defined above

- TEAEs
- TEAEs leading to study drug withdrawal
- TEAEs leading to dose reduction
- TEAEs leading to drug interruption
- Exposure adjusted special topics TEAEs per 100 patient years
- Prevalence rates for most common TEAEs
- Risk ratio for TEAEs
- Treatment-emergent SAEs (TESAEs)

Fracture events will be described by a cumulative incidence plot of fracture. A summary of treatment-emergent fracture by bone sparing agent use (bone sparing agent will be selected in CM dataset using ATC codes M05B drugs affecting bone structure and mineralization, A11CC Vitamin D and analogues, A12A Calcium, A12CD Fluoride and H05BA Calcitonins) at study entry will be provided. Association with weight decrease will be presented by histogram of fracture events by patient weight change (the weight collected closest to the start of fracture will be considered and compared to baseline weight).

A listing will be generated for patients with fall treatment-emergent events with syncope and/or loss of consciousness (using the MLG Syncope).

Timing of occurrence of fracture events based on first dose of study drug will be presented. A detailed summary of patients with fall and fracture events will be presented, in addition, a graph will be created to display the time to fall and time to fracture.

An overview of patients at risk for developing an AE seizure will be tabulated by displaying medical history.

To adjust for unequal lengths of study treatment period among patients, and potentially between treatment groups, an additional summary based on event rate per 100 patient years will be performed for all TEAEs, special topics TEAEs and all TESAEs occurring during the double-blind treatment. The event rate per patient is calculated as the total number of events divided by the total treatment duration in years. The treatment duration in years will be calculated as treatment duration in days divided by 365.25.

Additional AE groupings

The following additional tables for treatment-emergent adverse events (TEAEs) will be created. In all tables, results will be shown by treatment group. The double-blind part of the study will be considered only for risk difference, risk ratio, interval-specific incidence and prevalence.

To increase the sensitivity of the analysis, the synonymous or pathophysiologically related TEAEs were grouped using following approaches:

- Utilization of MedDRA hierarchy – i.e. using grouping on the level of system organ class (SOC) or high level group term (HLGT) or high level term (HLT),
- Utilization of predefined customized MedDRA queries denoted as MedDRA Labeling Groupings (MLG). These MLG are created and centrally maintained by Bayer internal coding experts,
- Data driven approach was used for grouping of TEAEs, for which the mentioned above predefined grouping were not available/suitable. For this purpose, a customized MedDRA query was created by selection of MedDRA terms after the review of safety data, as shown in Table 6–2.

Incidence proportions will be calculated as number of patients experiencing an event per number of patients exposed. Incidence proportions will be presented by worst CTCAE grade (grade 1 to grade 5, missing grade, any grade)

- by MedDRA SOC, HLGT, HLT, preferred term (PT) and worst CTCAE grade,
- by MLG, PT and worst CTCAE grade,
- by data driven grouping, PT and worst CTCAE grade.

A separate table will be created for PTs not covered by MLG or data driven grouping.
A definition table for the groupings will be provided, showing the AE grouping, the type of grouping entity used for definition (e.g. MLG, combination of MLGs, MLG + PT, data

driven), the corresponding preferred terms, and the reported terms. Preferred terms which occurred in the data are marked with preceding asterisk (***).

The risk difference 'Darolutamide – Placebo' and risk ratio 'Darolutamide/Placebo' will be calculated together with 95% confidence intervals (CIs). No zero-cell correction will be applied for calculation of the risk ratio, i.e. in case of no events in the placebo arm the risk ratio will not be calculated. The incidence rate ratio for Darolutamide/Placebo will also be calculated with 95% CI.

Risk difference, risk ratio, incidence rate ratio with 95% CIs will also be calculated for the predefined grouped terms defined as special topics, see Table 6–1.

Interval specific and cumulative incidence proportions will be created for grouped TEAEs (i.e. HLGT/HLT, MLG, data driven grouping) with a 5% incidence proportion of at least 5% in either treatment group. The prevalence of these events will also be displayed for the same intervals. Prevalence will be calculated as the number patients with the respective TEAE starting or ongoing in the specific time interval, divided by the number of patients still being in the study at the beginning of the time interval.

Table 6–2: TEAEs grouped by data driven approach:

Grouped term	MedDRA search criteria
Pneumonia and Pneumonitis	PT: Lower respiratory tract infection; Lung infection; Pneumonia; Pneumonia pneumococcal; Pneumonia staphylococcal; Lower respiratory tract inflammation; Pneumonitis
Diabetes mellitus and Hyperglycaemia MLG plus	MLG: Hyperglycaemia PT: Diabetes mellitus; Diabetes mellitus inadequate control; Diabetic metabolic decompensation; Type 2 diabetes mellitus; Diabetic ketoacidosis
Renal impairment MLG combo plus	MLG: Laboratory tests related to reduced renal function MLG: Renal impairment PT: Blood urea increased
Rash MLG combo plus	MLG: Rash MLG: Skin erythema PT: Dermatitis
Dizziness MLG plus	MLG: Dizziness PT: Vertigo
Cerebral ischaemia MLG combo plus	MLG: Cerebral infarction and stroke not specified as hemorrhagic or ischemic MLG: Cerebral ischemic infarction and stroke PT: Cerebral ischaemia: Transient ischaemic attack

MLG: MedDRA Labeling Grouping; PT: Preferred Term; HLT: High Level Term

The following tables will be created:

Incidence proportions by worst CTCAE grade: These tables are company standard tables showing the absolute and relative frequencies in each treatment group, by worst CTCAE grade.

- TEAEs by MedDRA (SOC/ HLGT/ HLT/ PT) and worst CTCAE grade,
- Treatment-emergent grouped adverse events by MedDRA labeling grouping (MLG), PT and worst CTCAE grade,
- Treatment-emergent grouped adverse events by data driven grouping, PT and worst CTCAE grade,
- Definition of adverse events MedDRA labeling groupings (MLG),
- Definition of adverse events data driven groupings,
- TEAEs - PTs not covered by MLG or data driven grouping and worst CTCAE grade.

Risk difference, risk ratio, exposure-adjusted rates and incidence rate ratio: These tables show columns for AE grouping, incidence proportion darolutamide, incidence proportion placebo, risk difference with 95% confidence interval (CI), risk ratio with 95% CI, exposure-adjusted incidence rate darolutamide, exposure-adjusted incidence rate placebo, incidence rate ratio with 95% CI.

- Incidence rates and overall risk ratio for grouped terms defined as special topics – double-blind treatment period
- Incidence rates and overall risk ratio for TEAEs by MedDRA SOC, HLGT and HLT – double-blind treatment period. All numbers will be shown on the HLGT and HLT level, not on the SOC level. The SOC is used for sorting only.
- Incidence rates and overall risk ratio for TEAEs by MLG – double-blind treatment period
- Incidence rates and overall risk ratio for TEAEs by data driven grouping – double-blind treatment period

TEAEs over time for common TEAEs (frequency > 5% in any treatment group)

- Interval-specific and cumulative event rates for most common TEAEs by HLGT/HLT grouping – double-blind treatment period
- Interval-specific and cumulative event rates for most common TEAEs by MLG grouping – double-blind treatment period
- Interval-specific and cumulative event rates for most common TEAEs by data driven grouping – double-blind treatment period
- Prevalence over time for most common TEAEs by HLGT/HLT grouping – double-blind treatment period
- Prevalence over time for most common TEAEs by MLG grouping – double-blind treatment period

- Prevalence over time for most common TEAEs by data driven grouping – double-blind treatment period

Adverse events within SOC Cardiac disorders

The following additional tables for TEAEs by history of cardiac disorders will be created:

- for the subgroup of patients with present medical history in the SOC 'Cardiac history',
- for the subgroup of patients without present medical history in the SOC 'Cardiac history'.

In all tables, results will be shown by treatment group.

Incidence proportions of the TEAEs in the system organ class (SOC) 'Cardiac disorders' will be presented by HLGT, HLT, PT and worst CTCAE grade.

Interval specific and cumulative incidence proportions for TEAEs will be presented for the HLGTs Cardiac arrhythmias, Coronary artery disorders, and Heart failures.

In addition, Exposure adjusted treatment-emergent AE (fracture, weight decrease, cardiac disorders, fall) per 100 patient years will be created for the treatment groups:

- Darolutamide – double-blind period
- Darolutamide – double-blind and open-label periods
- Placebo – double-blind period

Below selection will be used:

- Bone fractures: HLT: Fractures and dislocations NEC (without PTs: Joint dislocation, Joint dislocation pathological), HLT: Limb fractures and dislocations (without PT: Radial head dislocation), HLT: Pelvic fractures and dislocations, HLT: Skull fractures, facial bone fractures and dislocations, HLT: Spinal fractures and dislocations (without PT: Dislocation of vertebra), HLT: Thoracic cage fractures and dislocations (without PT: Dislocation of sternum)
- Weight decrease: MLG weight decrease
- Cardiac arrhythmia: HLGT cardiac arrhythmias
- Coronary artery disorders: HLGT coronary artery disorders
- Heart failures: HLGT Heart failures
- Fall: PT Fall and PT Accident

6.5.3 Deaths and Serious Adverse events

Serious adverse events (SAE) will be classified using the National Cancer Institute (NCI) Common Terminology Criteria Adverse Event (CTCAE), version 4.03 and MedDRA version 24.0 or most recent version

- TESAEs
- TESAEs leading to study drug withdrawal
- TESAEs leading to dose reduction
- TESAEs leading to drug interruption
- TESAEs leading to dose reduction and/or drug interruption
- Drug-related TESAEs
- Listing of TESAEs
- Listing of non-treatment-emergent SAEs

The incidence of deaths in the study and especially deaths up to 30 days of last dose of study drug will be summarized by each treatment group and cause of death. All deaths up to 30 days of last dose of study drug will be listed by patient with start and stop date of study medication, date of death, and cause of death. All deaths beyond 30 days after last dose of study drug will be displayed in a separate listing.

6.5.4 Clinical laboratory data

Descriptive statistics (number of observations, mean, standard deviation, minimum, median and maximum values) will be presented for clinical laboratory tests (hematology, clinical biochemistry and urinalysis), their changes from baseline (including baseline value), and their percent changes from baseline by treatment group at applicable visits.

Hematological and biochemical laboratory values will be graded based on NCI CTCAE version 4.03. CTCAE severity grading for laboratory abnormalities are based on applicable laboratory threshold values outlined in NCI CTCAE v4.03. It should be noted that in the present analysis of those laboratory parameters for which additional clinical information potentially can also influence the toxicity grade, this clinical information is in general not available and only the laboratory measurements are used for grading.

Any additional specific handling of the NCI CTCAE v4.03 toxicity grading assignments will be noted in the footnotes of the corresponding tables as applicable per the data collection in the study.

- In the event of overlapping CTCAE criteria ranges for specific lab tests, the algorithm assigns the worst grade
- If calcium type is not recorded (i.e. only “calcium” is recorded), then grading is done as if the calcium is total calcium. “Calcium corrected” is computed from total

calcium and serum albumin (if ≤ 4.0 g/dl) from the same time point based on CTCAE v3.0 guidance. If serum albumin (if ≤ 4.0 g/dl) from the same time point is not available or if "calcium, unspecified" was collected then grading is done as if the calcium is "corrected calcium."

- Results with special characters (such as ">" and "<") are not graded.

The frequency of laboratory abnormalities regarding hematology, coagulation panel, clinical chemistry, and urinalysis will be tabulated by treatment group. Worst grades for hematological and biochemical toxicities will be calculated according to CTCAE, version 4.03 based on laboratory measurements, and will be summarized by treatment group and NCI CTCAE v4.03 category and worst grade.

Clinical laboratory toxicities during treatment including a period of 30 days after last dose of treatment will be considered as "treatment-emergent.

The last non-missing value before or on the first day of study drug will be retained as "baseline" data. If several assessments are performed on the same day (without timing information) the average of the values will be considered.

Incidence tables (frequency and percentage of patients) as well as tables with change in NCI CTCAE v4.03 worst grade from baseline will be presented as following:

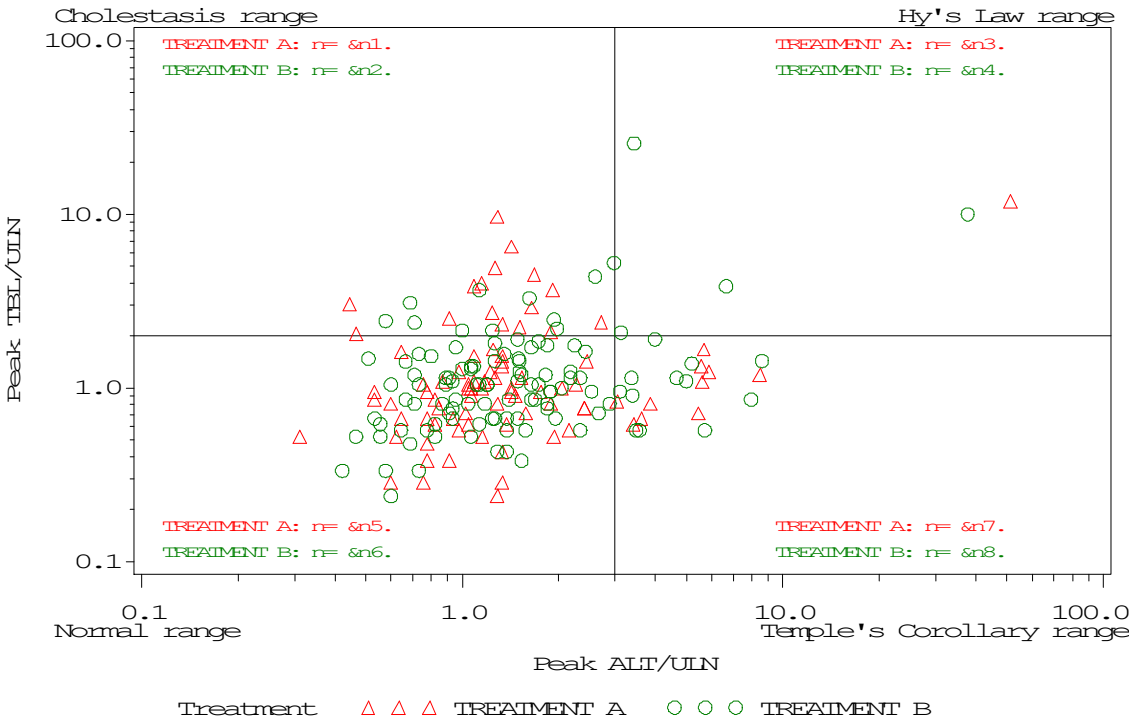
- Treatment-emergent hematological and biochemical toxicity.
- Treatment-emergent hematological and biochemical toxicities with incidence rate above 5% in any treatment group.
- Change in worst grade for hematological and biochemical toxicity from baseline.

The laboratory values will be also categorized into low, normal and high according to their reference ranges.

Descriptive statistics will be calculated by treatment group and time interval. 16-week time intervals will be used.

A listing will be provided for all patients possibly fulfilling Hy's Law criteria, i.e. patients with elevated AST and / or ALT > 3xULN, alkaline phosphatase < 2xULN and bilirubin ≥2xULN. For possible Hy's Law cases relevant laboratory data will also be displayed graphically within actual patient profiles (presenting total bilirubin, ALT, AST and ALP values in terms of ULN over time) by treatment group. Below example of Hy's law plot will be provided for peak total bilirubin vs ALT. If a patient has any total bilirubin ≥ 2xULN then peak bilirubin will be plotted versus the maximum ALT amongst the total bilirubin ≥ 2xULN. Otherwise, peak bilirubin will be plotted versus peak ALT.

Example for Hy's law plot:



Unscheduled laboratory data will be included in the descriptive tables.

6.5.5 12-Lead ECG, QTc

No analyses of ECG will be performed.

6.5.6 Other safety measures

For each treatment group, vital signs (i.e. blood pressure, heart rate, weight and BMI) will be tabulated and summarized by visit for observed values and changes from baseline using descriptive statistics, as appropriate. If more than one baseline assessment was collected, the most recent one will be used. If several assessments are performed on the same day (without timing information) the average of the values will be considered.

Outlier analyses will be conducted using the following limits:

- low systolic blood pressure: ≤ 90 mmHg and a decrease of ≥ 20 mmHg
- high systolic blood pressure: >190 mmHg and an increase of ≥ 20 mmHg
- low diastolic blood pressure: ≤ 50 mmHg and a decrease of ≥20 mmHg
- high diastolic blood pressure: > 105 mmHg and an increase of ≥ 20 mmHg

- low heart rate: < 50 bpm and a decrease of ≥ 15 bpm
- high heart rate: > 120 bpm and an increase of ≥ 15 bpm

The number and percentage of patients with outlying values will be tabulated by treatment group and time interval.

No subgroup analyses will be performed.

Unscheduled vital signs data will be included in the summary tables.

6.5.7 Physical examinations

The number and percent of patients with physical examination abnormalities will be summarized based on the safety analysis set for each treatment group and overall. Physical examination findings will be presented in a data listing.

No subgroup analyses will be performed.

7. Document history and changes in the planned statistical analysis:

SAP Version 4.2 dated 20 SEP 2018

Supplemental 4 SAP version 1.0 dated 04DEC2019

Supplemental 5 SAP version 1.0 dated 27JAN2020.

8. References

Refer to main SAP v4.2 dated 20 SEP 2018.

9. Appendices

Not applicable.

A multinational, randomised, double-blind, placebo-controlled, Phase III efficacy and safety study of darolutamide (ODM-201) in men with high-risk non-metastatic castration-resistant prostate cancer

Bayer study drug	BAY 1841788 / darolutamide		
Clinical study phase:	III	**Date:**	20 SEP 2018
Study No.:	17712	**Version:**	4.2
Author:	PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Table of Tables **3**

List of Abbreviations **5**

1. Introduction **7**

2. Study Objectives **7**

3. Study Design **8**
3.1 Determination of Sample Size 9

4. General Statistical Considerations **10**
4.1 General Principles 10
4.2 Handling of Dropouts 10
4.3 Handling of Missing Data 11
4.4 Interim Analyses and Data Monitoring 12
4.5 Data Rules 13
4.6 Blinded Review, Validity Findings, and Protocol Deviations 13
4.6.1 Blinded Review 14
4.6.2 Validity Findings 14
4.6.3 Protocol Deviations 14

5. Analysis Sets **14**

6. Statistical Methodology **15**
6.1 Population characteristics 15
6.1.1 Disposition of patients 15
6.1.2 Demographic and Baseline Characteristics 16
6.1.3 Medical history 18
6.1.4 Prior and concomitant medications 18
6.2 Efficacy 19
6.2.1 Primary Efficacy Analysis 19
6.2.1.1 Metastasis-free survival (MFS) 19
6.2.2 Analysis of secondary efficacy endpoints 28
6.2.2.1 Overall survival (OS) 31
6.2.2.2 Time to pain progression 33
6.2.2.3 Time to initiation of first cytotoxic chemotherapy 34
6.2.2.4 Time to first symptomatic skeletal event (SSE) 35
6.2.3 Tertiary or additional efficacy analysis 37
6.2.3.1 Progression-free Survival (PFS) 37
6.2.3.2 Time to first prostate cancer-related invasive procedures 39
6.2.3.3 Time to initiation of subsequent antineoplastic therapy 40
6.2.3.4 Time to PSA progression 41
6.2.3.5 Percent of patients with PSA response 42
6.2.3.6 Percent of patients with ECOG performance status deterioration 43
6.2.3.7 Time to ECOG performance status deterioration 43

6.2.3.8 Time to first opioid use for cancer pain 44
6.2.3.9 Health related Quality of Life and utility values. 44
6.2.3.9.1 Percent of patients with deterioration of FACT-P total score at 16 weeks 48
6.2.3.9.2 Time to deterioration in PCS subscale score 48
6.2.3.9.3 Percent of patients with improvement of EORTC-QLQ-PR25 urinary symptoms 49
6.2.3.9.4 Time to worsening of EORTC-QLQ-PR25 urinary symptom score 49
6.2.3.9.5 Percent of patients with deterioration of EQ-5D-3L utility score at 16 weeks 50
6.2.4 Subgroup analysis 50
6.3 Pharmacokinetics/pharmacodynamics 51
6.4 Biomarker evaluation 51
6.5 Safety 52
6.5.1 Extent of exposure 52
6.5.2 Adverse events 52
6.5.3 Deaths and Serious Adverse events 55
6.5.4 Clinical laboratory data 55
6.5.5 12-Lead ECG, QTc 57
6.5.6 Other safety measures 58
6.5.7 Physical examinations 59

7. Document history and changes in the planned statistical analysis: 59

8. References 60

9. Appendices 61
9.1 Study event schedule 61
9.2 Brief pain Inventory – Short Form (BPI-SF) questionnaire and scoring information 67
9.3 FACT-P questionnaire and scoring information 69
9.4 EORTC-QLQ-PR25 questionnaire and scoring information 77
9.5 EQ-5D-3L questionnaire and scoring information 80
9.6 Prohibited medication: metastasis-free survival 81
9.7 Cytotoxic chemotherapy: time to cytotoxic chemotherapy 82
9.8 Antineoplastic therapy: time to initiation of subsequent antineoplastic therapy 82
9.9 Statins medications 83
9.10 Opioid for cancer pain 83
9.11 Antiandrogen (AR inhibitors) therapy 83
9.12 GnRH agonist/antagonist therapy 83
9.13 Hormonal therapy 83
9.14 Bone sparing agent 84

Table of Tables

Table 1: Definition of primary and secondary analyses of MFS – handling of baseline metastasis 20
Table 2: Censoring Rules for MFS Based on independent blinded central reading Data 20

Table 3: Censoring Rules....24
Table 4: Censoring Rules....24
Table 5: Example for calculated α significance levels and standard normal thresholds (z) for OS....29
Table 6: Interim analysis for significance; number of OS events and standard normal thresholds are illustrating the example from Table 5....29
Table 7: Final Analysis for significance, scenario 1; number of OS events and standard normal thresholds are illustrating the example from Table 5....30
Table 8: Final analysis for significance, scenario 2; number of OS events and standard normal thresholds are illustrating the example from Table 5....30
Table 9: Censoring Rules for OS....31
Table 10: Censoring Rules for Pain Progression....33
Table 11: Censoring Rules for Cytotoxic Chemotherapy....34
Table 12: Censoring Rules for SSE....35
Table 13: Censoring Rules for SSE-free survival....36
Table 14: Censoring Rules for PFS Based on independent blinded central reading Data....37
Table 15: Censoring Rules for prostate cancer-related invasive procedure....39
Table 16: Censoring Rules for antineoplastic therapy....40
Table 17: Censoring Rules for PSA progression....42
Table 18: Censoring Rules for ECOG performance status deterioration....43
Table 19: Censoring Rules for Time to first opioid for cancer pain....44
Table 20: Censoring Rules for time to deterioration in PCS....48
Table 21: Censoring Rules for time to worsening of EORTC-QLQ-PR25 urinary symptoms 49

List of Abbreviations

AE	Adverse Event
ADT	Androgen Deprivation Therapy
ANCOVA	Analysis of covariance
AUC	Area under the curve
BMI	Body Mass Index
BPI-SF	Brief Pain Inventory – Short form
CYTOC	Cytotoxic Chemotherapy
CI	Confidence Interval
CRF	Case Report Form
DMC	Data Monitoring Committee
EBRT	External beam radiation therapy
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EORTC-QLQ-PR25	European Organization for Research and Treatment of Cancer quality of Life Questionnaire- Prostate Cancer Module
EQ-5D-3L	European Quality of Life 5 – Domain scale
EWB	Emotional well being
FACT-P	Functional assessment of cancer therapy Prostate
FAS	Full Analysis Set
FWB	Functional well being
HRQoL	Health Related Quality of Life
HLT	High level term
IRT	Interactive response technology
ITT	Intent-to-treat
IPE	Iterative Parametric Estimation
IVRS	Interactive voice response system
MedDRA	Medical Dictionary for Regulatory Activities
MFS	Metastasis-Free Survival
MLG	MedDRA Labelling Grouping
n	Number of non-missing values
N/A	Not Applicable
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
nmCRPC	Non-metastatic castration-resistant prostate cancer
OM	Operational manual
OS	Overall Survival
PCS	Prostate cancer subscale
PCWG2	Prostate Cancer Trials Working Group2
PDD	Protocol Deviations Document
PFS	Progression-free Survival

PK	Pharmacokinetics
PPS	Per-Protocol Set
PRO	Patient Reported Outcome
PSA	Prostate-specific antigen
PSADT	Prostate-specific antigen doubling time
PT	Preferred Term
PWB	Physical well being
QoL	Quality of Life
QTc	Corrected QT
QTcB	Corrected QT Bazaett's formula
QTcF	Corrected QT Fredericia's formula
RECIST	Response Evaluation Criteria in Solid Tumors
RPSFT	Rank-Preserving Structural Failure Time
SAE	Serious Adverse Event
SAC	Statistical Analysis Center
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SDG	Standardized Drug Grouping
SSE	Symptomatic skeletal event
SOC	System organ class
SWB	Social family well being
TEAE	Treatment-Emergent AE
TOI	Trial Outcome Index
US	United States
WHO-DD	World Health Organization Drug Dictionary

1. Introduction

Study 17712 (ARAMIS) is a multinational, randomized, double-blind, placebo-controlled, phase III efficacy and safety study of darolutamide in men with high-risk non-metastatic castration-resistant prostate cancer (nmCRPC).

This statistical analysis plan (SAP) describes the analyses and data presentations for final primary efficacy endpoint (metastasis-free survival) and the final overall survival analysis at the end of double-blind part when targeted number of primary endpoint events is collected.

The SAP contains definitions of analysis sets, derived variables, and statistical methods for the analysis of efficacy and safety.

This SAP version 4.2 is an updated of SAP version 4.1 dated 13SEP2018 which was an update of version 4.0 dated 10AUG2018 and version 3.0 dated 12MAR2018. SAP v3.0 is based on the integrated protocol version 4.0 (amendment 3), dated 26 FEB 2018, is an amendment of the abbreviated SAP version 2.1 dated 22 JUN2017, based on protocol version 3.0 (amendment2), dated 19 JUL 2016.

2. Study Objectives

The primary objective of the study is to demonstrate superiority of darolutamide over placebo in metastasis-free survival (MFS) in patients with high-risk non-metastatic castration-resistant prostate cancer (nmCRPC).

The secondary objectives of the study are:

- Overall survival (OS),
- Time to pain progression,
- Time to first symptomatic skeletal event (SSE),
- Time to initiation of first cytotoxic chemotherapy for prostate cancer,
- Characterize the safety and tolerability of darolutamide.

The additional objectives of the study are:

- Progression free survival (PFS),
- Time to first prostate cancer related invasive procedure,
- Time to initiation of first subsequent antineoplastic therapy,
- determine effect of darolutamide on prostate-specific antigen (PSA) progression and PSA response
- ECOG performance status deterioration

- Health-related quality of life (QoL)
- Evaluate pharmacokinetics (PK) of darolutamide and keto-darolutamide
- Explore possible relationships between exposure and safety and efficacy response.

3. Study Design

This is a randomized, phase III, multi-center, double-blind, placebo-controlled efficacy and safety study of oral darolutamide (600 mg twice a day [bid]) in patients with nmCRPC who are at high risk for developing metastatic disease.

Approximately 1500 patients will be randomized to receive darolutamide or placebo in a 2:1 ratio in a double-blind manner. Randomization will be stratified by prostate-specific antigen doubling time (PSADT) (≤ 6 vs. > 6 months) and use of osteoclast-targeted therapy (yes vs. no). Randomized patients will receive study treatment until confirmed metastasis or intolerable adverse event (AE).

The study has a design of 2 parallel groups. For each patient, the study may involve 1 to 2 variable length periods: a study treatment period (a double-blind part for darolutamide or placebo arms and an open-label darolutamide treatment phase) and a follow-up period.

The length of the periods for each patient will depend on the absence or presence of metastasis – after confirmed metastasis, the patient must be withdrawn from study treatment.

During the double-blind part, patients may be on study treatment (darolutamide or placebo arm) or at follow-up. After the double-blind part, patients may be on open-label study treatment if they are in the darolutamide arm or if they have received double-blind placebo treatment and have started open-label darolutamide treatment in case of positive study results. Once the study results are available, and if they support a positive benefit/risk assessment for darolutamide in the study by the judgment of the sponsor, also considering feedback from the study steering committee and/or health authorities, those patients who are on study treatment (darolutamide or placebo) will be offered the opportunity to receive darolutamide through open-label treatment in this study.

The double-blind treatment (darolutamide or placebo arm) is planned to be continued until the total number of events for the primary efficacy analysis (MFS analysis) had been reached (approximately 385 events). During this part the treatment code will remain blinded.

Randomization and blinding:

Randomization will be performed centrally blocking by center according to the design of the study using a 2-step procedure. Firstly, a separate master randomization schedule and study treatment package list will be created using randomly permuted blocks. Secondly, randomly permuted blocks from the master randomization schedule are assigned to the study centers. An interactive response technology (IRT) (also called interactive voice response system [IVRS]) system assigns patients to receive either darolutamide or matching placebo using allocation ratio 2:1, respectively.

The numbers are assigned to the unique patient number previously allocated by the investigator. The randomization will be stratified by:

- PSADT (≤ 6 months vs. > 6 months),
- Osteoclast-targeted therapy at randomization (yes vs. no).

Details of the randomization method will be given in the IRT user requirements.

The reference for this study will be placebo tablets that match the darolutamide tablets. During the double-blind part of the study at the time of radiographic metastasis patient's treatment assignment will remain blinded.

All patients, study personnel, and sponsor's personnel directly involved in the conduct of the study will be blinded to treatment assignments during the double-blind part of the study. After completing the double-blind part the study will be unblinded for the primary analyses.

Schedule of procedures:

Efficacy and safety measurements obtained during the course of the study are summarized in the schedule of assessments, see Appendix 9.1.

3.1 Determination of Sample Size

Several clinical publications were used as basis for the sample size calculation. The median MFS for placebo is based on denosumab phase III study results [8] at the time the ARAMIS protocol was designed, no data were available on the expected treatment effect size for androgen-receptor inhibitors in high-risk nmCRPC patients. Data published in 2016 and recently published data from SPARTAN and PROSPER phase III trials indicate, that the treatment effect of the class of novel androgen receptor inhibitors (enzalutamide, apalutamide) is stronger than initially assumed

Given the results from PROSPER and SPARTAN, the initial assumption for the treatment effect size of a hazard ratio of 0.75 was considered as too conservative.

All patients had to have an independent evaluation of radiological images for no baseline metastases at baseline to be eligible for randomization. However, for the central efficacy review, the baseline scans were again reviewed. During this review some patients were identified with metastases at baseline; to calculate sample size these patients are considered as having a MFS event on the date of randomization.
The number of events has been adjusted to account for the non-informative character of baseline metastases events and the dilution impact to the estimated MFS hazard ratio.

The assumptions used in determining the sample size for this study are as follows:

- Overall 2-sided type I error rate: 0.05,
- Assumed hazard ratio of 0.65,

- Accounting for 5% of the patients with baseline metastasis, diluted hazard ratio of 0.71,
- Randomization ratio: 2:1
- Statistical power at the final analysis: 91%,
- Median MFS for placebo: 25 months,

385 MFS events will provide approximately 91% power to detect a statistically significant difference in MFS times, with a two-sided log-rank test with a 0.05 level of significance. With 40 months accrual time and a dropout rate of 40%, the study will require approximately 1500 patients (1000 darolutamide patients, 500 placebo patients) to achieve the targeted 385 MFS events. The MFS analysis will be performed when the targeted number of approximately 385 MFS events has been observed.

This sample size calculation was performed with a simulation based algorithm.

4. General Statistical Considerations

4.1 General Principles

The statistical evaluation will be performed by using the software package SAS release 9.2 or higher (SAS Institute Inc., Cary, NC, USA).

All variables will be analyzed by descriptive statistical methods.

In general, continuous variables will be summarized using number of non-missing values (n), number of missing values, means, standard deviations, medians, maximum, minimum, and interquartile range.

Ordinal variables will be summarized using n, number of missing values, medians, maximum, minimum, and interquartile range.

Categorical variables will be summarized using n, number of missing values, and percentages.

Time-to-event variables will be summarized using Kaplan-Meier estimates.

Frequency tables will be generated for categorical data.

4.2 Handling of Dropouts

Patients withdrawn from study treatment will not be replaced. Refer to section 4.7 in the study protocol for withdrawal of patients from study.

All primary efficacy analyses are based on the Full Analysis Set (FAS) that comprises all randomized patients, including patients who withdraw regardless of the reason for withdrawal. The intent-to –treat (ITT) population set mentioned in the protocol is identical to

the FAS. See following chapters for more details on deriving efficacy endpoints in case of missing data.

4.3 Handling of Missing Data

All missing or partial data will be presented in the patient data listing as they are recorded on the Case Report Form (CRF).

When appropriate, the following rules will be implemented so as not to exclude patients from statistical analyses due to missing or incomplete data.

Missing or unevaluable tumor assessments "including scheduled assessments that were not done and incomplete assessments that did not result in an unambiguous occurrence of first metastasis or evaluation of local regional disease progression" will not be used in the calculation of derived efficacy variables related to tumor assessments unless a new lesion occurred or the lesions that were evaluated already showed metastasis. No imputation will be performed for unevaluable or missing tumor assessment. For example, if a patient misses a scan visit and metastasis is documented at the next available scan visit, the actual visit date of the first documented metastasis will be used to calculate MFS.

If a date is incomplete, (e.g. only year and month of date of tumor assessment or date of death is available), then day 15 of the month will be used for the calculation of, for example, MFS, OS.

If the actual scan date of the radiological metastasis is missing and radiological metastasis has been documented based on criteria specified in the protocol, the scheduled scan date will be used to calculate the time to metastasis.

If date for QoL questionnaire is not documented, then the corresponding visit date will be used.

Safety variables, Medical history, and Concomitant medications:

Treatment emergent AEs, treatment phases, period and relative days will be derived according to data management programming, the operational manual document and standard guidelines.

Missing Patient report Outcome (PRO) data

In case of missing responses for one or more items, subscale scores can be prorated (see scoring of FACT-P, EORTC-QLQ-PR25, Brief Pain Inventory – Short form (BPI-SF) and of EQ-5D-3L in Appendices 9.2 to 9.5).

For FACT-P this is done by multiplying the sum of the subscale by the number of items in the scale, then dividing the number of items actually answered. Prorating of scores is acceptable as long as more than 50% of the items are answered (assuming that the score of missing items are similar to those of non-missing items). If less than or equal to 50% of the items are answered for any domain, then the score of that domain is set to missing. The total score is then calculated as the sum of the un-weighted subscale scores. Moreover, the FACT-P total score is set to missing if the related overall item response rate is less than or equal to 80%.

For EQ-5D-3L, if there is a missing or ambiguous answer (i.e. marking of more than one answer) on the 5-dimension questions, then the index score is missing.

For BPI-SF 2 scores will be derived: 1/ for the pain severity score if one missing answer then scoring will be set to missing, 2/ for the pain interference score if 4 or more missing answers out of the seven questions then the score will be set to missing.

For EORTC-QLQ-PR25, six scales are created, if less than or equal to 50% of the items are answered for any scales, then the score of that scale is set to missing.

4.4 Interim Analyses and Data Monitoring

No formal interim analysis will be performed for primary endpoint MFS.

A formal interim analysis for secondary endpoints will be performed by the sponsor at the same time as the primary endpoint analysis. See more details in Section 6.2.

A Data Monitoring Committee (DMC) (also called Data and safety monitoring board (DSMB)) was instituted in order to ensure the ongoing safety of the patients. The operation of the DMC is guided by a DMC charter.

The DMC includes 3 independent disease experts (physicians) and one independent statistician. The DMC will operate independently of the Sponsor and Investigators. Data review meetings will be held periodically as per separate DMC charter. Enrollment into the study is to continue throughout the scheduled meetings of the DMC. The Statistical Analyst Center (SAC) is scheduled to receive the randomization list for their unblinding. Data which may compromise the integrity of the study (e.g., comparative data and/or any unblinded data) are to be analyzed and discussed only in the closed session of the DMC meetings. The closed session will be restricted to the DMC members and a non-voting facilitator (biostatistician from the SAC). Closed session minutes will be maintained by the DMC in confidence. All

data provided to the DMC and all deliberations of the DMC are to be considered privileged and confidential. DMC members signed confidentiality agreements.

4.5 Data Rules

Generally, for each date stored in the database, a set of organizational variables will be derived in order to describe the temporal context of that date in the specific study: phase of treatment (pre, during or post study treatment), day relative to the start of study treatment, day relative to the end of study treatment.

Additional contextual variables may be created in analysis datasets.

Refer to Section 4.3 for handling of missing data, as well as to Section 6, for specific endpoints data rules, for example censoring patients or definition of treatment-emergent adverse events (TEAEs).

All derived variables will be stored in analyses data sets (ADS) created by the statistician and / or the statistical analyst.

Since differences may occur between the values of the stratification variables entered by the investigator at the time of randomization (IVRS) and those collected on the CRF, the analysis will be performed using both assignments to the strata. However, the primary stratified analyses for the efficacy endpoints will be based on the information collected in the IVRS whereas stratification information entered in the CRF (considered the 'true' information) will be used for sensitivity analyses. In rare circumstances, if CRF data cannot be documented, then IVRS data will be used.

Baseline data for stratification will be taken from the last non-missing observation on or before the randomization date.

Baseline data for demographics and characteristics, PRO questionnaires and safety will be taken from the last non-missing observation on or before the first day of study drug intake.

Reported outcome questionnaires dates of completion were not collected therefore the dates of corresponding visit will be used for analyses.

Unscheduled visits will be considered in analyses, excepted for analyses of health-related QoL.

A data cut-off date for the final and the interim efficacy analyses will be chosen.

Time to event variables values will be displayed in months. Months will be calculated dividing days by 30.44.

4.6 Blinded Review, Validity Findings, and Protocol Deviations

Protocol deviations and validity review will be performed as described in the sponsor's operational instruction on conducting blind review meetings, cleaning, and reviewing study

data, in addition to detailed guidance on important deviations and validity findings. Any unscheduled assessments prior to randomization will be included in the determination of the baseline value.

4.6.1 Blinded Review

Blinded review will be conducted prior to database lock for each analysis, as described in the sponsor's operational instruction on conducting blinded review meetings. Reviewers will be tasked with evaluating the overall quality and reliability of the study data and its suitability for the planned statistical analysis. Blinded review will also assess protocol deviations and validity findings.

4.6.2 Validity Findings

Criteria for validity, i.e. eligibility for each analysis set, are described in Section 5. Details may be further described in a separate document. Validity findings will be obtained as described in both the sponsor's operational instruction on cleaning and reviewing study data, and the sponsor's detailed guidance on important deviations and validity findings, and includes major protocol deviations as described in Section 4.6.3 and the separate Protocol Deviations Document (PDD). The blinded review will review and finalize the validity findings prior to the database lock for each analysis.

4.6.3 Protocol Deviations

Protocol deviations will be classified as major or important. Major protocol deviations are deviations affecting analysis set eligibility or treatment group assignment and constitute validity findings as described in Section 4.6.2.

Major protocol deviations will be identified based on the criteria described in the sponsor's operational instruction on cleaning and viewing study data. Assessment criteria, deviation definitions, methods of evaluation, and further details will be described in a separate PDD.

5. Analysis Sets

All patients who were randomized are included in the **Full Analysis Set (FAS)**. Following the intent-to-treat (ITT) principle, the patients in this set will be grouped according to the treatment they were allocated to receive at randomization, irrespective of actual treatment. The FAS is the primary analysis set for the efficacy analyses.

In contrast to wording from the protocol, a Per Protocol data Set (PPS) will not be specified and any per protocol analysis will not be performed as its additional value is minor.

All patients who were randomized and received at least one dose of study treatment will be included in the **Safety Analysis Set (SAF)**. This set of patients will be grouped for analysis according to the treatment they actually received, as opposed to the treatment they were allocated to receive at randomization. Only patients who solely received placebo treatment will be analyzed under the placebo arm. Patients who received at least one dose of

darolutamide treatment will be analyzed under the darolutamide arm. The SAF will be used for all safety analyses.

6. Statistical Methodology

The formal statistical analyses will be both descriptive and inferential. Summaries will be provided for each of the treatment groups, darolutamide, and placebo. In addition, descriptive summaries of population characteristics will be provided for the total study population.

Stratification factors are:

- PSADT (≤ 6 months vs. > 6 months)
- use of osteoclast-targeted therapy (yes vs. no)

CRF baseline data are closely monitored and corrected if necessary therefore IVRS stratification variable data may differ from CRF data in some instances. For the purpose of stratified analyses per this SAP, IVRS data will be used. For the primary endpoint MFS, a sensitivity analysis using CRF stratification data will be created. For other purposes, like analysis of baseline patient information, CRF data will be used.

Patient's characteristics tables will be displayed by the subgroups:

Geographical region (North America, Asia Pacific, and Rest of the World (ROW)).

North America region is defined by countries:

- Canada, United States.

Asia Pacific region is defined by countries/regions:

- Japan, South Korea, and Taiwan.

ROW region is defined as all countries that are neither included in the North America region nor in the Asia Pacific region.

6.1 Population characteristics

6.1.1 Disposition of patients

The number of patients enrolled and included in each population will be tabulated by region, country, and center. A summary table will also be presented for the number of patients enrolled and the number and percentage of patients in each of the defined populations. Re-screened patients will only be counted once.

The reasons for patients excluded from each of the patient populations will also be tabulated. The reasons for discontinuation of study treatment will be tabulated.

The number of patients who discontinued study treatment due to PSA increased without documented metastasis per independent central reading will be displayed.

6.1.2 Demographic and Baseline Characteristics

Descriptive summaries of demographics and baseline characteristics will be presented by treatment group and overall for the FAS and SAF populations. Comparability of the treatment groups with respect to demographics and baseline characteristics will be assessed using descriptive summaries.

The following demographic data will be summarized:

- Age (years), calculated at the date of randomization using date of birth
- Age (years) (<65, 65 - 74, 75 - 84, ≥85)
- Sex (Male)
- Race (White, Asian, Black or African American, Other)
- Ethnicity (Hispanic or Latino)
- Height (cm)
- Weight (kg)
- Body Mass Index (BMI) (kg/cm^2) (< 20, 20 to < 25, 25 to < 30, ≥ 30)
- Geographical region (North America, Asia Pacific, ROW).

The following baseline characteristics will be summarized:

- PSADT (≤6, >6 months) - IVRS
- PSADT (≤6, >6 months) - CRF
- PSADT (months)
- Osteoclast-Targeted Therapy (yes, no) - IVRS
- Osteoclast-Targeted Therapy (yes, no) - CRF
- ECOG Performance status
- Gleason Score
- Primary tumor classification
- Regional lymph node classification
- Time since start of castration-resistant (months)
- Time since initial diagnosis (capture under clinical staging) (months)

- Time from start of first prior GnRH agonist/antagonist to start of study-drug (months) (for selection, see 9.12)
- Time from prior orchiectomy to start of study-drug (months)
- Time from start of first prior antiandrogen (AR inhibitors) to start of study-drug (months) (for selection, see 9.11)
- Time from the first prior ADT to start of study drug (months) (ADT is defined by GnRH agonist/antagonist, orchiectomy, antiandrogen (AR inhibitors))
- Renal function - eGFR at baseline:
    - Normal: eGFR ≥ 90 mL/min
    - Mildly impaired: 60 ≤ eGFR < 90 mL/min
    - Moderately impaired: 30 ≤ eGFR < 60 mL/min
    - Severely impaired: 15 ≤ eGFR < 30 mL/min
    - End stage renal disease: eGFR < 15 mL/min and not on dialysis, or requiring dialysis

  eGFR will be calculated according to the Modification of Diet in Renal Disease (MDRD) Formula:

  $$eGFR\ (mL/min/1.73m^2) = 186 \times SCR^{-1.154} \times age^{-0.203} \times (0.742, \text{if female}) \times (1.212 \text{ if blacks or African American}) \times (0.881 \text{ if Japanese}) \times (1.227 \text{ for Chinese (mainland China, Taiwan and Hongkong)}$$

  where SCR= serum creatinine measured in mg/dL
- Hepatic impairment at baseline:
    - Normal: Total bilirubin and AST ≤ upper limit of normal (ULN)
    - Mild hepatic impairment: Total bilirubin > ULN to 1.5 x ULN or (Total bilirubin ≤ ULN and AST > ULN)
    - Moderate impairment: Total bilirubin > 1.5 to 3 x ULN, any AST
    - Severe impairment: Total bilirubin > 3 x ULN, any AST.
- Baseline PSA from central laboratory (≤ 10, >10 to ≤20, >20 ng/mL)
- Baseline hemoglobin (g/dL)
- Baseline lactate dehydrogenase (U/L)
- Baseline alkaline phosphatase (U/L)
- Baseline testosterone (nmol/L)
- Prior hormonal therapies (1, ≥2) (for definition of hormonal therapies, refers to Section 9.13).

In case date is not completed the following rules will be applied for calculation:

- -if only year is documented, 1st January will be applied,
- -if month and year are available, day 1of the month will be used.

Demographic and baseline characteristics tables will be provided for the subgroups:

- patients with presence of baseline metastasis and absence of baseline metastasis, per independent blinded central reading for efficacy assessment,
- patients from United States (US) and non-US countries.

6.1.3 Medical history

Medical history will be summarized by Medical Dictionary for Regulatory Activities (MedDRA) version 20.0 or the most recent version terminology.

Summary statistics (frequency and percentage) will be provided by body system organ class (SOC), high level term (HLT) and preferred term (PT) for the FAS population, by treatment group.

Past and present finding will be displayed.

6.1.4 Prior and concomitant medications

Concomitant medications and procedures are:

- Prior prostate cancer therapy: frequency of patients for each drug category
- Radiotherapy (prior)
- Opioid treatment (prior, concurrent, and follow-up): frequency of patients – patient listings
- Diagnostic and therapeutic procedures: frequency of patients by procedure (prior, concurrent)
- Concomitant medications: frequency of patients for each drug category
- Concomitant statin medications: frequency of patients
- Diagnostic and therapeutic procedures related to AE: frequency of patients by procedure (concurrent)

The dictionary used for coding medications is the WHO Drug dictionary.

In addition to the FAS population, concomitant medication will also be summarized on the SAF population.

For selection of statins medications, please refer to Appendix 9.9.

6.2 Efficacy

After the end of randomization of patients into the study, it will be reviewed if all stratification factors are represented by a sufficient number of patients. In case at least one of the stratification factors does not have sufficient number of patients (less than 3%) and the inferential analyses would be negatively impacted, the stratification factor will not be included as a stratification factor in this analysis.

Efficacy analyses will be performed in the FAS population.

6.2.1 Primary Efficacy Analysis

The primary efficacy variable is MFS, as determined by the independent blinded central reading.

6.2.1.1 Metastasis-free survival (MFS)

MFS is defined as time from randomization to confirmed evidence of metastasis or death from any cause, whichever occurs first. The MFS analysis will be performed when approximately 385 events are observed. Patients not experiencing death or metastasis will be censored at the last tumor assessment.

Survival distribution function will be used as basis for statistical hypothesis. Hypothesis will be two sided, although superiority over placebo is anticipated. The two-sided hypothesis is formulated as follows:

H_0: $S_{DAROLUTAMIDE}(t) = S_{PBO}(t)$, for all $t > 0$ and

H_1: $S_{DAROLUTAMIDE}(t) \neq S_{PBO}(t)$, for some $t > 0$,

Where S(t) represents estimated survival distribution at time t for MFS.

Metastasis in bone is defined as appearance of 1 or more lesions that are confirmed by the central reading according to the one of the methods described below. If the central reading identifies changes on bone scan, confirmatory anatomic imagine CT/MRI or x-ray, of the area in question needs to be obtained. Anatomic imaging performed up to 2 weeks prior to bone scan or later may be used as a confirmatory scan. Appearance of bone metastasis is assigned to the date of the bone scan at which the lesion was first identified.
Metastasis in non-osseous tissue is defined as new distant pathologic lymph nodes (M1a) or other pathological lesion (M1c) according to Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. New or progressive regional pathologic lymph nodes will not be defined as metastasis.

As part of the blinded central imaging review for the efficacy assessment to determine distant metastases, the baseline scans, which were classified as metastasis free by the eligibility read,

were reviewed again, together with all later scans. During this review some patients were classified with metastases already at baseline.

CCI [redacted]

CCI [redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Censoring Rules

Time of MFS is calculated as time to MFS = End Date-Date of randomization +1.

Table 2: Censoring Rules for MFS Based on independent blinded central reading Data

Situation	End Date on or prior to cut-off date	Censored
Documented metastasis	Date of documented metastasis	No
Documented metastasis at baseline	Date of randomization	No /CCI [redacted]

Situation	End Date on or prior to cut-off date	Censored
		CCI)
Documented metastasis after two or more consecutively missed tumor assessments, i.e. metastasis later than last evaluable scan + (32 +1) weeks	Date of last tumor assessment that the patient was known to be metastasis-free	Yes
Death before documented metastasis and not later than last evaluable scan + (32+1) weeks	Date of death	No
Death before documented metastasis and after two or more consecutively missed tumor assessments, i.e. death later than last evaluable scan + (32+1) weeks	Date of last tumor assessment that the patient was known to be metastasis-free	Yes
Discontinued the study before any post-baseline tumor assessments	Date of randomization	Yes
Discontinued the study before any post-baseline tumor assessments and died within (32+1) weeks after randomization	Date of death	No
Discontinued the study before any post-baseline tumor assessments and died later than (32+1) weeks after randomization	Date of randomization	Yes
Discontinued the study, but no documented metastasis	Date of last tumor assessment before discontinuation	Yes
Prohibited new anticancer treatment started prior to documented metastasis	Date of last tumor assessment before start of prohibited new treatment	Yes
Patients still on treatment without documented metastasis as of data cut-off	Date of last tumor assessment	Yes

MFS will be compared between the treatment arms using a two-sided stratified log-rank test, stratified by PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization. A two-sided overall alpha of 0.05 will be used for the efficacy analysis of MFS.

The analysis will be performed according to treatment groups as randomized, with stratification as recorded in the IVRS data.

For the baseline metastasis non-censored primary analysis, this endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'MFS' (primary analysis) and 'SAMFSBAS' (secondary analysis).

For the baseline metastasis censored primary analysis, this endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'MFS' (secondary analysis) and 'SAMFSBAS' (primary analysis).

Sample code to produce stratified log rank test:

```
proc lifetest data=data;
time prog_time*cnsr(1);
strata psa_double ot_therapy / group=treat test=(logrank);
run;
```

Where:

prog_time contains metastasis free survival time for the patient, *cnsr* contains a value of one for censored records and zero otherwise, *psa_double* contains PSA doubling time category at randomization, *ot_therapy* contains category for osteoclast-targeted therapy use at randomization, *treat* contains assigned treatment group.

The hazard ratio as well as its 95% confidence interval (CI) will be presented based on fitting a Cox regression model with treatment as factor and stratified by PSA doubling time (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

Sample code to produce hazard ratio:

```
proc phreg data=data;
model prog_time *cnsr(1) = trtpn;
strata psa_double ot_therapy / missing;
run;
```

Where:

prog_time contains metastasis free survival time for the patient, *cnsr* contains a value of one for censored records and zero otherwise, *psa_double* contains PSA doubling time category at randomization, *ot_therapy* contains category for osteoclast-targeted therapy use at randomization, *treat* contains assigned treatment group.

The Kaplan-Meier estimate of the median survival time and first and third quartiles will be presented with 95% CI. The survival curves will also be plotted. 4, 8, 12, months MFS rate time-points will also be provided with 95% CI. Greenwood's formula will be used for the standard error of the Kaplan-Meier estimate in the calculation of these confidence limits. The number and percentage of patients in each treatment group who have progressed or died at 4, 8, 12 months will be summarized along with the number and percentage of censored patients.

Sample code to produce Kaplan Meier curves:

proc lifetest data=data method=km outsurv=outsurv;

*time prog_time *cnsr(1);*

strata treat;

run;

Where:

Prog_time contains metastasis free survival time for the patient, *cnsr* contains a value of one for censored records and zero otherwise, *treat* contains assigned treatment group.

If the complete date of the scan is not known, but the year and month are available, Day 15 of the month will be used for the calculation.

If the complete date of the death is not known, but the year and month are available, Day 15 of the month will be used for the calculation.

Per protocol, following medications are considered prohibited:

- Radiopharmaceuticals,
- Immunotherapy (e.g. sipuleul T),
- Cytotoxic chemotherapy and any other systemic antineoplastic therapy,
- Enzalutamide, ARN-509, bicalutamide, flutamide, nilutamide,
- Cyproterone acetate estrogen,
- 5 α-reductase inhibitor,
- Abiraterone acetate, TAK-700 or other CYP17 inhibitors,
- Systemic ketoconazole (as antineoplastic therapy),
- Osteoclast-targeted therapy such as bisphosphonate or denosumab given for preventing skeletal-related events. These drugs are allowed for treatment of osteoporosis,
- Continuous use of systemic corticosteroid.

Selection and list of these prohibited treatments are available in Appendix 9.6.

Description of the MFS events will be provided, ie number of metastasis with location (distant lymph nodes, bone lesions, and/or visceral and/or other soft tissue) and number of death.

Sensitivity analyses:

Sensitivity analysis 1: Analysis of time to metastasis will be conducted using only the appearance of documented metastases as event. In particular, deaths are not counted as events. Patients who died before documented metastasis will be censored. See Table 3 below with this change.

Table 3: Censoring Rules

Situation	End Date on or prior to cut-off date	Censored
Documented metastasis	Date of documented metastasis	No
Documented metastasis at baseline	Date of randomization	No [redacted]
Documented metastasis after two or more consecutively missed tumor assessments, i.e. metastasis later than last evaluable scan + (32 +1) weeks	Date of last tumor assessment that the patient was known to be metastasis-free	Yes
Death before documented metastasis	Date of last tumor assessment that the patient was known to be metastasis-free	Yes
Discontinued the study before any post-baseline tumor assessment	Date of randomization	Yes
Discontinued the study with at least one post-baseline tumor assessment and no documented metastasis	Date of last tumor assessment before discontinuation	Yes
Prohibited new anticancer treatment started prior to documented metastasis	Date of last tumor assessment before start of prohibited new treatment	Yes
Patients still on treatment without documented metastasis as of data cut-off	Date of last tumor assessment	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'SAMFS1' (baseline metastasis non-censored) and 'SAMFS2' (baseline metastasis censored).

Sensitivity analysis 2: Analysis of MFS will be conducted with all prohibited new treatment started prior to documented metastasis considered as an event. See Table 4 below with this change.

Table 4: Censoring Rules

Situation	End Date on or prior to cut-off date	Censored

Situation	End Date on or prior to cut-off date	Censored
Documented metastasis	Date of documented metastasis	No
Documented metastasis at baseline	Date of randomization	No [redacted]
Documented metastasis after two or more consecutively missed tumor assessments, i.e. metastasis later than last evaluable scan + (32 +1) weeks	Date of last tumor assessment that the patient was known to be metastasis-free	Yes
Death before documented metastasis and not later than last evaluable scan + (32+1) weeks	Date of death	No
Death before documented metastasis and after two or more consecutively missed tumor assessments, i.e. death later than last evaluable scan + (32+1) weeks	Date of last tumor assessment that the patient was known to be metastasis-free	Yes
Discontinued the study before the post-baseline tumor assessments	Date of randomization	Yes
Discontinued the study before the post-baseline tumor assessments but died later than (32+1) weeks after randomization	Date of randomization	Yes
Discontinued the study before the post-baseline tumor assessments but died within (32+1) weeks after randomization	Date of death	No
Discontinued the study, but no documented metastasis	Date of last tumor assessment before discontinuation	Yes
Prohibited new anticancer treatment started prior to documented metastasis	Date of start of prohibited new treatment	No
Patients still on treatment without documented metastasis as of data cut-off	Date of last tumor assessment	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'SAMFSPT1' (baseline metastasis non-censored) and 'SAMFSPT2' (baseline metastasis censored).

If the complete date of the prohibited treatment is not known, but the year and month are available, Day 15 of the month will be used for the calculation.

Sensitivity analysis 3: similar analysis to the one created for the primary analysis for MFS with a two-sided model stratified by the stratification data from CRF. This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'MFS' (baseline metastasis non-censored) and 'SAMFSBAS' (baseline metastasis censored).

Sensitivity analysis 4: similar analysis to the one created for the primary analysis for MFS with a two-sided log rank test without including stratification factors in the model. This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'MFS' (baseline metastasis non-censored) and 'SAMFSBAS' (baseline metastasis censored).

Sensitivity analysis 5: similar analysis to the one created for the primary analysis for MFS, but using investigator assessment of radiological imaging. This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'SAMFSIA1' (baseline metastasis non-censored) and 'SAMFSIA2' (baseline metastasis censored).

Sensitivity analysis 6: similar analysis to the one created for the primary analysis for MFS with all deaths independent when it occurred considered as an event. This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'SAMFSDH1' (baseline metastasis non-censored) and 'SAMFSDH2' (baseline metastasis censored).

Sensitivity analysis 7: similar analysis to the one created for the primary analysis for MFS with event at date of first post baseline scan with metastasis instead of event at randomization for patients with baseline metastasis. If no metastasis documented in post baseline scans such patient will be censored at last available scan date. In case such a patient did not have any post baseline scan, the patient will be censored at randomization. This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to SAMFSPB1' (baseline metastasis non-censored) and 'SAMFSPB2' (baseline metastasis censored).

Discordance of metastasis between independent blinded central reading and investigator assessment will be displayed.

Summary of baseline metastasis per independent blinded central reading and per charter version will be created.

Summary of eligibility read and efficacy read per charter version will be displayed.

Summary of adjudication reading will also be displayed.

The following flags will be available in the central imaging related datasets providing more detailed information on some specificities of the central reading process and versions of the imaging charter:

- Flags for charter version (flag_ch1, flag_ch2, flag_ch3),
- Flags for adjudication (flag_ad1, flag_ad2, flag_ad3),
- Flags for discordances between reviewers sets: eligibility, efficacy and confirmation of metastasis (flag_dc, flag_dc1, flag_dc2),
- Flags for change of metastasis determination after final read (flag_fb1, flag_fb2).

Description	Value	Flag name
Charter version flag for eligibility (Charter cut-off 01 FEB 2017)	• v1.0 • v4.0	flag_ch1
Eligibility review adjudication 0 = no adjudication 1 = adjudication	• 0 • 1	flag_ad1
Charter version flag for confirmation of metastasis review	• v1.0	flag_ch2
Confirmation of Metastasis adjudication 0 = no adjudication 1 = adjudication	• 0 • 1	flag_ad2
Charter version flag for efficacy	v4.0	flag_ch3
Independent Efficacy Review Adjudication 0 = no adjudication 1 = adjudication	• 0 • 1	flag_ad3
Change of baseline response per Independent Efficacy Review triggered at final visit evaluation 0 = no change 1 = change from no metastasis to metastasis 2 = change from metastasis to no metastasis	• 0 • 1 • 2	flag_fb1
Change of metastasis progression per Independent Efficacy Review triggered at final visit evaluation 0 = no change 1 = change	• 0 • 1	flag_fb2
Metastasis at baseline per efficacy review 0 = no metastasis 1 = metastasis	• 0 • 1	flag_dc
Discordance of metastasis at baseline (Eligibility Review vs. Independent Efficacy Review) 0 = same result 1 = different results	• 0 • 1	flag_dc1
Comparison of confirmation of metastasis under Charter v1.0 where metastasis confirmed to the same time point as efficacy. 0 = same result 1 = different results	• 0 • 1	flag_dc2
Initial date of metastasis per before final visit evaluation (can be missing if no metastasis seen)	DDM MMY YYY	IMETSDT
Final date of metastasis per independent efficacy review before final visit evaluation (can be missing if no metastasis seen)	DDM MMY YYY	FMETSDT

6.2.2 Analysis of secondary efficacy endpoints

Secondary endpoints will be tested with the hierarchical gatekeeping procedure. The order of secondary endpoints is

- (1) Overall survival (OS),
- (2) Time to pain progression (PP),
- (3) Time to cytotoxic chemotherapy (CYTOC),
- (4) Time to first symptomatic skeletal event (SSE).

Secondary endpoints are tested only if the primary endpoint MFS is significant. They will be tested according to the sequence given above. If significance of the primary analysis of MFS is considered unmet, each of the secondary objectives will be considered unmet. The same overall two-sided significance level of 0.05 as used for the primary endpoint will be used for the secondary endpoints.

All secondary endpoints OS, PP, CYTOC and SSE will be tested sequentially two times, in case the previous endpoint in the hierarchical order was significant. The first test for statistical significance will occur at the time of the MFS analysis and the final test for statistical significance will occur when approximately 240 OS events have been observed.

A rho-family spending function with parameter rho=10 will be used for OS, time to pain progression, time to cytotoxic chemotherapy and time to first SSE to determine the two stopping boundaries for efficacy at the interim and final analyses. The OS information fraction will be used for determining the alpha spending and significance threshold also for the other secondary endpoints at the interim analysis. For the final analysis the remaining alpha available and the actually observed numbers of events at interim and final for each endpoint will be used to determine the significance thresholds of secondary endpoints.

This rho spending function is slightly more conservative at the first test compared to the O'Brien-Fleming spending function and assigns slightly more alpha at the second test compared to the O'Brien-Fleming spending function. This choice is motivated by the low likelihood to achieve a significant result for the hierarchically first secondary endpoint OS at the first test for both spending functions.

The significance boundaries will be calculated with the EAST software package based on number of events actually observed in the analyzed data sets.

As an example for the planned sequential testing of the secondary endpoint OS, the table shows calculated alpha significance levels (α) and standard normal significance thresholds (z), assuming numbers of OS events of 140 at interim and 240 at final analysis. See Table 5 for details.

Table 5: Example for calculated α significance levels and standard normal thresholds (z) for OS

	O'Brien-Fleming		Rho family, rho=10	
CCI [redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		[redacted]	
[redacted]		[redacted]		[redacted]

The schema of testing the secondary endpoints is illustrated in Table 6 to Table 8.

Table 6: Interim analysis for significance; number of OS events and standard normal thresholds are illustrating the example from Table 5

	140 OS events
	Interim analysis
OS interim dataset	test at z1
	if positive ⇩
PP interim dataset	test at z1
	if positive ⇩
CYTOC interim dataset	test at z1
	if positive ⇩
SSE interim dataset	test at z1

Table 7: Final Analysis for significance, scenario 1; number of OS events and standard normal thresholds are illustrating the example from Table 5

Assuming OS interim analysis was not positive:

		240 OS events
		Final analysis
OS final dataset	OS interim not positive →	test at z2
		if positive ↓
PP interim dataset*		test at z2
		if positive ↓
CYTOC final dataset		test at z2
		if positive ↓
SSE final dataset		test at z2

*PP is a more subjective endpoint, therefore an analysis of PP which includes a large portion of data collected under open-label conditions is not considered meaningful. In case PP would be tested in the final analysis, the interim dataset would be used also for this test procedure.

Table 8: Final analysis for significance, scenario 2; number of OS events and standard normal thresholds are illustrating the example from Table 5

Assuming OS interim analysis was positive and PP interim analysis was not positive:

		240 OS events
		final analysis
OS final dataset	OS interim was positive	no test
PP interim dataset*	PPR interim not positive →	test at z2
		if positive ↓

CYTOC final dataset		test at z2
		if positive 
SSE final dataset		test at z2

In further scenarios where in the interim analysis after OS also PP was positive or where in the interim analysis after OS and after PP also CYTOC was positive, accordingly modified testing sequences would be applied.

The above testing schema is following the testing strategy 2 which is discussed in the publication of Hung et al. [10].

6.2.2.1 Overall survival (OS)

OS is defined as time from randomization to death due to any cause. OS of patients not known to have died will be censored at their last date of being known to be alive or at the database cutoff date, whichever comes first. OS will be analyzed on the FAS population.

If a patient died and the date of death is completely or partially missing, but there is an AE with the outcome as 'Death', the date of death will be replaced by the end date of the AE.

If the complete date of death is not known, but the year and month are available and there is no AE with outcome as 'Death', Day 15 of the month will be used for the calculation of the time to death.

Table 9 will be used for calculating the OS; time to death = End Date – Date of Randomization+1.

Table 9: Censoring Rules for OS

Situation	**End Date on or prior to cut-off date**	**Censored**
Death during study	Date of death	No
Patient still alive at data cut-off	Date of data cut-off	Yes
Patient lost to follow-up before data cut-off	Date last known to be alive	Yes
Patient lost to follow-up without contact after randomization	Date of randomization	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'DEATH'.

The Last Known Alive Date (LKAD) is derived from the main data sources. The last available date across all selected data panels listed below will be picked as the LKAD by patient. Information from selected data, i.e. visit dates, exposure information, demographics, laboratory measurements, tumor assessment dates, MFS dates, SSE dates, survival status date, and disposition events or follow up assessments will be used to determine survival status. Within all the dates from the selected data panels, identify the latest available date as the LKAD for each patient.

A log rank test stratified by stratification factor of PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization (i.e. from IVRS) will be used to compare the darolutamide treated and placebo groups.

The hazard ratio as well as its 95% CI will be presented based on fitting a Cox regression model with treatment as factor and stratified by PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm (including display of numbers at risk at various time points). The Kaplan-Meier estimate of the median survival time and first and third quartiles will be presented with 95% CI. Greenwood's formula will be used for the standard error of the Kaplan-Meier estimate in the calculation of these confidence limits.

The number and percentage of patients in each treatment group who have died will be summarized at yearly time points along with the number and percentage of censored patients. Summaries by cause of death will be provided too.

Descriptive statistics of survival follow-up time will be calculated by treatment arm and total. Follow-up time (days) is the time to death as described for the OS endpoint for censored and uncensored patients.

Once the study results are available, and if they support a positive benefit/risk assessment for darolutamide in the study by judgment of the Sponsor (considering feedback from the study steering committee and/or health authorities), those patients who are on study treatment (darolutamide or placebo) will be offered the opportunity to receive darolutamide through open-label treatment in this study. This may dilute the effect size for the secondary endpoints as the placebo arm outcomes will become more similar to the darolutamide arm outcomes. The amount of this bias can be approximately assessed by established crossover adjustment methods like Rank-Preserving Structural Failure Time (RPSFT) [7] or Iterative Parametric Estimation (IPE) [5]. These methods construct a placebo arm Kaplan-Meier curve which is expected to resemble the Kaplan-Meier (KM) curve which would be observed if placebo to darolutamide crossover would not had occurred. The data of the second analysis of OS will be analyzed in addition with the RPSFT and IPE methods. Only the unadjusted OS data will be used for testing for a statistically significant treatment effect.

6.2.2.2 Time to pain progression

Time to pain progression (PP) is defined as time from randomization to pain progression, where progression is defined as an increase of 2 or more points from baseline in question 3 of the Brief Pain Inventory-Short Form questionnaire (BPI-SF) related to the worst pain in the last 24 hours taken as a 7-day average for post-baseline scores, or initiation of short or long-acting opioids for pain, whichever comes first. Initiation or change in the use of other non-opioid analgesics is not used in the analysis of pain progression.

List of opioids for cancer pain is available in Appendix 9.10.

Questionnaire BPI-SF is detailed in Appendix 9.2.

A minimum number of 4 completed daily reports at each reporting time point out of the 7 days required per protocol are needed to consider valid the assessment.

If the complete date of the assessment is not known, but the year and month are available, day 15 of the month will be used for the calculation.

Table 10 will be used for calculating the time to first pain progression. Time to Pain progression = End Date – Date of Randomization +1.

Table 10: Censoring Rules for Pain Progression

Situation	End Date on or prior to cut-off date	Censored
Recorded pain progression during the study	Date of the first assessment that qualified as pain progression	No
No baseline BPI-SF assessments	Date of randomization	Yes
Discontinued the study before the post-baseline BPI-SF assessments	Date of randomization	Yes
Death during the study before pain progression	Date of last visit the patient was known not to have progressed or randomization date whatever comes later	Yes
Patient has no recorded pain progression at data cut-off	Date of last visit the patient was known not to have progressed or randomization date whatever comes later	Yes
Patient lost to follow-up before data cut-off	Date of last visit the patient was known not to have progressed or randomization date whichever comes later	Yes
Patient taking opioids for any reason	Date of randomization	Yes

within 4 weeks prior to randomization		

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'PP'.

A log rank test stratified by stratification factor of PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization will be used to compare the darolutamide treated and placebo groups.

The hazard ratio as well as its 95% CI will be presented based on fitting a Cox regression model with treatment as factor and stratified by PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm. The Kaplan-Meier estimate of the median survival time and first and third quartiles will be presented with 95% CI. Greenwood's formula will be used for the standard error of the Kaplan-Meier estimate in the calculation of these confidence limits.

6.2.2.3 Time to initiation of first cytotoxic chemotherapy

The time to cytotoxic chemotherapy (CYTOC) is defined as time from randomization to the start of the first cytotoxic chemotherapy cycle. Patients who have not taken cytotoxic chemotherapy will be right censored at their last visit.

Cytotoxic chemotherapy drugs are defined with ATC code L01A, L01B, L01C, L01D and L01X. The list of all cytotoxic chemotherapy is available under Appendix 9.7.

If the complete date of the cytotoxic chemotherapy is not known, but the year and month are available, day 15 of the month will be used for the calculation.

Table 11 will be used for calculating the time to first cytotoxic chemotherapy. Time to cytotoxic chemotherapy = End Date – Date of Randomization +1.

Table 11: Censoring Rules for Cytotoxic Chemotherapy

Situation	End Date on or prior to cut-off date	Censored
Recorded cytotoxic chemotherapy during the study	Date of the first assessment that qualified as cytotoxic chemotherapy	No
Death during the study before cytotoxic chemotherapy	Date of last visit at which cytotoxic chemotherapy question was collected or randomization date whatever comes later	Yes
Patient has no recorded cytotoxic chemotherapy at data cut-off	Date of last visit at which cytotoxic chemotherapy question was collected or randomization date whatever comes later	Yes

Patient lost to follow-up before data cut-off	Date of last visit at which cytotoxic chemotherapy question was collected or randomization date whatever comes later	Yes

Cytotoxic chemotherapy is a specific antineoplastic therapy and therefore the CRF question "Has the patient taken any new antineoplastic therapy since previous visit?" implies the question for cytotoxic chemotherapy.

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'CYTOC'.

A log rank test stratified by stratification factor of PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization will be used to compare the darolutamide treated and placebo groups.

The hazard ratio as well as its 95% CI will be presented based on fitting a Cox regression model with treatment as factor and stratified by PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm. The Kaplan-Meier estimate of the median survival time and first and third quartiles will be presented with 95% CI. Greenwood's formula will be used for the standard error of the Kaplan-Meier estimate in the calculation of these confidence limits.

6.2.2.4 Time to first symptomatic skeletal event (SSE)

The time to first SSE is defined as time from randomization to the occurrence of the first SSE. SSE is defined as external beam radiation therapy (EBRT) to relieve skeletal symptoms, new symptomatic pathologic bone fracture, occurrence of spinal cord compression, or tumor-related orthopedic surgical intervention.

Patients who do not reach the SSE event will be censored at their last visit (SSE assessment).

If the complete date of the assessment is not known, but the year and month are available, Day 15 of the month will be used for the calculation.

Table 12 will be used for calculating the time to first SSE. Time to first SSE = End Date – Date of Randomization +1.

Table 12: Censoring Rules for SSE

Situation	End Date on or prior to cut-off date	Censored
Recorded SSE event during the study	Date of the first assessment that qualified as SSE	No
Patient has no recorded SSE event at data cut-off	Date of last SSE assessment before data cut-off	Yes

Patient lost to follow-up before data cut-off	Date of last SSE assessment or randomization date, whichever comes later	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'SSEFS'.

A log rank test stratified by stratification factor of PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization will be used to compare the darolutamide treated and placebo groups.

The hazard ratio as well as its 95% CI will be presented based on fitting a Cox regression model with treatment as factor and stratified by PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm. The Kaplan-Meier estimate of the median survival time and first and third quartiles will be presented with 95% CI. Greenwood's formula will be used for the standard error of the Kaplan-Meier estimate in the calculation of these confidence limits.

The data of the second analysis of SSE will be analyzed in addition with the RPSFT and IPE methods for the same reasons as described for the analyses of OS. Only the unadjusted SSE data will be used for testing for a statistically significant treatment effect.

Sensitivity analysis 1: **SSE-free survival** will be analyzed like time to first SSE event, death will be considered as an event. See Table 13 describe censoring rules (SSE-free survival = End Date – Date of Randomization +1).

Table 13: Censoring Rules for SSE-free survival

Situation	End Date on or prior to cut-off date	Censored
Recorded SSE event during the study	Date of the first assessment that qualified as SSE	No
Death during the study before SSE assessment	Date of death	No
Patient has no recorded SSE at data cut-off	Date of last SSE assessment or randomization date whichever comes later	Yes
Patient lost to follow-up before data cut-off	Date of last SSE assessment or at the data cut-off, whichever occurs first	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'SASSEDTH'.

A summary description of skeletal events will be tabulated.

6.2.3 Tertiary or additional efficacy analysis

Calculating the p-value for tertiary endpoints has informative character only. These efficacy endpoints are:

6.2.3.1 Progression-free Survival (PFS)

PFS is defined as the time (days) from the date of randomization to the date of radiological disease progression based on independent blinded central reading, including progressing pelvic lymph nodes and new pathologic lymph nodes identified above or below the aortic bifurcation or death due to any cause, whichever occurs first. The radiological progression component of PFS will be derived by taking all distant metastasis events as determined for the MFS endpoint, adding all local radiological progression events per RECIST evaluation and choosing whatever comes first, in cases where both types of radiological progression were observed.

Censoring rules for PFS are derived from the rules used for MFS as in Table 14.

Censoring Rules

Time of PFS is calculated as time to PFS = End Date-Date of randomization +1.

Table 14: Censoring Rules for PFS Based on independent blinded central reading Data

Situation	End Date on or prior to cut-off date	Censored
Documented metastasis or local progression	Date of documented metastasis	No
Documented metastasis at baseline	Date of randomization	No CCI [redacted]
Documented metastasis or local progression after two or more consecutively missed tumor assessments, i.e. metastasis later than last evaluable scan + (32 +1) weeks	Date of last tumor assessment that the patient was known to be metastasis-free	Yes
Death before documented metastasis or local progression and not later than last evaluable scan + (32+1) weeks	Date of death	No
Death before documented metastasis or	Date of last tumor assessment that	Yes

Situation	End Date on or prior to cut-off date	Censored
local progression and after two or more consecutively missed tumor assessments, i.e. death later than last evaluable scan + (32+1) weeks	the patient was known to be metastasis-free	
Discontinued the study before any post-baseline tumor assessments	Date of randomization	Yes
Discontinued the study before any post-baseline tumor assessments and died within (32+1) weeks after randomization	Date of death	No
Discontinued the study before any post-baseline tumor assessments and died later than (32+1) weeks after randomization	Date of randomization	Yes
Discontinued the study, but no documented metastasis or local progression	Date of last tumor assessment before discontinuation	Yes
Prohibited new anticancer treatment started prior to documented metastasis or local progression	Date of last tumor assessment before start of prohibited new treatment	Yes
Patients still on treatment without documented metastasis or local progression as of data cut-off	Date of last tumor assessment	Yes

If the complete date of the scan is not known, but the year and month are available, Day 15 of the month will be used for the calculation.

The date used for the calculation of PFS is the actual (rather than scheduled) date of assessment (i.e. actual scan date); incomplete scans are not considered unless they showed progression.

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to ‘PFS’ (baseline metastasis non-censored) and ‘SAPFSBAS’ (baseline metastasis censored).

The two treatment groups will be compared using a stratified two-sided log-rank test with an alpha of 0.05 stratified by the same stratification factors used in the MFS analysis.

The product-limit estimates of the survival distribution functions (Kaplan-Meier) will be presented for each treatment group: N, total censored, total failed, time to progression (median and its 95% CI, first and third quartiles, range), rate of patients who did not yet progressed at 4,8, 12, etc. months, KM curves.

The hazard ratio (darolutamide over placebo) and its 95% CI will be generated with the Cox model as for MFS.

6.2.3.2 Time to first prostate cancer-related invasive procedures

Time to first prostate cancer-related invasive procedure is defined as time from randomization to date of first prostate cancer-related invasive procedure.

Prostate cancer-related invasive procedure is defined as any procedure needed for alleviation of symptoms, signs or findings caused by progression of prostate cancer (e.g. catheterization of the bladder, percutaneous drainage of hydronephrosis, palliative electro resection of the prostate, etc.).

Status of prostate cancer-related invasive procedures will be assessed from randomization until the first prostate cancer-related invasive procedure, i.e. until end of follow-up period.

Patient with no prostate cancer-related invasive procedure will be censored.

Table 15 will be used for calculating the time to first prostate cancer-related invasive procedure. Time Cancer-related invasive procedure = End Date-Date of randomization +1.

Table 15: Censoring Rules for prostate cancer-related invasive procedure

Situation	End Date on or prior to cut-off date	Censored
Recorded prostate cancer-related invasive procedure during the study	Date of the first assessment that qualified as prostate cancer-related invasive procedure	No
Death during the study before prostate cancer-related invasive procedure	Date of the last visit at which prostate cancer-related invasive procedure question was collected or randomization date whatever comes later	Yes
Patient has no recorded prostate cancer-related invasive procedure at data cut-off	Date of last visit at which prostate cancer-related invasive procedure question or randomization date whatever comes later	Yes
Patient lost to follow-up before data cut-off	Date of last visit at which prostate cancer-related invasive procedure question or at the data cut-off, whichever occurs first	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to ‘CRIP’.

A log rank test stratified by same stratification factors used in the MFS analysis will be used to compare the darolutamide treated and placebo groups.

The hazard ratio as well as its 95% CI will be generated with a Cox regression model as for MFS.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm. The Kaplan-Meier estimate of the median time to first prostate cancer-related invasive procedure and first and third quartiles will be presented with 95% CI.

6.2.3.3 Time to initiation of subsequent antineoplastic therapy

Time to initiation of subsequent antineoplastic therapy is defined as time from randomization to initiation of first antineoplastic therapy.

Subsequent antineoplastic therapy will be assessed from the first follow-up visit until the first use of such therapy. Patients who do not reach the endpoint or died will be censored.

Antineoplastic therapy will be selected using:

-ATC code class L (antineoplastic and immunodulating agents): L01 Antineoplastic agents (except cytotoxic chemotherapy L01A, L01B, L01C, L01D and L01X), L02 endocrine therapy and L03 immunostimulants,

-ATC code class H: H02Corticosteroids for systemic use.

The list of antineoplastic therapy is available in Appendix 9.8.

Table 16 will be used for calculating the time to initiation of subsequent antineoplastic therapy. Time to initiation of subsequent antineoplastic therapy event = End Date-Date of randomization +1.

Table 16: Censoring Rules for antineoplastic therapy

Situation	End Date on or prior to cut-off date	Censored
Recorded antineoplastic therapy during the study	Date of the first assessment that qualified as antineoplastic therapy	No
Death during the study before antineoplastic therapy	Date of last visit at which antineoplastic therapy question was collected or randomization date whatever comes later	Yes
Patient has no recorded antineoplastic therapy at data cut-off	Date of last visit at which antineoplastic therapy question was collected or randomization date whatever comes later	Yes
Patient lost to follow-up before data cut-off	Date of last visit at which antineoplastic therapy question was collected or at the data cut-off, whichever occurs first	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'ISAT'.

A log rank test stratified by same stratification factors used in the MFS analysis will be used to compare the darolutamide treated and placebo groups.

The hazard ratio as well as its 95% CI will be generated with a Cox regression model as for MFS.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm. The Kaplan-Meier estimate of the median time to initiation of subsequent antineoplastic therapy and first and third quartiles will be presented with 95% CIs.

A summary of first subsequent antineoplastic therapy will be provided.

6.2.3.4 Time to PSA progression

Time to PSA progression is defined as the time (days) from the date of randomization to the date of first PSA progression.

PSA progression is defined according to the Consensus Guidelines of the Prostate Cancer Trials Working Group2 (PCWG2) as:

For patients with no decline from baseline at week 16, the PSA progression is defined as the date that a ≥ 25% PSA increase in PSA along with an absolute increase of ≥ 2 ng/ml above the baseline is documented, which is confirmed by a second consecutive value obtained 3 or more weeks later,

Or

For patients with declines from baseline at week 16, the PSA progression is defined as the date that a ≥ 25% PSA increase and an absolute increase of ≥ 2 ng/ml above the nadir is documented, which is confirmed by a second consecutive value obtained 3 or more weeks later.

The above PSA definition deviate from PCWG2, as confirmation by a second value is requested here for a patient with no decline from baseline.

Early increases in PSA values before the 16 weeks are not considered as PSA progression.

PSA documented post end of study-treatment visit will not be considered. Central laboratory assessment will be considered.

Baseline PSA value is the last non-missing observation on or before the first day of study drug intake.

Table 17will be used for calculating the time to PSA progression. Time to PSA progresion =End Date-Date of randomization +1.

Table 17: Censoring Rules for PSA progression

Situation	End Date on or prior to cut-off date	Censored
No baseline or post-baseline PSA assessment	Date of randomization	Yes
No PSA progression	Last PSA assessment date	Yes
Patient had an PSA progression	The date of first observation of PSA progression	No
PSA progression after two or more consecutive missed PSA assessments	Date of last PSA assessment before missed assessments	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'TTCPP'.

Time to PSA progression will be analyzed with the same methods as the primary variable MFS. A log rank test stratified by the same factors as used for randomization will be used to compare the darolutamide treated and placebo groups. Kaplan-Meier curves, median survival times, and their 95% CI will be presented. Hazard ratios will be calculated with the Cox model.

6.2.3.5 Percent of patients with PSA response

The percentage change of PSA from baseline at any time point will be calculated and the proportion of patients achieving a decline of ≥ 50% from baseline will be determined.

PSA values will be collected until the end-of-study treatment visit.

PSA response rate will be compared between treatment groups using Cochran-Mantel-Haenszel test adjusting for same stratification factors as for primary endpoint. Estimates and the 95% CI will be computed for each treatment group.

Sample code for Cochran-Mantel-Haenszel test:

```
proc freq data=data;
tables psa_double*ot_therapy*PSA_Response*trtpn /CMH;
run;
```

Where:

PSA_Response contains corresponding PSA Response values, *trtpn* contains assigned treatment group, *psa_double* contains PSA doubling time category at randomization, *ot_therapy* contains category for osteoclast-targeted therapy use at randomization.

In addition, descriptive statistics will be provided for PSA at baseline, 16 weeks, change from 16 weeks to baseline, maximum percent decline from baseline at any time on study.

6.2.3.6 Percent of patients with ECOG performance status deterioration

ECOG performance status (PS) is assessed at screening and every 16 weeks until the end of follow-up period.

ECOG PS will be tabulated and summarized by visit for observed values and changes from baseline using descriptive statistics, as appropriate. If more than one baseline assessment was collected, the most recent one will be used. If several assessments were performed on the same day (without timing information) the most severe score will be used.

ECOG performance status deterioration is defined as an increase to grade 3 or higher, with an increase of at least 2 from baseline.

Deterioration will be compared between treatment groups using Cochran-Mantel-Haenszel test adjusting for same stratification factors as for primary endpoint. Estimates and the 95% CI will be computed for each treatment group.

6.2.3.7 Time to ECOG performance status deterioration

Time from randomization to ECOG performance status deterioration is defined as time from randomization to ECOG PS deterioration.

Table 18 will be used for calculating the time to ECOG performance status deterioration. Time to ECOG performance status deterioration event = End Date-Date of randomization +1.

Table 18: Censoring Rules for ECOG performance status deterioration

Situation	End Date on or prior to cut-off date	Censored
Recorded ECOG PS deterioration during the study	Date of the first ECOG PS deterioration	No
No ECOG PS deterioration during study	Date of last ECOG PS assessment	Yes
Patient has no post baseline ECOG PS assessment at data cut-off	Date of randomization	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to ‘ECOGDET’.

A log rank test stratified by same stratification factors used in the MFS analysis will be used to compare the darolutamide treated and placebo groups.

The hazard ratio as well as its 95% CI will be generated with a Cox regression model as for MFS.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm. The Kaplan-Meier estimate of the median time to ECOG PS deterioration and first and third quartiles will be presented with 95% CIs.

6.2.3.8 Time to first opioid use for cancer pain

Time to first opioid use for cancer pain is defined as time from randomization to first opioid use for cancer pain.

List of opioids for cancer pain is available in Appendix 9.10.

Table 19 will be used for calculating the time to Time to first opioid use for cancer pain. Time to first opioid use for cancer pain = End Date-Date of randomization +1.

Table 19: Censoring Rules for Time to first opioid for cancer pain

Situation	End Date on or prior to cut-off date	Censored
Recorded opioid use for cancer pain during the study	Date of first opioid use	No
No opioid use for cancer pain during study	Date of last visit	Yes
Patient has no post baseline assessment at data cut-off	Date of randomization	Yes

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'OCP'.

A log rank test stratified by same stratification factors used in the MFS analysis will be used to compare the darolutamide and placebo groups.

The hazard ratio as well as its 95% CI will be generated with a Cox regression model as for MFS.

Kaplan-Meier survival curves (product-limit estimates) will be presented by treatment arm. Kaplan-Meier estimates of the median time and first and third quartiles times will be presented with 95% CIs.

6.2.3.9 Health related Quality of Life and utility values.

1) The **EQ-5D-3L** is a generic quality of life preference based instrument which has been validated in cancer population to measure both utility and health status.

The EQ-5D-3L contains a descriptive system which measures 5 health dimensions: mobility, self-care, usual activity, pain/discomfort, and anxiety/depression. Each dimension contains 3 levels of response to reflect the degree of problems patients have experienced: no problem (level 1), some problems (level 2), and extreme problems (level 3). These five health

dimensions are summarized into a single score, the EQ-5D-3L index score. The EQ-5D-3L index score ranges -0.59 to 1 with higher scores representing better health states.

The EQ-5D-3L also contains a visual analog scale (EQ-VAS), which records the respondents' self-rated health status on a vertical graduated visual analogue scale ranging from 0 (worst imaginable health state) to 100 (best imaginable health state).

For details on scoring, please refer to the EQ-5D-3L user guide Appendix 9.5.

Since EQ-5D-3L is a descriptive system based on five independent dimensions, a missing answer or ambiguous answer (i.e. marking of more than one level on scale) will lead to complete rejection of the questionnaire, but the VAS score will be retained if available.

2) The **FACT-P** questionnaire assesses prostate cancer-related quality of life and has been validated in the prostate cancer population.

This questionnaire contains 5 domains: Physical Well-Being (PWB), Social/Family Well-Being (SWB), Emotional Well-Being (EWB), Functional Well-Being (FWB), and Additional Concerns (also called Prostate cancer subscale [PCS]).

According to the FACT-P scoring guide, all subscale items are summed to a total which is the subscale (=domain) score. The FACT-P total score is the sum of the scores of 39 items of the questionnaire and ranges from 1 to 156; the higher the score, the better the quality of life of prostate cancer patients.

Each item can be answered on a 5-point (0–4) scale.

The sum of the scores on the first four domains (PWB, SWB, EWB, and FWB) constitutes the FACT-G (General). The sum of the scores on all five domains constitutes the FACT-P. The trial outcome index (TOI) is derived by the sum of PWB, FWB, and PCS scores.

For details on scoring, please refer to the FACT-P user guide Appendix 9.3.

In case of missing responses for one or more items, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the scale, then dividing the number of items actually answered. Prorating of scores is acceptable as long as more than 50% of the items are answered (assuming that the score of missing items are similar to those of non-missing items). If less than or equal to 50% of the items are answered for any domain, then the score of that specific domain is set to missing. The FACT-P total score is then calculated as the sum of the un-weighted subscale scores. Moreover, the FACT-P total score is set to missing if the related overall item response rate is less than or equal to 80%.

The total score is then calculated as the sum of the un-weighted subscale scores. A total score is calculated if all of the component subscales have a valid score.

Rate of the question GS7 'I am satisfied with my sex life' in the SWB domain will be used independent of the answer documented in the question Q1 "Regardless of your current level of sexual activity, please answer the following question. If you prefer not to answer it, please check this box and go to the next section".

3) The **BPI-SF** questionnaire assesses clinical pain related to cancer, it is a validated tool. Two scores will be derived: the pain severity and the pain interference.

Questionnaire is completed daily over a 7-day period. Averages during the period of reporting are calculated. If fewer than 4 reports out of the 7 days are available at the planned time point for assessment, this would constitute missing data.

The BPI assesses pain at its "worst" "least" "average" and "right now" (current pain). To derive the "pain severity" score a mean score of the 4 questions (questions 3 to 6 from the BPI-SF) related to pain will be created. In case of one or more missing information the "pain severity" score will be assessed to missing.

The BPI measures how much pain has interfered with seven daily activities, including general activity, walking ability, normal work, mood, enjoyment of life, relations with others, and sleep (question 9 from the BPI-SF). BPI "pain interference" is scored as the mean of the seven interference items. This mean can be used if more than 50% or four of seven, of the total items have been completed on a given administration.

Rate of pain entered in questions 3 to 9 will be used independent of the answer documented in the question 1 (have you had pain other than these everyday kinds of pain today) of the BPI-SF.

For details on scoring, please refer to the BPI-SF user guide Appendix 9.2.

4) The **EORTC-QLQ-PR25** questionnaire assesses prostate cancer-related quality of life and has been validated in the prostate cancer population.

The prostate cancer module is a 25-item questionnaire designed for use among patients with localized and metastatic prostate cancer. It includes subscales assessing urinary symptoms (8 items), bowel symptoms (4 items), hormonal treatment-related symptoms (6 items), incontinence aid (1 item), sexual activity (2 items) and sexual functioning (4 items). The sexual activity and sexual functioning scales constitute the functional scales. The urinary symptoms, bowel symptoms, hormonal treatment-related symptoms, and incontinence aid constitute the symptom scales.

If less than or equal to 50% of the items are answered for any subscales, then the score of that specific subscale is set to missing.

For details on scoring, please refer to the EORTC-QLQ-PR25 user guide Appendix 9.4.

PRO analyses:

PRO data as measured by the FACT-P, the EQ-5D-3L , the EORTC-QLQ-PR25 and the BPI-SF will be analyzed to assess differences in HRQoL and health utility values between treatment arms based on time-adjusted area under the curve (AUC) using all available data.

The PRO analyses will be performed for the patients in the FAS. Statistical tests will be performed with a 2 sided type I error of 5%.

Descriptive statistics on observed data will be presented for the FACT-P questionnaire (each domain score including the PCS and the FACT-P total score), for the EQ-5D-3L index score

(utility value) and visual analog scale score (VAS), for the EORTC-QLQ-PR25 questionnaire (each subscale score) and for BPI-SF questionnaire (pain severity and pain interference scores) at each assessment time and for change from baseline by treatment group. Questionnaires under unscheduled visits and not planned visits per protocol will not be displayed in the descriptive tables. Analyses will be done for patients with baseline assessments.

The frequency for missing health related quality of life (HRQoL) assessment by treatment group will be summarized.

An analysis of covariance (ANCOVA) model will be used to estimate the mean difference in the time-adjusted Area under Curve (AUC) between the two treatment groups, with covariates for baseline PRO scores and stratification factors as recorded in the IVRS data.

Least-square mean estimates, standard errors and 95% CIs will be estimated for each treatment group and for the treatment group difference.

Calculation of Time (AUC):

AUC will not be calculated if baseline data is missing.

The trapezoidal rule will be used to derive the AUC for a patient for the FACT-P total and subscale score, EQ-5D-3L index score, VAS score, BPI-SF pain severity and interference scores and EORTC-QLQ-PR25 urinary symptoms score. The time-adjusted AUC will be calculated by dividing the AUC by the duration (in days) over the treatment period. When calculating the AUC, the exact date of completion of the questionnaires should be plotted and the duration will be calculated in days. The time adjusted AUC for the FACT-P subscale scores and total score, EQ-5D-3L index score, VAS score, BPI-SF pain severity and interference scores and EORTC-QLQ-PR25 urinary symptoms score for an individual patient over a period of time [Ta,Tb] will be calculated as follows:

$$AUC_{a\text{-}b} = \frac{1}{2(T_b - T_a)} \sum_{i=a}^{b-1} (PRO_i + PRO_{i+1})(T_{i+1} - T_i)$$

Where PRO_i is the PRO measurement at time T_i (i=a… b).

To test the treatment effect, a mixed linear model (random coefficient model) will be used for the EQ-5D-3L index score and VAS score, and for the FACT-P total score, BPI-SF pain severity and interference scores and the subscale scores and the EORTC-QLQ-PR25 urinary symptoms score.

For patients with missing baseline FACT-P, PCS, EORTC-QLQ-PR25 urinary symptoms, EQ-5D-3L scores, time to deterioration will not be calculated.

6.2.3.9.1 Percent of patients with deterioration of FACT-P total score at 16 weeks

Patients will be defined as having total QoL deterioration, if they experience a decrease of ≥ 10 points in FACT-P total score at 16 weeks compared with baseline.

Deterioration from baseline in FACT-P total score will be compared between treatment arms using a Cochran-Mantel-Haenszel test stratified by stratification factor of PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

Analysis will be created on both sub-sets of patients:

- FAS,
- FAS excluding patients with no visit at week16.

6.2.3.9.2 Time to deterioration in PCS subscale score

Time to deterioration for the FACT-P PCS score: a deterioration being defined as a 3 point or more decline in the PCS score.

A patient's time to deterioration will be defined as the time from randomization to the deterioration date. For patients with no symptomatic deterioration at the time of the analysis, the time to deterioration will be censored at the date of the last FACT-P PCS valid assessment.

Table 20 will be used for calculating the time to deterioration in PCS. Time to Deterioration in PCS = End Date - Date of randomization + 1

Table 20: Censoring Rules for time to deterioration in PCS

Situation	End Date on or prior to cut-off date	Censored
No baseline or post-baseline PCS assessment	Date of randomization	Yes
No deterioration in PCS	Last PCS assessment date	Yes
Patient had a deterioration in PCS	The date of first observation of deterioration in PCS	No

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'FACTDETP'.

Times to deterioration will be displayed by the product-limit estimates of the survival distribution functions (Kaplan-Meier) for each treatment group: N, total censored, total failed, survival time (median and its 95% CI, range), deterioration-free rate at 12, 24, 36 etc.,

months, KM curves. The hazard ratio darolutamide over placebo, its 95% CI will be generated from the Cox model.

6.2.3.9.3 Percent of patients with improvement of EORTC-QLQ-PR25 urinary symptoms

Patients will be defined as having EORTC-QLQ-PR25 urinary improvement, if they experience a decrease of ≥ 8 points in EORTC-QLQ-PR25 urinary symptoms score, scale (PRURI), from baseline.

Improvement from baseline in EORTC-QLQ-PR25 urinary symptoms score will be compared between treatment arms using a Cochran-Mantel-Haenszel test stratified by stratification factor of PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

6.2.3.9.4 Time to worsening of EORTC-QLQ-PR25 urinary symptom score

A patient's time to worsening of EORTC-QLQ-PR25 urinary symptoms will be defined as the time from randomization to the deterioration date. For patients with no symptomatic deterioration at the time of the analysis, the time to deterioration will be censored at the date of the last EORTC-QLQ-PR25 (PRURI) valid assessment.

Patients will be defined as having EORTC-QLQ-PR25 urinary symptoms deterioration if they experience an increase of greater or equal to 8 points in EORTC-QLQ-PR25 urinary symptoms score, scale (PRURI), from baseline.

Table 21 will be used for calculating the time to worsening of EORTC-QLQ-PR25 urinary symptoms. Time to worsening of EORTC-QLQ-PR25 urinary symptoms = End Date - Date of randomization +1.

Table 21: Censoring Rules for time to worsening of EORTC-QLQ-PR25 urinary symptoms

Situation	End Date on or prior to cut-off date	Censored
No baseline or post-baseline EORTC-QLQ-PR25 urinary assessment	Date of randomization	Yes
No worsening in EORTC-QLQ-PR25 urinary symptoms score	Last EORTC-QLQ-PR25 urinary symptoms assessment date	Yes
Patient had a worsening in EORTC-QLQ-PR25 urinary symptoms	The date of first observation of worsening in EORTC-QLQ-PR25 urinary symptoms	No

This endpoint is selected from the ADTTE dataset using variable PARAMCD equal to 'EORTCWOR'.

Times to worsening will be displayed by the product-limit estimates of the survival distribution functions (Kaplan-Meier) for each treatment group: N, total censored, total failed, survival time (median and its 95% CI, range), deterioration-free rate at 12, 24, and 36 etc.… months, KM curves. The hazard ratio darolutamide over placebo, its 95% CI will be generated from the Cox model.

6.2.3.9.5 Percent of patients with deterioration of EQ-5D-3L utility score at 16 weeks

Patients will be defined as having a deterioration in EQ-5D-3L index if they experience a deterioration of greater or equal 0.06 points compared to baseline at 16 weeks after start of treatment.

Deterioration from baseline in EQ-5D-3L index score will be compared between treatment arms using a Cochran-Mantel-Haenszel test stratified by stratification factor of PSADT (≤6 vs. >6 months) and use of osteoclast-targeted therapy (yes vs. no) at randomization.

Analysis will be created on both sub-sets of patients:

- FAS,
- FAS excluding patients with no visit at week16.

6.2.4 Subgroup analysis

There are specific subgroup analyses of interest which will be performed for the primary endpoint (MFS) and the secondary endpoints (OS).

Subgroup analyses will be based on FAS analysis set.

Descriptive statistics and hazard ratio with 95% CI will be provided within each category, provided there are a sufficient number of patients in total within the subgroup across the treatment arms. All subgroups analyses will be done on non-stratified Cox model and log-rank test.

Subgroup analyses of interest:

- Baseline CRF PSADT (≤6 and >6 months),
- Baseline CRF osteoclast-targeted therapy (yes vs. no),
- Baseline PSA (ng/mL) (≤10; >10 to ≤20; >20) from central laboratory,
- Baseline PSA (ng/mL) (at or below median vs. above the median [median of all patients]) from central laboratory,
- Gleason score at diagnosis (≥7 vs. <7),

- Age (years) (<65, 65- 74, 75 - 84, ≥85),
- Geographical region (North America, Asia Pacific, ROW),
- Baseline presence of regional pathological lymph nodes (yes vs. no) by central imaging review,
- Baseline ECOG performance status (0, 1),
- Race: White, Asian, Black or African American, Other; ethnicity 'Hispanic or Latino' will be shown as well as it cannot be mapped to the Bayer race subgroups,
- Number of prior hormonal therapies (1, ≥2) (for definition of hormonal therapies, refers to Section 9.13).

Further important baseline cancer characteristics may also be considered.

Subgroup analyses will not be performed for the sensitivity analyses.

6.3 Pharmacokinetics/pharmacodynamics

PK sampling will be undertaken at centers participating in PK sample collection. It is estimated that approximately 600 patients will participate in the PK sampling, resulting in approximately 400 patients randomized to receive darolutamide being included in the population PK analysis.

Venous blood samples will be collected for determination of darolutamide, the diastereomers (*S*,*R*)-darolutamide and (*S*,*S*)-darolutamide and the metabolite keto-darolutamide concentrations in plasma.

Pharmacokinetics of darolutamide: PK data will be analyzed using a population PK model to explore possible relationships between selected demographics and clinical covariates and drug/metabolite exposure.

Pharmacokinetic analyses and results will be provided in a separate report. Full details of the modelling will be provided in a separate Modelling & simulation analysis plan.

6.4 Biomarker evaluation

Blood (plasma) samples will be collected for tumour-related biomarker assessments and a whole blood sample for pharmacogenetic (PG) assessments. Patients' biomarker status will be correlated with clinical treatment effect to explore which targets may be important in defining the appropriate therapeutic population for the agent.

Since this clinical study is not powered to specifically address biomarker questions, and since the analysis will be conducted retrospectively, these biomarker analyses are considered exploratory.

Biomarker analyses and results will be provided in a separate report.

6.5 Safety

No formal statistical tests will be done for the safety endpoints. All analyses for safety will be performed in the SAF population.

An updated analysis of safety events will be performed as of the final OS analysis.

Specific tables of adverse events will be displayed by following subgroups:

- Age (years) :<65, 65 - 74, 75 - 84, ≥85,
- Geographical region (North America, Asia Pacific, ROW).
- Renal function - eGFR at baseline (normal vs.mildly impaired vs. moderately impaired, severely impaired and end stage renal disease) (for calculation see Section 6.1.2),
- Hepatic impairment at baseline (normal vs mild vs moderate and severe impairment) (for calculation see section 6.1.2),
- Concomitant statin use (no vs. yes), to be determined by concomitant medication (see Appendix 9.9).

6.5.1 Extent of exposure

Extent of exposure will be summarized for the SAF by treatment group, using descriptive statistics.

Duration of study treatment will be calculated in days and presented in months as the date of the last dose of any study treatment – date of the first dose of any study treatment + 1.

Dose modification will be summarized.

Extent of exposure will be described by age groups and geographical regions.

6.5.2 Adverse events

All adverse events (AE) whether considered drug-related or not, will be reported on the CRF with diagnosis, start/stop dates, dates of any grade change, action taken, whether treatment was discontinued, any corrective measures taken, and outcome. For all events, the relationship to treatment and the severity of the event will be determined by the Investigator. AEs will be classified and coded using the National Cancer Institute (NCI) Common Terminology Criteria Adverse Event (CTCAE), version 4.03.

The treatment period for this study, for purposes of safety analyses, extends from the initiation of study treatment until 30 days after the last administration of study treatment.

Each change in AE grade was entered as a separate record with no automatic link to the original AE record, which may lead to variations of the verbatim for the same AE. This makes a correct grouping of AE grades, which actually belong together, not always possible. An AE is considered as treatment-emergent (TEAE) if there is an AE record which starts during treatment or within the post-treatment time window of 30 days.

For further definitions of the terms AE, SAE, seriousness, intensity, causal relationship with treatment, causal relationship to protocol-required procedures, action taken, and outcome; see Protocol Section 6.6.1.

Descriptive summary tables (frequency and percentage of patients, not of events) will be presented by treatment group and MedDRA version 20.0, or the most recent version terminology, for the following:

- AEs during screening
- Treatment-emergent AEs
- Treatment-emergent AEs with grade 3, 4, or 5
- Treatment-emergent AEs occurring in at least 1% of patients
- Treatment-emergent AEs occurring in at least 5% of patients
- Treatment-emergent AEs leading to study drug withdrawal
- Treatment-emergent AEs leading to dose reduction
- Treatment-emergent AEs leading to drug interruption
- Treatment-emergent AEs leading to dose reduction and/or drug interruption
- Treatment-emergent drug-related AEs
- Treatment-emergent drug-related AEs with grades 3, 4, or 5
- Treatment-emergent drug-related AEs occurring in at least 5% of patients in any treatment arm
- Treatment-emergent drug-related AEs leading to study drug withdrawal
- Treatment-emergent drug-related AEs leading to dose reduction
- Treatment-emergent drug-related AEs leading to drug interruption
- Treatment-emergent drug-related AEs leading to dose reduction and/or drug interruption

Interval specific and cumulative event rates for treatment-emergent adverse events (for AE with at least a 5% total incidence rate (any grade)).

Subgroup analyses will be performed for TEAEs. Tables showing differences of at least 5% points in incidence proportions between any of the subgroup categories will be produced.

Following treatment-emergent adverse events are considered as special topics:

- Bone fracture (selected using HLT: Fractures and dislocations NEC (without PTs: Joint dislocation, Joint dislocation pathological), Limb fractures and dislocations (without PT: Radial head dislocation), Pelvic fractures and dislocations, Skull fractures, facial bone fractures and dislocations, Spinal fractures and dislocations (without PT: Dislocation of vertebra),Thoracic cage fractures and dislocations (without PT: Dislocation of sternum))
- Fall (selection of MLG Fall (only PT Fall))
- Seizure (selection of MLG Seizures)
- Fatigue (selection of MLG decreased general strength and energy, selection of PT Lethargy, Malaise, Chronic fatigue syndrome)
- Weight decreased (selection of MLG Weight decreased)

Following tables will be created for the TEAE of special topics

- Treatment-emergent AEs
- Treatment-emergent AEs leading to study drug withdrawal
- Treatment-emergent AEs leading to dose reduction
- Treatment-emergent AEs leading to drug interruption
- Treatment-emergent SAEs

Incidence of TEAEs of special topics will be displayed for the subgroup of concomitant statin use.

Fracture events will be described by a cumulative incidence plot of fracture. A summary of treatment-emergent fracture by bone sparing agent use (for selection of bone sparing agent see Appendix 9.14) at study entry will be provided. Association with weight decrease will be presented by histogram of fracture events by patient weight change (the weight collected closest to the start of fracture will be considered and compared to baseline weight).

A listing will be generated for patients with fall treatment-emergent events with syncope and/or loss of consciousness (using the MLG Syncope).

An overview of patients at risk for developing an AE seizure will be tabulated by displaying medical history.

Listing of non-treatment-emergent AEs will be created.

Incidence of post treatment non-treatment-emergent AEs will be tabulated.

To adjust for unequal lengths of study treatment period among patients, and potentially between treatment groups, an additional summary based on event rate per 100 patient year will be performed for all AEs, special topics AEs and all SAEs occurring after the first dose

of study treatment. The event rate per patient is calculated as the total number of events divided by the total treatment duration in years.

6.5.3 Deaths and Serious Adverse events

Serious adverse events (SAE) will be classified using the National Cancer Institute (NCI) Common Terminology Criteria Adverse Event (CTCAE), version 4.03 and MedDRA version 20.0 or most recent version

- Treatment-emergent SAEs
- Treatment-emergent SAEs leading to study drug withdrawal
- Treatment-emergent SAEs leading to dose reduction
- Treatment-emergent SAEs leading to drug interruption
- Treatment-emergent SAEs leading to dose reduction and/or drug interruption
- Treatment-emergent drug-related SAEs
- Listing of treatment-emergent SAEs
- Listing of non-treatment-emergent SAEs

The incidence of deaths in the study and especially deaths up to within 30 days of last dose of study drug will be summarized by each treatment group and cause of death. All deaths up to within 30 days of last dose of study drug will be listed by patient with start and stop date of study medication, date of death, and cause of death. All deaths beyond 30 days after last dose of study drug will be displayed in a separate listing.

6.5.4 Clinical laboratory data

Descriptive statistics (number of observations, mean, standard deviation, minimum, median and maximum values) will be presented for clinical laboratory tests (hematology, clinical biochemistry and urinalysis), their changes from baseline (including baseline value), and their percent changes from baseline by treatment group at applicable visits.

Hematological and biochemical laboratory values will be graded based on NCI CTCAE version 4.03. CTCAE severity grading for laboratory abnormalities are based on applicable laboratory threshold values outlined in NCI CTCAE v4.03. It should be noted that in the present analysis of those laboratory parameters for which additional clinical information potentially can also influence the toxicity grade, this clinical information is in general not available and only the laboratory measurements are used for grading.

Any additional specific handling of the NCI CTCAE v4.03 toxicity grading assignments will be noted in the footnotes of the corresponding tables as applicable per the data collection in the study.

- In the event of overlapping CTCAE criteria ranges for specific lab tests, the algorithm assigns the worst grade
- If calcium type is not recorded (i.e. only "calcium" is recorded), then grading is done as if the calcium is total calcium. "Calcium corrected" is computed from total calcium and serum albumin (if ≤ 4.0 g/dl) from the same time point based on CTCAE v3.0 guidance. If serum albumin (if ≤ 4.0 g/dl) from the same time point is not available or if "calcium, unspecified" was collected then grading is done as if the calcium is "corrected calcium."
- Results with special characters (such as ">" and "<") are not graded.

The frequency of laboratory abnormalities regarding hematology, coagulation panel, clinical chemistry, and urinalysis will be tabulated by treatment group. Worst grades for hematological and biochemical toxicities will be calculated according to CTCAE, version 4.03 based on laboratory measurements, and will be summarized by treatment group and NCI CTCAE v4.03 category and worst grade.

Clinical laboratory toxicities during treatment including a period of 30 days after last dose of treatment will be considered as "treatment-emergent.

The last non-missing value before or on the first day of study drug will be retained as "baseline" data. If several assessments are performed on the same day (without timing information) the average of the values will be considered.

Incidence tables (frequency and percentage of patients) as well as tables with change in NCI CTCAE v4.03 worst grade from baseline will be presented as following:

- Hematological and biochemical toxicity during screening (the last evaluation available before assignment to treatment is taken into account).
- Treatment-emergent hematological and biochemical toxicity.
- Treatment-emergent hematological and biochemical toxicities with incidence rate above 5% in any treatment arm.
- Change in worst grade for hematological and biochemical toxicity from baseline.

The laboratory values will be also categorized into low, normal and high according to their reference ranges.

Descriptive statistics will be calculated by treatment group and time interval. 16-week time intervals will be used.

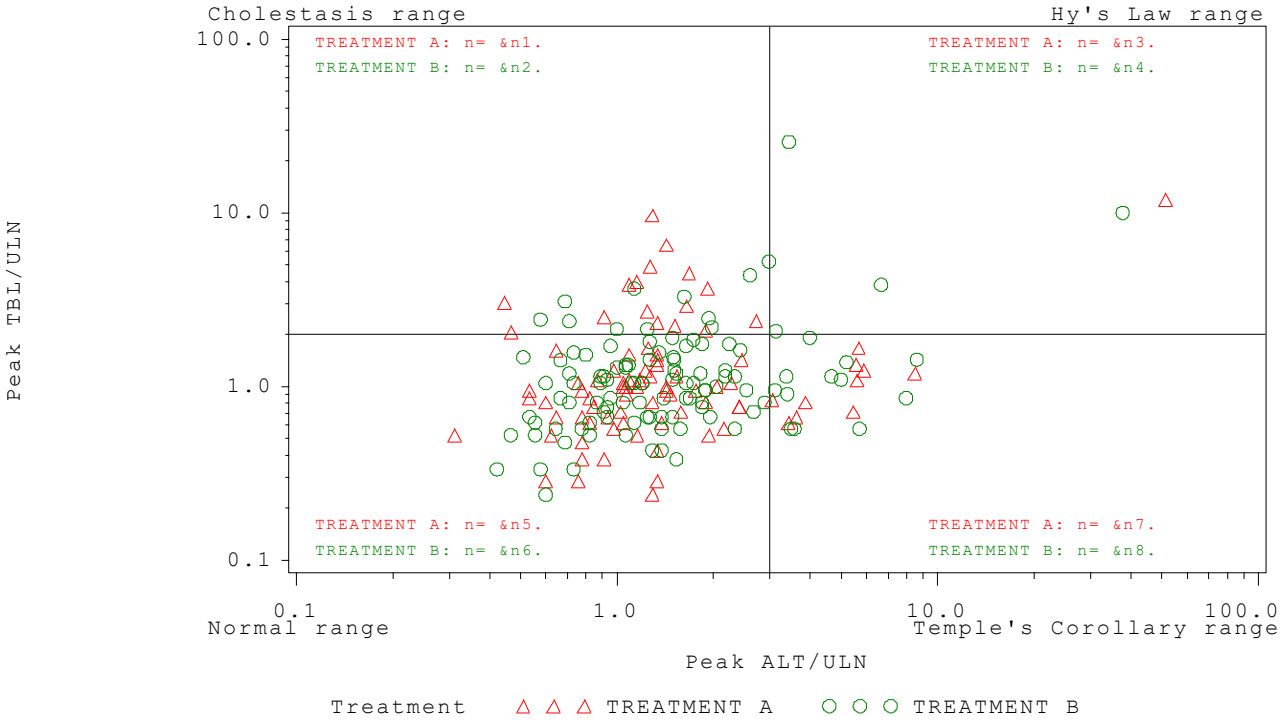
Only for ALT, descriptive statistics will be displayed for subgroups of statin users and non-statin users. Clinical laboratory tables will not be described by subgroups.

The data will also be displayed graphically using box plots and laboratory shift plots. These will show the baseline value and post baselines values up-to end of treatment.

A listing will be provided for all patients possibly fulfilling Hy's Law criteria, i.e. patients with elevated AST and / or ALT > 3xULN, alkaline phosphatase < 2xULN and bilirubin

≥2xULN (ref. 9). For possible Hy's Law cases relevant laboratory data will also be displayed graphically within actual patient profiles (presenting total bilirubin, ALT, AST and ALP values in terms of ULN over time) by treatment group. Below example of Hy's law plot will be provided for peak total bilirubin vs ALT. If a patient has any total bilirubin ≥ 2xULN then peak bilirubin will be plotted versus the maximum ALT amongst the total bilirubin ≥ 2xULN. Otherwise, peak bilirubin will be plotted versus peak ALT.

Example for Hy's law plot:



Unscheduled laboratory data will be listed (Section 16 of the CSR) but will be included in the descriptive tables.

6.5.5 12-Lead ECG, QTc

Analyses of ECG and QTc will be performed on the SAF population; no subgroup tables will be created.

Corrected QT (QTc) will be calculated using Bazett's (QTcB) and Fridericia's (QTcF) formula.

Standard 12-lead ECGs will be performed in a supine position after at least 10 minutes rest. Three consecutive 12-lead ECG recordings will be performed within approximately 5

minutes at screening and on day 1 visit before the first study treatment administration, at other visits, 12-lead ECG will be obtained once.

An ECG with a QTcF > 500 ms should be confirmed by a second ECG taken 1-2 hours later.

Descriptive statistics including arithmetic mean, SD, median, minimum, and maximum will be presented for the following ECG parameters: HR, RR, PR, QT, QRS, QTcB and QTcF. Parameters will be summarized for actual results and the change from baseline for the safety analysis set by treatment group at each scheduled visit. Baseline for 12-lead ECG will be mean of separate recordings. Similarly, post baseline values of 12-lead ECG for patients will be mean of separate recordings.

Mean time courses +/- one standard deviation will be displayed graphically.

The last non-missing average value before the first study drug intake will be considered as "baseline" data.

If several assessments are performed on the same day (without timing information) the average of the values will be considered.

In case, timing of ECG is missing at day1 then screening assessment will be considered as baseline data.

The number and percent of patients with absolute QTc interval prolongation or QTc increase from baseline will be defined and summarized by treatment arms. The criteria are:

QTc prolongation: QTc interval > 450 msec, QTc interval > 480 msec and QTc interval > 500 msec.

Increase from baseline in QTc interval: QTc interval increases from baseline > 30 msec and QTc interval increases from baseline > 60 msec.

The number and percent of patients with new clinically significant abnormalities on ECGs per the investigator's assessment at post-baseline time points will be summarized by treatment arms.

ECG data will be presented in a listing. Values with QTc interval > 450 msec,> 480 msec, and >500 msec, and change from baseline in QTc interval >30 msec and >60 msec will be flagged. In addition, abnormalities or clinically significant abnormalities found in the 12-lead ECGs at each scheduled time point will be provided in a listing.

ECG data collected under unscheduled visit will be included in the analyses.

6.5.6 Other safety measures

For each treatment group, vital signs (i.e. blood pressure, heart rate, weight and BMI) will be tabulated and summarized by visit for observed values and changes from baseline using descriptive statistics, as appropriate. If more than one baseline assessment was collected, the most recent one will be used. If several assessments are performed on the same day (without timing information) the average of the values will be considered.

In addition to descriptive analysis box plots will be created.

For weight parameter a graph representing mean changes in weight from baseline by visit will be created.

Outlier analyses will be conducted using the following limits:

- low systolic blood pressure: ≤ 90 mmHg and a decrease of ≥ 20 mmHg
- high systolic blood pressure: >190 mmHg and an increase of ≥ 20 mmHg
- low diastolic blood pressure: ≤ 50 mmHg and a decrease of ≥20 mmHg
- high diastolic blood pressure: > 105 mmHg and an increase of ≥ 20 mmHg
- low heart rate: < 50 bpm and a decrease of ≥ 15 bpm
- high heart rate: > 120 bpm and an increase of ≥ 15 bpm

The number and percentage of patients with outlying values will be tabulated by treatment group and time interval.

No subgroup analyses will be performed.

Unscheduled vital signs data will be listed (Section 16 of the CSR) but will be included in the summary tables.

6.5.7 Physical examinations

The number and percent of patients with physical examination abnormalities will be summarized based on the safety analysis set for each treatment group and overall. Physical examination findings will be presented in a data listing.

No subgroup analyses will be performed.

7. Document history and changes in the planned statistical analysis:

Abbreviated SAP Version 1.0 dated 14 MAY 2015

Abbreviated SAP Version 2.1 dated 22 JUN 2017

SAP Version 3.0 dated 12 MAR 2018

SAP Version 4.0 dated 10 AUG 2018

SAP Version 4.1 dated 13 SEP 2018

8. References

1. FACT-P guideline refers to the Administration and Scoring Guidelines in the manual or on-line at www.facit.org.
2. BPI user guide, 2009Charles S. Cleveland.
3. EORTC QLQ-C30 scoring manual - Third edition, 2001.
4. Pain Palliation Measurement in Cancer Clinical Trials – the US Food and Drug Administration Perceptive. Ethan Basch. Cancer, March 1, 2014.
5. Branson M, Whitehead J: Estimating a Treatment Effect in Survival Studies in Which Patients Switch Treatment. Statistics in Medicine 2002, 21:2449-2463.
6. Penson DF et al. (2016) Enzalutamide versus Bicalutamide in Castration-Resistant Prostate Cancer: The STRIVE Trial. J Clin Oncol. 34 (18): 2098-106.
7. Robins JM, Tsiatis AA Correcting For Non-Compliance in Randomized Trials Using Rank Preserving Structural Failure Time Models. Commun. Statist.-Theory Meth. 1991; 20(8):2609-2631.
8. Smith MR et al. (2012) Denosumab and bone-metastasis-free survival in men with castration-resistant prostate cancer: results of a phase 3, randomised, placebo-controlled trial. Lancet. 379 (9810): 39-46.
9. U.S. Department of Health and Human Services, Food and Drug Administration, Guidance for Industry Drug-Induced Liver Injury: Premarketing Clinical Evaluation, July 2009.
10. Hung HM et al. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. J Biopharm Stat. 2007; 17(6):1201-10.

9. Appendices

9.1 Study event schedule

Study event schedule during the double-blind study treatment period

Protocol activities	**Screening period**	**Study treatment period (Darolutamide (ODM-201) or placebo)**				
		Visit 1	**Visit 2**	**Visit 3**	**Visit 4 and subsequent visits**	**End-of-study treatment visit**
	Within 28 days before randomization	Day 1[1]	Day 15 (±3 days)	Day 29 (±5 days)	Week 16 (±7 days) and every subsequent 16 weeks	28 days (+7 days) after last dose
FACT-P or PCS of FACT-P	x (FACT-P)	x (FACT-P)			x[2]	x (FACT-P)
EQ-5D-3L	x	x			x[3]	x[3]
Physical examination, weight	x[4]	x		x	x	x
12-lead ECG and vital signs	x[5]	x[5]	x	x	x	x
Laboratory safety assessments	x	x	x	x	x	x
Serum PSA[14]	x	x			x	x
Testosterone	x	x			x	
PK samples at participating centres[6]			x	x	x	

Protocol activities	**Screening period**	**Study treatment period (Darolutamide (ODM-201) or placebo)**				
		Visit 1	**Visit 2**	**Visit 3**	**Visit 4 and subsequent visits**	**End-of-study treatment visit**
PK diary dispensing[6]		x	x	x		
PK diary collection[7]			x	x	x	
ECOG performance status	x				x	x
Pain diary dispensing	x			x	x	x
Pain diary collection[7]		x			x	x
Chest, abdomen and pelvic CT/MRI	x[8]				x[8]	
Bone scan[9]	x[8]				x[8]	
First SSE			x	x	x	x
First prostate cancer-related invasive procedure			x	x	x	x
Survival status			x	x	x	x
Adverse events[10]	x	x	x	x	x	x
Concomitant treatments[11]	x	x	x	x	x	x
Study drug dispensing and accountability		x		x	x	x[15]

Protocol activities	Screening period	Study treatment period (Darolutamide (ODM-201) or placebo)				
		Visit 1	Visit 2	Visit 3	Visit 4 and subsequent visits	End-of-study treatment visit
PG IC	x[12]					
PG sample		x[16]				
Tumor-related biomarker blood sample[13]		x				x

Abbreviations: IC = informed consent; ECG = electrocardiogram; ECOG = Eastern Cooperative Oncology Group; PG = pharmacogenomics; HR = heart rate; PK = pharmacokinetic; PSA = prostate-specific antigen; SSE = symptomatic skeletal event

[1)] In case the start of study treatment is not logistically feasible on the same day as randomization can be performed 1 day before the initiation of study treatment.

[2)] FACT-P (Appendix 5a) will be assessed at the week 16 visit only, and PCS subscale of FACT-P (Appendix 5b) will be assessed at the subsequent 16-week visits thereafter.

[3)] EQ-5D-3L (Appendix 6) will be assessed at the week 16 visit only, and end-of-study treatment visit.

[4)] Also height will be recorded at the screening visit.

[5)] A triplicate 12-lead ECG including HR in a supine position after at least 10 minutes rest. The 3 consecutive recordings will be performed within approximately 5 minutes. BP and HR (unless recorded by ECG) in a supine position after at least 10 minutes rest. The assessments will be done in the following order: ECGs, vital signs, and any type of blood draw as the last assessment.

[6)] See Section 6.2 for PK procedures (blood sampling and recording of 12-lead ECG and vital signs) at study centers participating in PK sample collection. If PK sampling is not feasible during these visits, later visits may be used for PK procedures.

[7)] Patients will be instructed to complete the diary for 6 days (-6 to -1 day) before day 1 and every 16-week visits until documented pain progression.

[8)] CT/MRI and bone scans are acceptable if performed within 42 days prior to start of study treatment. A local qualified site physician (e.g. site radiologist or PI, at the PI's discretion) must first review the screening scans and confirm the patient is non-metastatic before submitting the scans to central review to confirm eligibility. If the local qualified site physician detects metastasis, the scans should not be submitted to central review. For guidance on CT/MRI and bone scans performed at Week 16 and subsequent visits see Section 6.1.2.4.

[9)] New lesion(s) in bone scan must be confirmed by CT/MRI (whichever imaging was conducted at screening for soft tissue) or x-ray. If confirmatory scan is negative (does not confirm progression), the patient will continue study treatment and visit study center as specified in the study event table.

[10)] All AEs including SAEs will be recorded from the date of signing IC and continuing until the end-of-study treatment visit. Any AEs or SAEs that occur after the end-of-study treatment visit and which are considered to be related to the study treatment or any study procedures have to be reported and entered into the eCRF. When the eCRF is not available anymore, all SAEs considered study treatment related should be reported to Bayer using a paper SAE form. AEs related to study procedures and all SAEs will be recorded for screening failures.

[11)] Once the patient has been withdrawn from study treatment, concomitant treatments will be recorded if used to treat new related or unresolved related AEs or if it is an antineoplastic therapy.

[12)] An optional PG IC for pharmacogenetics research study will be obtained at screening visit (or later during the study if feasible) unless precluded by local guidelines, e.g. IEC/IRB or regulatory authorities.

[13)] Blood (plasma) sampling for tumor-related biomarker evaluation will be collected on day 1 visit before study treatment administration and at the end-of-study treatment visit unless precluded by local guidelines, e.g. IEC/IRB or regulatory authorities.

[14)] The local PSA result taken at screening should be used to confirm the patient has CRPC (inclusion criterion 4) and meets inclusion criterion 6.

[15)] Only study drug accountability performed at the end-of-study treatment visit.

[16)] A blood sample for DNA extraction will be taken at day 1 visit (or later during the study if feasible) only from patients who have signed the PG IC unless precluded by local guidelines, e.g. IEC/IRB or regulatory authorities.

Study event schedule during the follow-up period after withdrawal of study treatment

Protocol activities	For patients during double-blind period before confirmed metastasis and not initiating any prohibited treatment Visits Every 16 weeks (±7 days) from end-of-study treatment visit[1]	For patients after confirmed metastasis or for patients in the placebo arm at unblinding Visits or phone contacts Every 16 weeks (±7 days) from end-of-study treatment visit[2]
BPI-SF	X	X
Dispensing of pain diary	X	X
Collection of pain diary and pain medications	X[3]	X[3]
PCS subscale of FACT-P	X	X
ECOG performance status	X	X
Use of subsequent antineoplastic therapies	X	X
Collection of AEs and SAEs considered to be related to study treatment or procedures	X	X
Collection of concomitant treatments used to treat new related or unresolved related AEs	X	X
First SSE	X	X
First prostate cancer-related invasive procedure for prostate cancer	X	X

Protocol activities	For patients during double-blind period before confirmed metastasis and not initiating any prohibited treatment Visits Every 16 weeks (±7 days) from end-of-study treatment visit[1]	For patients after confirmed metastasis or for patients in the placebo arm at unblinding Visits or phone contacts Every 16 weeks (±7 days) from end-of-study treatment visit[2]
Survival status	X	X
Serum PSA	X	
Chest, abdomen, and pelvic CT/MRI	X	
Bone scan[4]	X	

[1)] Until confirmed metastasis, death or the end of the study
[2)] Until death or the end of the study
[3)] Patients will be instructed to complete the diary for 6 days before each 16-week visit until documented pain progression. Pain medications will be recorded on CRFs at each 16-week visit until documented pain progression.
[4)] New lesion(s) in bone scan must be confirmed by CT/MRI (whichever imaging was conducted at screening for soft tissue) or x-ray. If confirmatory scan is negative (does not confirm progression), the patient will continue to visit study center as specified in the study event table.

9.2 Brief pain Inventory – Short Form (BPI-SF) questionnaire and scoring information

STUDY ID# _____ HOSPITAL # _____

DO NOT WRITE ABOVE THIS LINE

Brief Pain Inventory (Short Form)

Date: _____/_____/_____ Time: _____

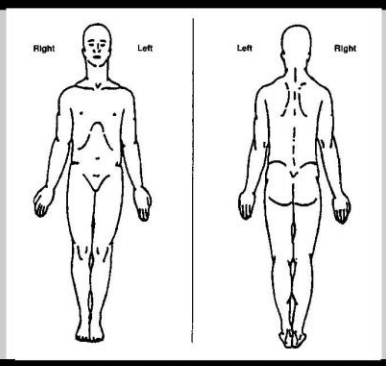
Name: _____ _____ _____

Last First Middle Initial

1. Throughout our lives, most of us have had pain from time to time (such as minor headaches, sprains, and toothaches). Have you had pain other than these every-day kinds of pain today?

   1. Yes 2. No

2. On the diagram, shade in the areas where you feel pain. Put an X on the area that hurts the most.



3. Please rate your pain by circling the one number that best describes your pain at its worst in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

4. Please rate your pain by circling the one nuimber that best describes your pain at its least in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

5. Please rate your pain by circling the one number that best describes your pain on the average.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

6. Please rate your pain by circling the one number that tells how much pain you have right now.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

7. What treatments or medications are you receiving for your pain?

8. In the last 24 hours, how much relief have pain treatments or medications provided? Please circle the one percentage that most shows how much relief you have received.

0% No Relief	10%	20%	30%	40%	50%	60%	70%	80%	90%	100% Complete Relief

9. Circle the one number that describes how, during the past 24 hours, pain has interfered with your:

A. General Activity

0 Does not Interfere	1	2	3	4	5	6	7	8	9	10 Completely Interferes

B. Mood

0 Does not Interfere	1	2	3	4	5	6	7	8	9	10 Completely Interferes

C. Walking Ability

0 Does not Interfere	1	2	3	4	5	6	7	8	9	10 Completely Interferes

D. Normal Work (includes both work outside the home and housework)

0 Does not Interfere	1	2	3	4	5	6	7	8	9	10 Completely Interferes

E. Relations with other people

0 Does not Interfere	1	2	3	4	5	6	7	8	9	10 Completely Interferes

F. Sleep

0 Does not Interfere	1	2	3	4	5	6	7	8	9	10 Completely Interferes

G. Enjoyment of life

0 Does not Interfere	1	2	3	4	5	6	7	8	9	10 Completely Interferes

Scoring:

Pain Severity Score = mean of items 3-6
(pain at its worst, pain at its least, pain on the average, pain for right now)

Pain Interference Score = mean of items 9A-9G (interference of pain with: general activity, mood, walking ability, normal work, relations with other people, sleep, and enjoyment of life).

9.3 FACT-P questionnaire and scoring information

PCS subscale of FACT-P concerns the “additional concerns” questions.

FACT-P (Version 4)

Below is a list of statements that other people with your illness have said are important. **By circling one (1) number per line, please indicate how true each statement has been for you during the past 7 days.**

	PHYSICAL WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GP1	I have a lack of energy	0	1	2	3	4
GP2	I have nausea	0	1	2	3	4
GP3	Because of my physical condition, I have trouble meeting the needs of my family	0	1	2	3	4
GP4	I have pain	0	1	2	3	4
GP5	I am bothered by side effects of treatment	0	1	2	3	4
GP6	I feel ill	0	1	2	3	4
GP7	I am forced to spend time in bed	0	1	2	3	4

	SOCIAL/FAMILY WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GS1	I feel close to my friends	0	1	2	3	4
GS2	I get emotional support from my family	0	1	2	3	4
GS3	I get support from my friends	0	1	2	3	4
GS4	My family has accepted my illness	0	1	2	3	4
GS5	I am satisfied with family communication about my illness	0	1	2	3	4
GS6	I feel close to my partner (or the person who is my main support)	0	1	2	3	4
Q1	*Regardless of your current level of sexual activity, please answer the following question. If you prefer not to answer it, please check this box* ☐ *and go to the next section.*					
GS7	I am satisfied with my sex life	0	1	2	3	4

FACT-P (Version 4)

By circling one (1) number per line, please indicate how true each statement has been for you during the past 7 days.

	EMOTIONAL WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GE1	I feel sad	0	1	2	3	4
GE2	I am satisfied with how I am coping with my illness	0	1	2	3	4
GE3	I am losing hope in the fight against my illness	0	1	2	3	4
GE4	I feel nervous	0	1	2	3	4
GE5	I worry about dying	0	1	2	3	4
GE6	I worry that my condition will get worse	0	1	2	3	4

	FUNCTIONAL WELL-BEING	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
GF1	I am able to work (include work at home)	0	1	2	3	4
GF2	My work (include work at home) is fulfilling	0	1	2	3	4
GF3	I am able to enjoy life	0	1	2	3	4
GF4	I have accepted my illness	0	1	2	3	4
GF5	I am sleeping well	0	1	2	3	4
GF6	I am enjoying the things I usually do for fun	0	1	2	3	4
GF7	I am content with the quality of my life right now	0	1	2	3	4

FACT-P (Version 4)

By circling one (1) number per line, please indicate how true each statement has been for you during the past 7 days.

	ADDITIONAL CONCERNS	**Not at all**	**A little bit**	**Some-what**	**Quite a bit**	**Very much**
C2	I am losing weight	0	1	2	3	4
C6	I have a good appetite	0	1	2	3	4
P1	I have aches and pains that bother me	0	1	2	3	4
P2	I have certain parts of my body where I experience significant pain	0	1	2	3	4
P3	My pain keeps me from doing things I want to do	0	1	2	3	4
P4	I am satisfied with my present comfort level	0	1	2	3	4
P5	I am able to feel like a man	0	1	2	3	4
P6	I have trouble moving my bowels	0	1	2	3	4
P7	I have difficulty urinating	0	1	2	3	4
BL2	I urinate more frequently than usual	0	1	2	3	4
P8	My problems with urinating limit my activities	0	1	2	3	4
BL5	I am able to have and maintain an erection	0	1	2	3	4

FACT-P Scoring Guidelines (Version 4)

Instructions:
1. Record answers in "item response" column. If missing, mark with an X

2. Perform reversals as indicated, and sum individual items to obtain a score.

3. Multiply the sum of the item scores by the number of items in the subscale, and then divide by the number of items answered. This produces the subscale score.

4. Add subscale scores to derive total scores (TOI, FACT-G & FACT-P).

5. The higher the score, the better the QOL.

Subscale	**Item Code**	**Reverse item?**		**Item response**	**Item Score**
PHYSICAL	GP1	4	-	_____	=_____
WELL-BEING	GP2	4	-	_____	=_____
(PWB)	GP3	4		_____	=_____
	GP4	4	-	_____	=_____
Score range: 0-28	GP5	4	-	_____	=_____
	GP6	4	-	_____	=_____
	GP7	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____= **PWB subscale score**

SOCIAL/FAMILY	GS1	0	+	_____	=_____
WELL-BEING	GS2	0	+	_____	=_____
(SWB)	GS3	0	+	_____	=_____
	GS4	0	+	_____	=_____
Score range: 0-28	GS5	0	+	_____	=_____
	GS6	0	+	_____	=_____
	GS7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____= **SWB subscale score**

EMOTIONAL	GE1	4	-	_____	=_____
WELL-BEING	GE2	0	+	_____	=_____
(EWB)	GE3	4	-	_____	=_____
Score range: 0-24	GE4	4	-	_____	=_____
	GE5	4	-	_____	=_____
	GE6	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 6: _____
Divide by number of items answered: _____= **EWB subscale score**

FUNCTIONAL	GF1	0	+	_____	=_____
WELL-BEING	GF2	0	+	_____	=_____
(FWB)	GF3	0	+	_____	=_____
Score range: 0-28	GF4	0	+	_____	=_____
	GF5	0	+	_____	=_____
	GF6	0	+	_____	=_____
	GF7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____= **FWB subscale score**

FACT-P Scoring Guidelines (Version 4) – Page 2

Subscale	**Item Code**	**Reverse item?**		**Item responds**	**Item Score**
PROSTATE	C2	4	-	_____	=_____
CANCER	C6	0	+	_____	=_____
SUBSCALE	P1	4	-	_____	=_____
(PCS)	P2	4	-	_____	=_____
	P3	4	-	_____	=_____
Score range: 0-48	P4	0	+	_____	=_____
	P5	0	+	_____	=_____
	P6	4	-	_____	=_____
	P7	4	-	_____	=_____
	BL2	4	-	_____	=_____
	P8	4	-	_____	=_____
	BL5	0	+	_____	=_____

Sum individual item scores: _____

Multiply by 12: _____

Divide by number of items answered: _____**=PC Subscale score**

To derive a FACT-P Trial Outcome Index (TOI):
Score range: 0-104
_____ + _____ + _____ =_____=**FACT-P TOI**

(**PWB** score) (**FWB** score) (**PCS** score)

To Derive a FACT-G total score:
Score range: 0-108
_____ + _____ + _____ + _____=_____=**FACT-G Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score)

To Derive a FACT-P total score:
Score range: 0-156

_____ + _____ + _____ + _____ + _____ =_____=**FACT-P Total score**

(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score) (**PCS** score)

*For guidelines on handling missing data and scoring options, please refer to the Administration and Scoring Guidelines in the manual or on-line at www.facit.org.

9.4 EORTC-QLQ-PR25 questionnaire and scoring information

ENGLISH

EORTC QLQ - PR25

Patients sometimes report that they have the following symptoms or problems. Please indicate the extent to which you have experienced these symptoms or problems during the past week. Please answer by circling the number that best applies to you.

During the past week	Not at all	A little	Quite a bit	Very much
31. Have you had to urinate frequently **during the day**?	1	2	3	4
32. Have you had to urinate frequently **at night**?	1	2	3	4
33. When you felt the urge to pass urine, did you have to hurry to get to the toilet?	1	2	3	4
34. Was it difficult for you to get enough sleep, because you needed to get up frequently at night to urinate?	1	2	3	4
35. Have you had difficulty going out of the house because you needed to be close to a toilet?	1	2	3	4
36. Have you had any unintentional release (leakage) of urine?	1	2	3	4
37. Did you have pain when you urinated?	1	2	3	4
38. Answer this question only if you wear an incontinence aid. Has wearing an incontinence aid been a problem for you?	1	2	3	4
39. Have your daily activities been limited by your urinary problems?	1	2	3	4
40. Have your daily activities been limited by your bowel problems?	1	2	3	4
41. Have you had any unintentional release (leakage) of stools?	1	2	3	4
42. Have you had blood in your stools?	1	2	3	4
43. Did you have a bloated feeling in your abdomen?	1	2	3	4
44. Did you have hot flushes?	1	2	3	4
45. Have you had sore or enlarged nipples or breasts?	1	2	3	4
46. Have you had swelling in your legs or ankles?	1	2	3	4

Please go to the next page

ENGLISH

During the last 4 weeks…	Not at all	A little	Quite a bit	Very much
47. Has weight **loss** been a problem for you?	1	2	3	4
48. Has weight **gain** been a problem for you?	1	2	3	4
49. Have you felt less masculine as a result of your illness or treatment?	1	2	3	4
50. To what extent were you interested in sex?	1	2	3	4
51. To what extent were you sexually active (with or without intercourse)?	1	2	3	4

PLEASE ANSWER THE NEXT FOUR QUESTIONS ONLY IF YOU HAVE BEEN SEXUALLY ACTIVE OVER THE LAST 4 WEEKS

52. To what extent was sex enjoyable for you?	1	2	3	4
53. Did you have difficulty getting or maintaining an erection?	1	2	3	4
54. Did you have ejaculation problems (eg dry ejaculation)?	1	2	3	4
55. Have you felt uncomfortable about being sexually intimate?	1	2	3	4

Questions are re-numbered from 1 to 25.

Scoring

	Scale name	Number of items	Item range	QLQ-PR25 item numbers
Functional scales				
Sexual activity	PRSAC	2	3	20,21
Sexual functioning	PRSFU	4	3	22 – 25
Symptom scales				
Urinary symptoms	PRURI	8	3	1 – 7,9
Bowel symptoms	PRBOW	4	3	10 – 13
Hormonal treatment-related symptoms	PRHTR	6	3	14 – 19
Incontinence aid	PRAID	1	3	8

Item range is the difference between the possible maximum and the minimum response to individual items. Values from 1 to 4 gives range=3.

Remarks

- Items 20 and 21 can be completed by all patients
- Items 22-25 are conditional on being sexually active, and thus will only be completed by a subgroup of patients. This will require reversing the response categories of questions 23-25 but not of 22.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$RawScore = RS = (I1 + I2 + ... + In)\ n$

Then for **Functional scales**:

$Score = 1 - ((RS - 1)\ /range) \times 100$

and for **Symptom scales / items:**

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

Missing items are simply ignored when making the calculations for scores, the above equations for multi-item scales can be used whenever at least half of the items are completed.

All of the scales and single-item measures range in score from 0 to 100. A high scale score represents a higher response level.

9.5 EQ-5D-3L questionnaire and scoring information

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

Mobility

I have no problems in walking about ❑

I have some problems in walking about ❑

I am confined to bed ❑

Self-Care

I have no problems with self-care ❑

I have some problems washing or dressing myself ❑

I am unable to wash or dress myself ❑

Usual Activities (*e.g. work, study, housework, family or leisure activities)*

I have no problems with performing my usual activities ❑

I have some problems with performing my usual activities ❑

I am unable to perform my usual activities ❑

Pain/Discomfort

I have no pain or discomfort ❑

I have moderate pain or discomfort ❑

I have extreme pain or discomfort ❑

Anxiety/Depression

I am not anxious or depressed ❑

I am moderately anxious or depressed ❑

I am extremely anxious or depressed ❑

Values for the 243 health states defined by the EuroQol classification have been calculated using a regression model. The following worked example indicates how these coefficients are to be used so as to compute the estimated values for each state.

Calculating EQ-5D-3L state scores - a worked example

EuroQol dimension	Level 2	Level 3
Mobility	0.069	0.314
Self-care	0.104	0.214
Usual activity	0.036	0.094
Pain / discomfort	0.123	0.386
Anxiety / depression	0.071	0.236
	Constant = 0.081	N3 = 0.269

The arithmetic needed to recover the estimated value for any health state from this table of decrements is given by the following example:

- Taking health state 1 1 2 2 3
- Full health (1 1 1 1 1) = 1.0
- Constant term (for any dysfunctional state)(subtract 0.081)
- Mobility. level 1 (subtract 0)
- Self-care. level 1 (subtract 0)
- Usual activity. level 2 (subtract 0.036)
- Pain / discomfort. level 2 (subtract 0.123)
- Anxiety / depression. level 3 (subtract 0.236)
- Level 3 occurs within at least 1 dimension (subtract N3 parameter 0.269)
- Hence the estimated value for state 1 1 2 2 3 is given by
  1.0- 0.081 - 0.036 - 0.123 - 0.236 - 0.269 = .25

9.6 Prohibited medication: metastasis-free survival

Following medications considered prohibited was selected in CM dataset:

- Radiopharmaceuticals (ATC codes V10BX, V10XX03),
- Immunotherapy (e.g. sipuleul T) (ATC code L03),
- Cytotoxic chemotherapy and any other systemic antineoplastic therapy (ATC code class L: antineoplastic and immunodulating agents , with the exception of L02AE (Gonadotropin-releasing hormone analogues)),
- Enzalutamide, ARN-509, bicalutamide, flutamide, nilutamide (ATC code L02B),
- Cyproterone acetate estrogen (ATC codes G03H, G03C),
- 5 α-reductase inhibitor (ATC code G04CB),
- Abiraterone acetate, TAK-700 or other CYP17 inhibitors (ATC code L02B),
- Systemic ketoconazole (as antineoplastic therapy) (ATC codes D01AC08; G01AF11, J02AB02),
- Osteoclast-targeted therapy such as bisphosphonate or denosumab (ATC code M05B) given for preventing skeletal-related events. These drugs are allowed for treatment of osteoporosis,
- Continuous use of systemic corticosteroid (ATC code H02).

Sponsor's assessment of prohibited medication is presented in a supplementary document.

9.7 Cytotoxic chemotherapy: time to cytotoxic chemotherapy

The cytotoxic chemotherapy was selected in CM dataset (ATC code L01A, L01B, L01C, L01D, and L01X).). Sponsor's assessment of cytotoxic chemotherapy is presented in a supplementary document.

9.8 Antineoplastic therapy: time to initiation of subsequent antineoplastic therapy

- Antineoplastic therapy treatment was selected in CM dataset
- ATC code class L (antineoplastic and immunodulating agents): L01 Antineoplastic agents (except cytotoxic chemotherapy L01A, L01B, L01C, L01D and L01X), L02 endocrine therapy and L03 immunostimulants
- ATC code class H: H02Corticosteroids for systemic use.

Sponsor's assessment of antineoplastic therapy is presented in a supplementary document.

9.9 Statins medications

Statins will be identified using the Standardized Drug Grouping (SDG) classification 'Statins'.

9.10 Opioid for cancer pain

The opioid treatments were selected in CM dataset (ATC code starting with N02A).

Sponsor's assessment of opioid treatments is presented in a supplementary document.

9.11 Antiandrogen (AR inhibitors) therapy

The antiandrogen (AR inhibitors) therapy treatments were selected in CM dataset (ATC code starting with L02BB and G03HA01).
L02BB antiandrogens (L02BB01 flutamide, L02BB02 nilutamide, L02BB03 bicalutamide, L02BB04 enzalutamide), G03HA antiandrogens plain (G03HA01 cyproterone).

9.12 GnRH agonist/antagonist therapy

The GnRH antagonist treatments were selected in CM dataset (ATC code starting with L02AE, H01CA and L02BX).

L02AE Gonadotropin releasing hormone analogues (L02AE01 buserelin, L02AE02 leuprorelin, L02AE03 goserelin, L02AE04 triptorelin, L02AE05 histrelin), H01CA Gonadotropin releasing hormones, L02BX Other hormone antagonists and related agents (L02BX01 abarelix, L02BX02 degarelix, L02BX03 abiraterone (except if used as previous therapy).

9.13 Hormonal therapy

Hormonal therapy will be were selected in CM dataset:

The GnRH antagonist treatments (ATC code starting with L02AE, H01CA and L02BX).

L02AE Gonadotropin releasing hormone analogues (L02AE01 buserelin, L02AE02 leuprorelin, L02AE03 goserelin, L02AE04 triptorelin, L02AE05 histrelin), H01CA Gonadotropin releasing hormones, L02BX Other hormone antagonists and related agents (L02BX01 abarelix, L02BX02 degarelix, L02BX03 abiraterone (except if used as previous therapy).

The antiandrogen (AR inhibitors) therapy treatments (ATC code starting with L02BB and G03HA01).

L02BB antiandrogens (L02BB01 flutamide, L02BB02 nilutamide, L02BB03 bicalutamide, L02BB04 enzalutamide), G03HA antiandrogens plain (G03HA01 cyproterone).

9.14 Bone sparing agent

Bone sparing agent will be were selected in CM dataset, using ATC codes:

M05B drugs affecting bone structure and mineralization, A11CC Vitamin D and analogues, A12A Calcium, A12CD Fluoride, H05BA Calcitonins.