Document Type:	Statistical Analysis Plan
Official Title:	A phase III randomized, double-blind, placebo-controlled trial of radium-223 dichloride in combination with abiraterone acetate and prednisone/prednisolone in the treatment of asymptomatic or mildly symptomatic chemotherapy-naïve subjects with bone predominant metastatic castration-resistant prostate cancer (CRPC)
NCT Number:	NCT02043678
Document Date:	10-JUL-2017

Title page - amended

A phase III randomized, double-blind, placebo-controlled trial of radium-223 dichloride in combination with abiraterone acetate and prednisone/prednisolone in the treatment of asymptomatic or mildly symptomatic chemotherapy-naïve subjects with bone predominant metastatic castration-resistant prostate cancer (CRPC)

Short title: Evaluation of **R**adium-223 dichloride in combination with **A**biraterone in CRPC - ERA 223

BSP Study drug	BAY 88-8223/Radium-223 dichloride
Study purpose:	To compare the symptomatic skeletal event-free survival of subjects with asymptomatic or mildly symptomatic chemotherapy-naïve bone predominant metastatic castration-resistant prostate cancer (CRPC) treated with radium-223 dichloride or placebo, in combination with abiraterone and prednisone/prednisolone

Clinical study phase:	III	**Date:**	July 10th 2017
Study No.:	BAY 88-8223 / 15396	**Version:**	V5.0

Author:

PPD
PPD
PPD
Bayer HealthCare
Pharmaceuticals
Whippany, NJ 07981

PPD
PPD
Bayer HealthCare
Pharmaceuticals
Whippany, NJ 07981

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

Abbreviations

AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
BPI-SF	Brief Pain Inventory – Short Form
Ca	Calcium
Cl	Chloride
C_{max}	Maximum plasma concentration
CRPC	Castration -resistant prostate cancer
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events; version 4.03
EBRT	External beam radiation therapy
ECOG	Eastern Co-operative Oncology Group
ECOG PS	Eastern Co-operative Oncology Group Performance Status
EOD	Extent of disease
eCRF	Electronic case report form
ePRO	Electronic subject report outcome
EQ-5D	European Quality of Life - 5 Dimensions
EU	European Union
GCP	Good Clinical Practice
Hb	Hemoglobin
HR	Hazard ratio
HRQoL	Health related quality of life
ICF	Informed consent form
IDMC	Independent data monitoring committee
IMP	Investigational medicinal product
ITT	Intent-to-Treat
IV	Intravenous
IXRS	Interactive Voice/Web Response System
K	potassium
kBq	KiloBecquerel; SI unit of radioactivity
kg	Kilogram
LDH	Lactate dehydrogenase
mCi	Millicuries
MedDRA	Medical Dictionary for Regulatory Activities
mL	Milliliter
MRI	Magnetic resonance imaging
Na	Sodium
NCA	Non-compartmental analysis
NCI	National Cancer Institute
OS	Overall survival

PD	Progressive disease
PS	Performance status
PSA	Prostate specific antigen
qd	Every day (quaque die)
QoL	Quality of life
rPFS	Radiological progression-free survival
PKS	Pharmacokinetic analysis set
RBC	Red blood cell
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Statistical analysis software
SSE	Symptomatic skeletal event
SSE-FS	Symptomatic skeletal event-free survival
ULN	Upper limit of normal
WBC	White blood cell
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary
WPS	Worst pain subscale

Definitions of terms

Radium-223 dichloride	The investigational product, a targeted alpha particle emitting radiopharmaceutical, is a ready-to-use solution for intravenous injection containing the drug substance radium-223 dichloride. The active moiety is the alpha particle emitting nuclide radium-223, present as a divalent cation ($^{223}Ra^{2+}$).
Dose	Doses are given as kiloBecquerel (kBq) per kilogram body weight, with the corresponding dose given in millicurie (mCi) per kilogram in parenthesis. The term “dose” is used to describe the quantity of radioactivity from radium-223 administered.
Bone scan	Whole body technetium-99m bone scan

1. Study Objectives

1.1 Primary Objective

The primary objective is to compare, in subjects with asymptomatic or mildly symptomatic chemotherapy-naïve bone predominant metastatic CRPC, the clinical benefit of radium-223 dichloride versus placebo in combination with abiraterone and prednisone/prednisolone, with the primary efficacy endpoint being:

- Symptomatic skeletal event-free survival (SSE-FS)

1.2 Secondary Objective

The secondary objectives are to compare the radium-223 dichloride and placebo treatments to establish additional clinically relevant improvements in subjects with CRPC bone predominant metastasis using the variables below:

- OS
- Time to opiate use for cancer pain
- Time to pain progression
- Time to cytotoxic chemotherapy
- rPFS
- Safety, acute and long term, including new primary malignancies and hematopoietic reserve for tolerability of subsequent chemotherapy

1.3 Exploratory Objective

The study will also include the following exploratory endpoints:

- Time to first on-study SSE
- Percentage change in total ALP from baseline
- Time to ALP progression
- Time to PSA progression
- ALP response

- PSA response
- Bone scan-specific rPFS
- Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during the treatment period
- Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during period between start of treatment and end of active follow-up with clinic visits
- Assessment of pharmacokinetics (PK) of abiraterone
- Resource utilization
- Biomarker assessments

An additional explorative objective is to evaluate the impact of baseline total body weight (TBW) and ideal body weight (IBW) on SSE-FS and adverse events.

2. Study Design

2.1 Design overview

This study is a phase III multinational, multicenter, randomized, double blind, placebo-controlled, parallel group study with a randomization allocation ratio of 1:1 (radium-223 dichloride plus abiraterone acetate plus prednisone/prednisolone : placebo plus abiraterone acetate plus prednisone/prednisolone). Randomization will be stratified by:

- Geographical regions (Western Europe/North America/Australia vs. Asia vs. rest of world)
- Concurrent use of denosumab or bisphosphonates or none
- Total ALP < 90 U/L versus total ALP ≥ 90 U/L

To prevent sparse data, strata for current use of denosumab and current use of bisphosphonates will be combined in all stratified analysis for this study.

This study will be conducted at approximately 170 investigative study centers and approximately 800 subjects will be enrolled.

2.2 Determination of sample size

Sample size is calculated based on the primary endpoint SSE-FS. Using a test with a two-sided alpha of 0.05, power of 90%, and a randomization ratio of 1:1 between the experimental (radium-223 dichloride) and control (placebo) arms, 389 events are required to detect a 39% increase in SSE-FS (i.e., an overall 0.05 level two-sided log rank test has approximately 90% power to detect a difference between the two SSE-FS curves if the alternative hypothesis HR is 0.72 [assuming the median SSE-FS is 29.2 months for radium-223 dichloride vs. 21.0 for control]). The expected study duration for SSE-FS is 37 months, assuming subjects enroll at a rate of 50 subjects per month, an enrollment ramp-up time of 9 months, a dropout rate of 10%, exponentially distributed event time, 21 months in SSE-FS median time for the control group, and a total of 800 subjects in the 2 treatment groups combined.

This study is also powered (~70%) for the analysis of OS if OS is tested at a 0.05 level (2-sided) For the concluding analysis of OS, 500 deaths are projected to occur by approximately 71.4 months after the first subject is randomized, assuming the median OS for the control arm is 35.3 months and a 25% improvement for the radium-223 dichloride arm. If the final analysis of OS after 500 deaths reveals that the experimental treatment is statistically significantly better than treatment with control, then the OS endpoint will be declared positive for the final analysis.

3. General Statistical Considerations

3.1 General Principles

The statistical evaluation will be performed by using the software package SAS version 9.2 or higher (SAS Institute Inc., Cary, NC, USA). Unless otherwise noted, data will be analyzed by descriptive statistical methods: The number of data available and missing data, mean, standard deviation, minimum, quartiles, median, and maximum will be calculated for metric data. Frequency tables will be generated for categorical data.

Definition of efficacy and safety endpoints, analysis strategies, structure of analysis datasets and layout of analysis data displays are following Bayer Healthcare Pharmaceuticals (BHP) standards: Xofigo Project Standards, the Therapeutic Area Oncology Standards (TAS) and the Global Medical Standards (GMS), respectively. Where the given ordering reflects the priority of the different standards, means specifications of the latter ones have to be followed only if not specified in standards mentioned before.

3.2 Adjustments for Covariates

The three categorical variables that define the randomization strata: geographical regions (Western Europe/North America/Australia vs. Asia vs. rest of world), concurrent use of denosumab or bisphosphonates (yes vs. no), and total ALP < 90 U/L versus total ALP ≥ 90 U/L will be accounted for in all statistical tests and models using appropriate methods as defined in Section 5.2.

The following additional baseline covariates which are considered to be potentially prognostic will also be included in the statistical modeling of the primary efficacy data, as an exploratory analysis:

- Baseline albumin value (< or ≥ median);
- Baseline hemoglobin value (< or ≥ median);
- Baseline LDH value (< or ≥ median);
- Baseline ECOG performance status (0, 1);
- Baseline PSA value (< or ≥ median);
- Age (<65, ≥65).

A step-wise selection method will be used to choose the final statistical model with entry alpha level 0.1 and exit alpha level 0.1. The stratification factors are always included as covariates in the final model. If deemed necessary, additional variables may also be added in the model selection process.

3.3 Handling of Dropouts

A "dropout" is defined as a subject who has been randomized and discontinues study participation prematurely for any reason. Subjects withdrawn from study treatment will not be replaced. Refer to Section 5.2.2 in the study protocol for withdrawal of subjects from study.

All efficacy analyses are based on the intent-to-treat population, that comprises all randomized subjects, including subjects who withdraw regardless of the reason for withdrawal. See following chapters for more details on deriving efficacy endpoints in case of missing data.

3.4 Handling of Missing Data

In order to achieve the goal of a well conducted clinical trial according to ICH Good Clinical Practice (ICH-GCP), every effort should be made to collect all data. However, despite best efforts, it may be inevitable that missing or incomplete data are reported. All missing or partial data will be presented in the subject data listing, as they are recorded on the CRF. Except as noted, missing data will not be imputed or carried forward in any statistical analysis.

Adverse event and concomitant medication partial missing start/stop dates will be imputed and the imputation rule will be specified in the data specification.

3.5 Interim Analyses and Data Monitoring

3.5.1 Interim Analyses

There is no formal interim analysis for efficacy planned for the primary endpoint SSE-FS.

For the secondary efficacy endpoint of OS, one interim analysis (to be performed at the same time as the final SSE-FS analysis) and one final analysis are planned. At the interim OS analysis, it is expected that 275 deaths will have occurred (assuming 35.3 months median survival in the control arm).

If SSE-FS is statistically significant at the 0.05 level (2-sided) and rPFS, time to pain progression and time to cytotoxic chemotherapy are all statistically significant at the 0.049 level (2-sided), then the OS will be formally tested at an overall alpha=0.05 (2-sided). If the analysis of OS after 275 deaths following treatment with radium-223 dichloride plus abiraterone plus prednisone/prednisolone is statistically significantly better compared to control (two-sided $p \leq 0.005$, based on O'Brien-Fleming alpha spending function (1, 2)), then OS will be declared positive for the interim analysis, assuming the final event number for OS is 500. The actual nominal alpha levels will be calculated based on the actual number of events accrued at the OS interim analysis. If the interim OS analysis is not statistically significant, the final analysis for OS will be performed when approximately 500 deaths have occurred, corresponding to an approximately 70% power to detect a 25% improvement (in OS with Ra-223 Cl_2 compared with placebo) with a two-sided alpha of 0.05.

If at least one of the secondary endpoints, rPFS, time to pain progression or time to cytotoxic chemotherapy, is not statistically significant at the 0.049 level (2-sided), then the OS will be formally tested at an overall alpha=0.001 (2-sided). Refer to scenario 2 in section 5.2.2.2.1 for details.

3.5.2 Data Monitoring

As described previously in protocol Section **3.1**, an IDMC will be instituted for independent review of ongoing data from this trial in accordance with a separate IDMC Charter. The IDMC will operate independently of the sponsor and investigators.

An independent data monitoring committee (IDMC) will be established for this study to review accumulating efficacy and safety data at regular intervals throughout the study and monitor overall study conduct. The IDMC will include experts in oncology, biostatistics, and safety who are not participating in this trial and do not have affiliation with the Investigators or the Sponsor or other significant conflicts of interest. Their main objective will be to protect the interests of the subjects in the study and of those still to be entered. They will do this by monitoring the study periodically for safety, study progress, and protocol compliance, as well as assessing the risk/benefit of the trial. IDMC meetings will be held as per separate DMC charter, approximately every 6 months throughout the blinded trial phase. Ad hoc meetings will take place if needed.

Specific areas of concern for the IDMC are:

- Subject safety (unblinded safety data will be reviewed at each meeting and any cases of unexpected AEs will be considered)
- Accrual factors that may potentially impact on randomization balance

- Subject eligibility and adequacy of follow-up
- Protocol compliance (deviations from outcome assessment schedules which may lead to biases will be evaluated)

The specific duties of the IDMC as well as statistical monitoring guidelines and procedures are described in the IDMC Charter.

3.6 Data Rules

Generally, for each date stored in 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.

The baseline for all efficacy endpoints is defined as the last non-missing assessment taken on or before the randomization date.

Unless otherwise specified, the baseline for all safety data, is the last non-missing assessment taken on or before first treatment date.

4. Analysis Sets

4.1 Assignment of analysis sets

Intent-to-treat analysis set (ITT)

All randomized subjects will be included in the ITT analysis set. The ITT population will be used in the analysis of all efficacy endpoints. Subjects will be included in all ITT analyses according to the treatment to which they are randomized.

Safety analysis set (SAF)

All randomized subjects who received at least one dose of any study drug will be included in the safety analysis set. This safety population will be used in the analyses of all safety endpoints. Subjects will be included in the analyses according to the treatment they received.

Pharmacokinetic analysis set (PKS)

All subjects of Subgroup 1 defined in the protocol with a valid pharmacokinetic profile for non-compartmental analysis (NCA) will be included in the pharmacokinetic analysis set (PKS).

5. Statistical Methodology

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

5.1 Population Characteristics

5.1.1 Disposition of Subjects

The number and percentage of subjects screened, randomized, and treated will be presented by treatment group and overall. The number of subjects will also be tabulated by the total number of radium 223 injections received. The reasons for subjects discontinued from the treatment will be summarized by treatment group. In addition, the number of subjects screened, screen failures, and included in each analysis population will be displayed by region, country and center. The screen failure reasons will be summarized by treatment group.

5.1.2 Demographic and Baseline Characteristics

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

The following demographic data will be summarized:

- Sex
- Age at screening (years)
- Age category (< 65, 65 – 74, 75-84, ≥85 years)
- Race and ethnicity
- Height (cm)
- Weight (kg) at baseline
- BMI

The following baseline characteristics will be summarized:

- Stage of prostate cancer at diagnosis
- Treatment of prostate cancer before enrollment (e.g., surgery, radiation, etc.)
- Gleason score at diagnosis of prostate cancer
- PSA and total ALP at diagnosis of prostate cancer
- PSA and total ALP at randomization
- Vital signs: blood pressure (mm Hg), heart rate (bpm), respiratory rate (rpm), and temperature (°C)
- ECOG PS
- Cancer pain assessment

Categorical summaries of each of three randomization stratification factors: geographical regions (Western Europe/North America/Australia vs. Asia vs. rest of world), concurrent use of denosumab or bisphosphonates (yes vs. no), and total ALP (< 90 U/L versus ≥ 90 U/L) at randomization will also be presented.

5.1.3 Medical History

Medical history will be summarized by body system for the ITT population by treatment group and overall.

5.1.4 Extent of Exposure

Extent of exposure will be summarized for the safety populations by treatment group.

Duration of treatment will be calculated in days as the date of the last dose of study treatment (when abiraterone as IMP) – date of the first dose of study treatment + 1. This will be summarized using descriptive statistics.

Number of radium 223/placebo injections received will be summarized by treatment group.

Total dose of radium-223 dichloride and abiraterone as IMP will be summarized separately.

5.1.5 Prior and Concomitant Medications

All non-study medications taken during the study will be coded using the World Health Organization Drug Dictionary (WHO-DD) 2005 Q3 and the Anatomical Therapeutic Chemical (ATC) classification system. Coding will include the drug class and generic drug name.

Non-study medications taken during the study will be categorized as prior medications, concomitant medications during the treatment period, and post treatment medications during the active follow-up and long-term follow-up.

Prior medications will be defined as a non-study medication with a stop date prior to the first dose of study treatment.

Concomitant medications will be defined as:

- Non-study medications with a start or stop date on or after the date of the first dose of study treatment;

- Non-study medications that started prior to the first dose of study treatment and are ongoing during the treatment period;

- Non-study medications with partial start dates that indicate that the medication could be concomitant in relation to the date of the first dose of study treatment;

- Non-study medications with completely missing start dates, unless their stop dates confirm otherwise (i.e. the stop date is before the first injection of study treatment).

Post treatment medications are defined as non-study medications taken after the treatment period.

All concomitant medications will be listed, including verbatim descriptions and coded terms, and flags for prior medications. Prior, concomitant, and post treatment medications including anti-cancer therapies will be summarized using frequencies of subjects reporting each drug category and generic drug name. For each subject, multiple records of the same concomitant medication will be counted once within a drug class and generic drug name.

5.2 Efficacy Analysis

5.2.1 Primary Efficacy Analysis

The primary efficacy endpoint is symptomatic skeletal event-free survival (SSE-FS). It is defined as the time from randomization to the earliest occurrence of the following:

(1) An on-study SSE, which is defined as:

a. the use of external beam radiotherapy (EBRT) to relieve skeletal symptoms
b. the occurrence of new symptomatic pathological bone fractures (vertebral or non-vertebral)
c. the occurrence of spinal cord compression
d. a tumor related orthopedic surgical intervention.

(2) Death from any cause

The censoring rules for SSE-FS is summarized in Table 1:

Table 1:Symptomatic Skeletal Event Free Survival (SSE-FS) censoring rule.

Situation	End Date	Censored	Reason for Censoring
No post-baseline SSE assessment and no death	Date of Randomization	Yes	No post-baseline SSE assessment and no death.
Subject had an SSE event	Date of first SSE	No*	N/A
Death without prior SSE (<13 weeks between last SSE assessment and death) #	Date of Death	No*	N/A
Death without prior SSE (≥ 13 weeks between last SSE assessment and death) #	Last SSE assessment before the missing SSE assessments #	Yes	≥ 13 weeks between last SSE assessment and death
Neither SSE nor death at data cutoff	Last SSE assessment	Yes	Neither SSE nor death

Symptomatic Skeletal Event Free Survival (SSE-FS) = End Date – Date of Randomization +1
*The earliest end date in the table is used in calculating the SSE-FS.
#: use randomization date instead of last SSE assessment date if no post-baseline SSE assessment.
Note: SSE events immediately after missing SSE assessments are still counted as events in the analysis of SSE-FS.

SSE-FS analysis will be performed for the ITT population using a stratified log-rank test stratified by the three randomization strata from IxRS: geographical regions (Western Europe/North America/Australia vs. Asia vs. rest of world), concurrent use of denosumab or bisphosphonates [yes vs. no], and total ALP < 90 U/L versus total ALP ≥ 90 U/L. The treatment variable is binary and will be coded as 1 if the patient is randomized to receive radium-223 dichloride and 0 if the patient is randomized to receive placebo. The hazard ratio (radium-223 dichloride/placebo) will be computed together with the two sided 95% CI using a stratified Cox regression model with the three randomization factors as strata in the model.

SSE-FS will also be summarized using the Kaplan-Meier estimates. Median survival time together with the 25th and 75th percentiles and associated 95% CI will be presented by treatment group. Kaplan-Meier curves will be generated for both treatment groups.

The contribution of each component of the composite SSE between the arms will be evaluated. Descriptive statistics will be presented.

In order to assess the impact of death censoring rule on the analysis of SSE-FS, a sensitivity analysis will be performed for SSE-FS by counting all deaths without prior SSE as SSE-FS events.

Sensitivity analyses for SSE-FS will be performed concerning initiation of new systemic anticancer therapy or initiation/change of bone targeted treatment (i.e. bisphosphonates/denosumab) separately. In the sensitivity analyses, SSE-FS will be censored at the start of new systemic anticancer therapy or initiation/change of bone targeted treatment if the subject received such treatment (see Table 2: for detailed censoring rules).

Another sensitivity analysis will be done by considering the initiation of systemic anticancer therapy or initiation/change of bone targeted treatment as a SSE-FS event.

Another supportive analysis for SSE-FS will be performed using a non-stratified log-rank test. The treatment effect (hazard ratio) will be estimated using the Cox proportional hazards regression model without accounting the three stratification factors.

As an exploratory analysis, the stratified Cox proportional hazards regression model may be fitted including other baseline covariates considered to be of prognostic importance (see section 3.2).

In addition, analyses examining the relationship between body weight (total dose delivered) and SSE-FS and the ideal body weight with SSE-FS will be performed. Details regarding the subgroup analysis for SSE-FS is specified in Section 5.5.

Table 2: Censoring rule for SSE-FS sensitivity analysis

Situation	End Date	Censored	Reason for Censoring
No post-baseline SSE assessment and no death	Date of Randomization	Yes	No post-baseline SSE assessment and no death.
Subject had a SSE event	Date of first SSE	No*	N/A
Death without prior SSE (<13 weeks last SSE assessment and death)#	Date of Death	No*	N/A
New systemic anticancer therapy started (or initiation/change of bone targeted treatment) prior to on study SSE or death	Start date of new systemic anticancer therapy (or bone targeted treatment)	Yes	Start of new systemic anticancer therapy (or initiation/change of bone targeted treatment)
Death without prior SSE (≥ 13 weeks between last SSE assessment and death)#	Last SSE assessment before missing SSE assessment (use randomization date if no post-baseline SSE assessment)#	Yes	≥ 13 weeks between last SSE assessment and death
Neither SSE nor death at data cutoff	Last SSE assessment	Yes	Neither SSE nor death

Symptomatic Skeletal Event Free Survival (SSE-FS) = End Date – Date of Randomization +1
*The earliest end date in the above table is used in calculating the SSE-FS.
#: use randomization date instead of last SSE assessment date if no post-baseline SSE assessment.

All the sensitivity analyses and supportive analyses will be done for the ITT population only.

5.2.2 Secondary efficacy Analysis

5.2.2.1 Secondary efficacy endpoints

The secondary efficacy variables are specified below:

- Overall survival (OS)
- Time to opiate use for cancer pain
- Time to pain progression
- Time to cytotoxic chemotherapy
- Radiological progression free survival(rPFS)

Overall survival is defined as the time (days) from the date of randomization to the date of death due to any cause. For subjects who are still alive, their OS will be censored at the last known alive date or the database cutoff date, whichever occurs first.

Time to opiate use for cancer pain is defined as the interval from the date of randomization to the date of opiate use. Subjects who have no opiate use at the time of analysis will be censored at the last known date of no opiate use. Subjects with no on-study assessment or no baseline assessment will be censored at the date of randomization.

Analgesic use will be captured via three methods:

- analgesic concomitant medication case report form, where the physician records the analgesic medication prescribed to manage pain.
- 24 hour analgesic consumption case report form, in which all analgesic medication taken in the last 24 hours, including dose of medication and number of pills consumed, will be recorded. Patients will be asked to bring all of their medication to the clinic visit and complete this form with the assistance of the clinician.
- Opiate use case report form, in which the information of opiate pain medication since last assessment was collected.

For the time to first opiate use for cancer pain secondary endpoint, the earliest date of the first opiate use recorded via any of the following three methods will be used: the 24 hour analgesic consumption case report form, analgesic concomitant medication case report form or opiate use case report form.

Patients who had opiate analgesics at baseline are not eligible for the analysis of this endpoint.

Time to pain progression is defined as the interval from randomization to the first date a subject experiences pain progression based on worst pain score (WPS):

For asymptomatic patients (WPS 0 at baseline), pain progression is defined as an increase of 2 or more points in the average (i.e., average of 7-day assessments) "worst pain in 24 hours" score from baseline observed at 2 consecutive evaluations ≥4 weeks apart

OR

initiation of short or long-acting opioid use for pain.

Assessments will occur daily for one week (including the visit date). An evaluable pain assessment interval requires completion of a minimum of 4 out of 7 daily questions. Subjects who have not experienced pain progression at the time of analysis will be censored on the last date the subject was known to have not progressed. Subjects with no on-study assessment or no baseline assessment will be censored at the date of randomization.

For mildly symptomatic patients (WPS 1-3 at baseline), pain progression is defined as an increase of 2 or more points in the average (i.e., average of 7-day assessments) "worst pain in 24

hours" score from baseline observed at 2 consecutive evaluations ≥4 weeks apart and an average worst pain score of ≥ 4.

OR

initiation of short or long-acting opioid use for pain.

Assessments will occur daily for one week (including the visit date). An evaluable pain assessment interval requires completion of a minimum of 4 out of 7 daily questions. Subjects who have not experienced pain progression at the time of analysis will be censored on the last date the subject was known to have not progressed. Subjects with no on-study assessment or no baseline assessment will be censored at the date of randomization.

For both asymptomatic and mildly symptomatic patients, the opioid use component of this endpoint will be based on the same source data for time to opiate use endpoint. Patients who had opiate analgesics at baseline are not eligible for the analysis of this endpoint.

Time to cytotoxic chemotherapy is defined as the time (days) from the date of randomization to the date of the first cytotoxic chemotherapy. Subjects who have not started cytotoxic chemotherapy during the study will be censored at the last assessment date.

Radiological progression-free survival (rPFS) is defined as the time (days) from the date of randomization to the date of confirmed radiological progression or death (if death occurs before progression). Subjects without confirmed radiological progression or death at the time of analysis will be censored at their last date of radiological tumor assessment (See Table 3 for detailed censoring rules). Bone scans and CT/MRIs will be read both locally and by blinded independent central assessment. If there is an inconsistency between local assessment and central review, then the assessment will be based on central review. If progression is detected by bone scan, a confirmatory scan is required at least 6 weeks later. A single SPECT or MRI (with and without contrast media) should be obtained to confirm any suspicious bone scan findings. The date of confirmed radiological progression will be the date of first observation of radiological progression.

Radiological bone progression is determined if at least one of the following criteria is met:

A subject is considered to have progressed by bone scan if:

- The first bone scan with ≥2 new lesions compared to baseline is observed <12 weeks from randomization and is confirmed by a second bone scan taken ≥6 weeks later showing ≥2 additional new lesions (a total of ≥4 new lesions compared to baseline); or

- The first bone scan with ≥2 new lesions compared to baseline is observed ≥12 weeks from randomization and the new lesions are verified on the next bone scan ≥6 weeks later (a total of ≥2 new lesions compared to baseline).

If bone scans are repeated, the best scan should be submitted to the core imaging lab as per judgment of the site.

Progression in soft tissue and occurrence of visceral disease will be assessed by CT or MRI every 8 weeks then every 12 weeks (± 7 days) based on modified RECIST 1.1 (protocol section 14.3). CT or MRI scans will be read both locally and by blinded independent central assessment. The RECIST will be used to assess soft tissue and visceral disease only. Bone lesions detected by CT or MRI cannot be selected as target or non-target lesions. Bone lesions will be evaluated by bone scans.

Table 3: Radiological progression free survival censoring rule

Situation	End Date	Censored	Reason for Censoring
No baseline or post-baseline radiological assessment	Randomization Date	Yes	No baseline or post-baseline tumor assessment.
Subject had a radiological assessment of PD (no more than one missed radiological assessment)	Date of first PD	No*	N/A
Subject discontinued from study for other than PD or death	Last radiological assessment without PD	Yes	Subject discontinued from study due to a reason other than PD or death
Death during the study (no more than one missed radiological assessment) without radiological PD assessment before death	Date of Death	No*	N/A
Subject discontinued from study due to PD, but no documented PD date	Date of last radiological assessment	Yes	Subjects discontinued from study due to PD, but no documented date of PD
Subject still on study at the time of data cutoff without PD	Last radiological assessment before data cutoff	Yes	Subject is still alive without PD
Death or PD after more than one missed radiological assessment	Date of last radiological assessment before missed assessment	Yes	Missed more than one tumor assessments
New systemic anticancer treatment started prior to radiological PD or death	Date of last radiological assessment before starting new systemic anticancer treatment	Yes	New systemic anticancer treatment started

Radiological Progression Free Survival (rPFS) = End Date – Date of Randomization +1

*Earliest end dates in the above table are used in calculating the PFS.

5.2.2.2 Analysis of secondary efficacy endpoints

A hierarchical testing procedure will be followed for the analysis of secondary endpoints. All time to event efficacy analysis will be performed for the ITT population using a stratified log-rank test accounting for the three randomization stratification factors. The treatment variable is binary and will be coded as 1 if the subject was randomized to receive radium-223 dichloride and 0 if the subject was randomized to receive placebo. The treatment effect (hazard ratio) will be estimated using the Cox proportional hazards regression model stratified by the three randomization stratification factors. For rPFS, analyses will be done based on both the investigator's and the independent assessments. Analysis based on the independent assessment will be considered as the primary analysis for rPFS. Analysis based on the investigator's assessment will be considered as supportive analysis.

Time to event endpoints will also be summarized using the Kaplan-Meier estimates. Median survival time together with the 25th and 75th percentiles and associated 95% confidence interval (CI) will be presented by treatment group as well as the number and percentage of censored observations. A Kaplan-Meier curve will be generated for each treatment group.

Sensitivity analysis will be conducted for rPFS without considering new systemic anti-cancer treatment received in the censoring rules.

Sensitivity analysis will be done for time to pain progression without considering initiation of short or long-acting opioid use for pain as event.

Sensitivity analysis will be performed for time to opiate use by only using the analgesic use data on the Opiate Use eCRF page. Similarly sensitivity analysis will be performed for time to pain progression using pain score data and analgesic use data only based on the Opidate Use eCRF page.

Although patients who had opiate analgesics at baseline are excluded from the primary analysis of time to pain progression, they will be included in the analysis as a sensitivity analysis. For patients who took a weak opiate at baseline, the initiation of any strong opioid would be counted as pain progression; For patients who took a strong opioid at baseline, the initiation of an additional strong opioid would be counted as pain progression.

The pain score change at week 12 will also be analyzed using an ANCOVA model with baseline pain score and study treatment as covariates.

Per protocol, individual subjects may be unblinded per investigators request after they have an SSE, allowing the investigator to determine if the subject would be a candidate to receive treatment with radium 223 dichloride. This unblinding criteria is considered to be the only feasible option to allow patients to have access to Xofigo in the placebo arm, after they have experienced an SSE, thereby becoming symptomatic. For all the secondary efficacy endpoints, analyses will be done based on the randomized treatment. Further evaluation might be done to evaluate the potential impact of unblinding on the analysis of secondary endpoints.

5.2.2.2.1 Hierachial testing procedure

The overall type I error rate for the primary endpoint and selected secondary endpoints will be controlled at 2-sided 0.05 or less using a hierarchical testing procedure as below:

If the primary endpoint SSE-FS is statistically significant at the 0.05 level (2-sided), then the alpha level will be split between OS (alpha=0.001) and the secondary endpoints rPFS, TtPP and TtCC (alpha=0.049).

The rPFS, TtPP and TtCC will be formally tested at 0.049 using a hierarchical testing procedure in this order (Figure 1):

- rPFS: Radiographic progression-free survival
- TtPP: Time to pain progression
- TtCC: Time to cytotoxic chemotherapy

There are two possible scenarios for testing secondary endpoints rPFS, TtPP, TrCC and OS, depending on the analysis results:

Scenario 1:

If rPFS, TtPP and TtCC are all statistically significant at the 0.049 level (2-sided), then the OS will be formally tested at an overall alpha=0.05 (2-sided). OS will be formally tested at the 0.005 level (2-sided) at the interim OS analysis with 275 deaths or at the 0.048 level (2-sided) at the final OS analysis with 500 deaths. To maintain an overall 2-sided significance level of 0.05, these levels will be adjusted based on the number of deaths observed at the interim OS analysis. Figure 1 shows the details of the testing sequence for scenario 1.

Scenario 2:

If at least one of the secondary endpoints, rPFS, TtPP or TtCC, is not statistically significant at the 0.049 level (2-sided), then OS will be formally tested at an overall alpha=0.001 (2-sided). OS will be formally tested at the 5.3×10^{-6} level (2-sided) at the interim OS analysis with 275 deaths or at the 0.001 level (2-sided) at the final OS analysis with 500 deaths. Figure 2 shows the details of the testing sequence for scenario 2.

Figure 1. Pre-specified testing procedure (scenario 1)

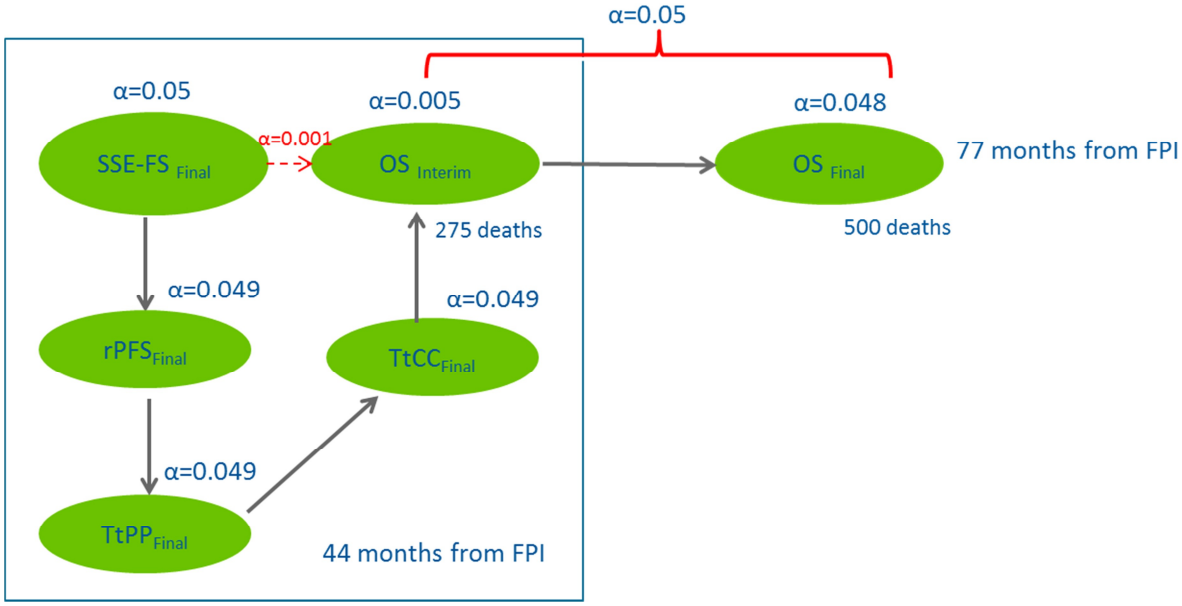
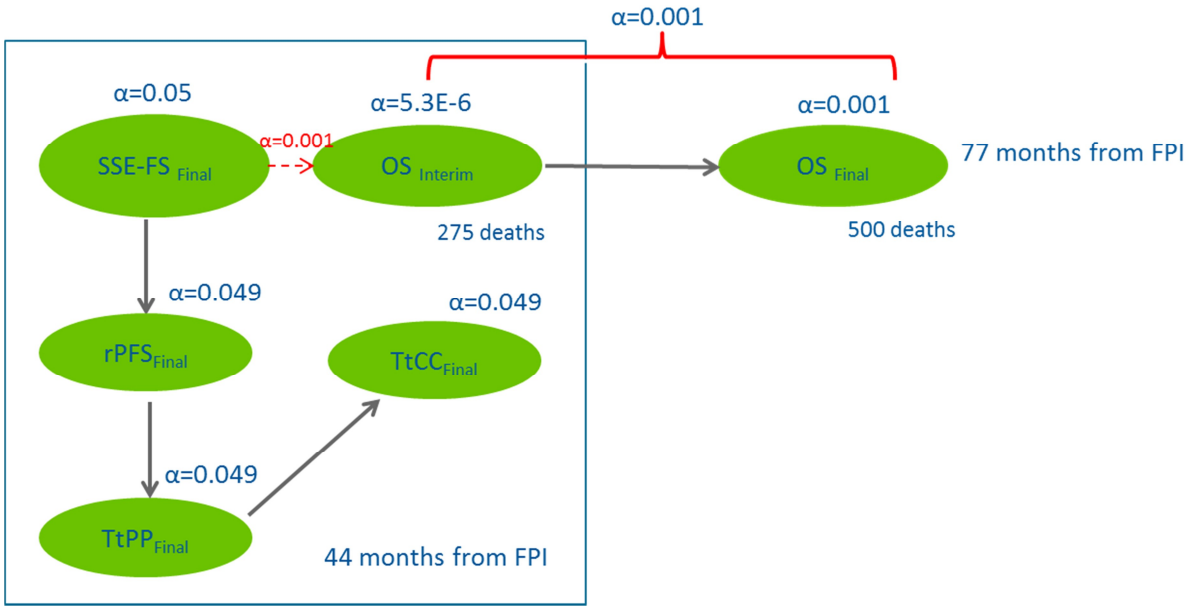


Figure 2. Pre-specified testing procedure (scenario 2)



If SSE-FS is not statistically significant at the 0.05 level (2-sided), then no formal testing will be performed for any other endpoints.

5.2.3 Exploratory efficacy analysis

5.2.3.1 Exploratory efficacy endpoints

There are multiple exploratory endpoints specified for this study as listed below.

Time to first on-study SSE is defined as the time (days) from the date of randomization to the date of the first on-study SSE.

Percentage change in total ALP from baseline to 12 weeks and the percentage change in total ALP from baseline to 24 weeks will be calculated for each subject, based on laboratory data collected as of the 12- and 24-week assessments, respectively. The last observation (including baseline) carried forward method will be used to impute the missing value at week 12 or week 24.

Time to ALP progression is defined as the time (days) from the date of randomization to the date of first ALP progression. Alkaline phosphatase progression is defined as ≥ 25% increase from the baseline value, at least 12 weeks from baseline in subjects with no ALP decline from baseline; or ≥ 25% increase above the nadir value, which is confirmed by a second value obtained 3 or more weeks later in subjects with an initial ALP decline from baseline.

Time to PSA progression is defined as the time (days) from the date of randomization to the date of first PSA progression. Prostate specific antigen progression is defined as ≥ 25% increase from the baseline value and an increase in absolute value of ≥2 ng/mL, at least 12 weeks from baseline in subjects with no PSA decline from baseline; or ≥25% increase and an absolute increase of ≥2 ng/mL above the nadir value, which is confirmed by a second value obtained 3 or more weeks later in subjects with an initial PSA decline from baseline.

Alkaline phosphatase response is defined as ≥30% reduction of the blood level, compared to the baseline value. Confirmed ALP response is defined as a ≥30% reduction of the blood level, compared to the baseline value, confirmed by a second ALP value 4 or more weeks later. ALP responses will be evaluated at week 12 and week 24 after treatment is started.

Prostate specific antigen response is defined as ≥30% reduction of the blood level, compared to the baseline value. A confirmed PSA response is defined as ≥30% reduction of the blood level, compared to the baseline value, confirmed by a second PSA value 4 or more weeks later. PSA responses will be evaluated at week 12 and week 24 after treatment is started.

Radiological progression free survival based on bone scans (bone rPFS) is defined as the time (days) from the date of randomization to the date of confirmed radiological progression detected by bone scans or death (if death occurs before progression). Subjects without confirmed radiological progression or death at the time of analysis will be censored at their last date of radiological tumor assessment. Bone scans will be read both locally and by blinded independent central assessment. If there is an inconsistency between local assessment and central review, then the assessment will be based on central review. If progression is detected by the independent central assessment, a confirmatory scan is required at least 6 weeks later. The date of confirmed

radiological progression will be the date of first observation of central assessment of radiological progression. Radiological progression is determined if at least one of the following criteria is met:

- A subject is considered to have progressed by bone scan if:
    - The first bone scan with ≥2 new lesions compared to baseline is observed <12 weeks from randomization and is confirmed by a second bone scan taken ≥6 weeks later showing ≥2 additional new lesions (a total of ≥4 new lesions compared to baseline); or
    - The first bone scan with ≥2 new lesions compared to baseline is observed ≥12 weeks from randomization and the new lesions are verified on the next bone scan ≥6 weeks later (a total of ≥2 new lesions compared to baseline).

If bone scans are repeated the best scan should be submitted to the core imaging lab as per judgment of the site.

HRQoL Endpoints: Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during the treatment period will be assessed using the NCCN-FACT FPSI-17 questionnaire.

The NCCN-FACT FPSI-17 is a validated instrument that was developed to assess symptoms of prostate cancer, symptoms of treatment of prostate cancer, and health related quality of life of prostate cancer patients (3)The instrument was developed in accordance recent FDA guidance for development of instruments for ePRO. The instrument contains 17 items, each of which utilize a Likert scale with 5 possible responses. The ten items reflect disease related physical symptoms of disease and the responses on the items are be summed to calculate a disease related physical symptom subscale score. One item represents emotion symptom of disease and the response to that item is used to calculate a disease related emotional symptom subscale score. Four items represent treatment related symptoms and the responses to these items are summed to calculate a treatment side effect subscale score. Finally, two items represent functional well-being and response to those items are summed calculated a functions/well-being subscale score.

Time to deterioration in HRQoL will be the endpoint to evaluate the HRQoL FPSI-DRS-P subscale score. Deterioration is defined as a 2 point drop in DRS-P score that persists for two consecutive assessments at least 4 weeks apart, if two consecutive assessment are available. If there is a 2 point drop in the DRS-P score and the next assessment is not available due to death then that single 2 point drop would count as a deterioration. The date of the event is the first assessment with at least 2 points decrease from the baseline. If no deterioration observed, time to deterioration in HRQoL will be censored at the last HRQoL observation.

Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during the period between start of treatment and end of active follow-up with clinic visit.

5.2.3.2 Analysis of exploratory efficacy endpoints

Similar to the primary efficacy endpoint, the stratified log-rank test stratified by the three randomization strata will be performed for the time to event variables (time to first on-study SSE, percentage change in total ALP from baseline, time to ALP progression, time to PSA progression and bone rPFS, etc.) based on the ITT population. The censoring rules for bone rPFS will be the same as defined for rPFS in Table 33. The treatment effect (hazard ratio) and its 95% CI will be estimated using the Cox proportional hazards regression model stratified by the three randomization stratification factors. The Kaplan-Meier estimates and plots will also be presented for each endpoint.

Sensitivity analysis of Time to total ALP progression will be performed by droping the 12 week restriction in the definition.

To assess the potential bias due to missing data , FPSI-DRS-P data will also be analyzed using a pattern mixture model. Patients will be ranked based on the percentage of completed HRQoL assessments (where the numerator for a patient is the number of assessments and the denominator is the number of assessments that the patient should have had if the completed the questionnaire at every scheduled visit prior to their death or prior to leaving the study due to administrative censoring). The median value of the percentage of completed HRQoL assessment will be used to categorize patients into "patients with higher degree of missing data" or "patients with lower degree of missing data". An index variable for the two groups (0= patient with a "high degree of missing data" and 1 = patient with a "lower degree of missing data") will be created. A stratified Cox regression model will be used which contains the treatment variable, the index variable, and a term for the interaction between treatment and the index variable. The stratification factors for the Cox model will be the stratification factors used at randomization. If the interaction term is statistically significant, the analysis will indicate that there might be evidence of informative censoring.

CMH test adjusting the three randomization strata will be used for the ALP response and PSA response to evaluate difference between treatment group. Response rate difference and its exact 95% CI will be provided. ANCOVA with randomization strata as covariates will be done for the percentage change in total ALP from baseline.

Summary statistics will be provided for the HRQoL scores and subscales. See Section 5.4.2 for additional information on the analysis of HRQoL data.

5.2.3.2.1 Exploratory endpoint for labeling purposes

For labeling purposes, analysis for time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during the treatment period will be included in the multiplicity adjustment for this study. This exploratory endpoint will be tested at a two-sided alpha level of 0.05 if the primary endpoint and the selected secondary efficacy endpoints specified in section 5.2.2.2.1 all have statistically significant results based on the preplanned method for alpha adjustment.

No multiplicity adjustment will be done for other exploratory endpoints. Further analyses may be identified and performed for exploratory purposes.

5.3 Safety Analysis

All safety analyses will be done based on the safety population. No formal statistical test will be done for the safety endpoints.

5.3.1 Adverse events

All adverse events (AE) will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) Version 19 or higher. The intensity of an AE will be documented using the NCI-CTCAE v4.03.

Each AE recorded on the CRF will be classified as either a pre-treatment AE, a treatment-emergent AE (TEAE) or a post-treatment AE.

The treatment period for this study extends from the initiation of treatment until 4 weeks after the last administration of abiraterone as IMP and prednisone/prednisolone or 6 months after last administration of radium-223/placebo, whichever occurs later unless a new systemic anti-cancer therapy is initiated. If a new anti-cancer therapy is initiated before the end of the period defined above, the treatment period extends from the initiation of treatment until 4 weeks after the start date of the new anti-cancer therapy.

Pre-treatment AEs
Pre-treatment AEs will be defined as AEs that started and either stopped before the first dose of study treatment or continued after and did not worsen in intensity(i.e. increase in CTC toxicity grade or became serious) during the treatment period.

Treatment-emergent AEs
All other AE starting or worsening within the treatment period will be considered TEAEs for example:

- Events that started on or after the first dose and within the treatment period and are not a continuation of a pre-treatment event.
- Events that started before the first dose and worsened after the first dose or the treatment period.

Post-treatment AEs
Post-treatment follow-up AEs will be defined as AEs that started after the treatment period. Note that the intention of this study is that only related AEs will be collected after the treatment period.

An overall summary of AEs will be provided to present the number and percentage of subjects with

- any pre-treatment AEs, TEAEs, or post-treatment AEs
- any study drug related TEAEs
- any serious TEAEs,
- any serious related TEAEs
- any CTC Grade ≥3 TEAEs
- any TEAEs leading to discontinuation
- any TEAEs leading to death

TEAEs and post-treatment AEs will be summarized by MedDRA system organ class and preferred term. For each subject, multiple occurrences of the same event will be counted once within a system organ class and preferred term.

The same summaries will be repeated for related TEAEs, serious TEAEs, serious related TEAEs, CTC Grade ≥3 TEAEs, CTC Grade 5 TEAEs, TEAEs leading to discontinuation and TEAEs leading to death.

The maximum severity of the TEAEs and post-treatment AEs will be summarized according to the NCI-CTCAE toxicity criteria. For each subject, multiple occurrences of the same event will be counted once at their maximum severity within a system organ class and preferred term. TEAE will also be summarized by NCI-CTCAE toxicity criteria and relationship to study medication.

The numbers of TEAEs, and subjects experiencing TEAEs will be presented.

A data listing will be produced for all AEs, TEAEs leading to discontinuation and drug related TEAEs. Verbatim descriptions and coded terms will be listed for all AEs.

5.3.2 Deaths

Deaths reported during the study period will be tabulated by treatment group. A data listing of all deaths will be provided.

5.3.3 Clinical Laboratory Data

The following laboratory parameters will be summarized:

- Hematology parameters: Hematocrit, Hb, Platelet Counts, RBC, WBC, WBC differential.
- Chemistry parameters: Na, K, Cl, Ca, ALT, AST, LDH, Total ALP, Creatinine, Blood Urea Nitrogen (BUN), Total Bilirubin, and Albumin.
- Other parameters: PSA, Testosterone.

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

generated for each laboratory parameter by visit and treatment group to investigate trends over time and outliers in the data.

In addition, changes from baseline will be summarized in shift tables according to severity during treatment and baseline CTC grade.

If more than one assessment occurred at any visit (i.e. repeat samples taken), the last valid (non-missing) value will be used in the summaries. Unscheduled laboratory data will be listed but will not be included in the summary tables.

5.3.4 ECG

Changes from baseline at week 4 and at end of treatment (4 weeks post last injection) in ECG data during the treatment period will be summarized for each treatment group using shift tables.

5.4 Analysis of Other Endpoints

Other endpoints will be summarized for the ITT population..

5.4.1 ECOG Performance Status

The number and percentage of subjects in each category will be presented. Changes from baseline in PS on the ECOG scale will be summarized in shift tables by treatment group.

5.4.2 Quality of Life

Summaries of HRQoL data will be performed for the ITT population and will be based on an as observed basis (i.e. no imputation for missing data performed) unless otherwise specified.

5.4.2.1 EQ-5D

EQ-5D questionnaire consists of five ordinal categorical responses and a visual analogue scale (VAS). The responses to the five ordinal questions are used to calculate a utility score and the response to the VAS item is used to calculate a self-reported health status score. UK tariffs (i.e. weights) and the standard EQ-5D scoring algorithm will be used to calculated the utility scores.

For each treatment group, responses to the ordinal questions at each assessment time point will be described with frequency tables.

Summary statistic, including mean and change from baseline, for the utility score and the VAS score will be presented by treatment group for each assessment time points.

5.4.2.2 BPI-SF

The BPI-SF (protocol Section 14.9) is a short, self-administered questionnaire with 11 items, which was designed to evaluate the intensity of, and the impairment caused by pain. All BPI-SF

items are scored using rating scales. Four items measure pain intensity (pain now, average pain, worst pain, and least pain) using 0 ("no pain") to 10 ("pain as bad as you can imagine") numeric rating scales, and seven items measure the level of interference with function caused by pain (general activity, mood, walking ability, normal work, relations with other people, sleep and enjoyment of life) using 0 (no interference) to 10 (complete interference) rating scales.

The items are aggregated into two dimensions, (1) Pain severity index, using the mean of the four items on the pain intensity, and (2) Function interference index, using the mean of the seven pain interference items. All four severity items must be completed for aggregating the pain severity index. The function interference index is scored as the mean of the item scores multiplied by seven, given that more than 50% or four of seven, of the items have been completed.

Summary statistic, including mean and change from baseline, for the pain severity and pain intensity and the items make up these indices will be provided by visit and treatment.

5.4.2.3 NCCN-FACT FPSI-17

The NCCN-FACT FPSI-17 (Comprehensive Cancer Network - Functional Assessment of Cancer Therapy - Prostate Symptom Index-17) instrument consists of 17 questions relating to four domains: Disease-related symptoms subscale – physical (DRS-P), Disease-related symptoms subscale – emotional (DRS-E), Treatment side effects subscale (TSE), Function and well-being subscale (FWB).

Summary statistic, including mean and change from baseline, for each domain will be summarized using descriptive statistics by study visit and treatment group.

5.4.2.4 Long term safety

Descriptive summaries will be provide for the long term safety endpoints including new primary malignancies and hematopoietic reserve for tolerability of subsequent chemotherapy

5.4.2.5 Resource utilization

Resource utilization will be provided in the data listing.

5.5 Examination of Subgroups

Subgroup analyses will be conducted for the primary efficacy endpoint SSE-FS, rPFS and OS based on the ITT population. Descriptive statistics and hazard ratio estimates with 95% CI will be provided at least for the subgroups listed below, provided there is a sufficient number of events in total within the subgroup across the treatment arms.

- ECOG performance status at baseline (0 vs 1)
- Extent of Disease (number of bone lesions <6 vs. 6-20 vs. >20 at baseline)
- Asymptomatic vs mildly symptomatic
- Ethnicity (collected as Caucasian, Hispanic, Black, Asian and Other)

- Age group (<65, 65-74, 75-84, >=85)
- Body weight and ideal body weight normalized dose quartiles
- Randomization stratification factors:
  - geographical regions (Western Europe/North America/Australia vs. Asia vs. rest of world)
  - concurrent use of denosumab or bisphosphonates (combined strata) (Yes vs. No)
  - total ALP < 90 U/L versus total ALP ≥ 90 U/L.
- US and EU
- Randomization strata concurrent use of denosumab versus bisphosphonates
- BMI < 30 and >=30
- Number of Radium-223 dichloride injections received (by number of injections - 1- 4, vs. 5-6)
- Gleason Score at the time of diagnosis (<8, >=8, Missing)
- Baseline PSA values by median
- Received prior Antiandrogens (enzalutamide vs. first generation Antiandrogens vs. none)
- Baseline LDH by median
- Baseline tumor location (bone only vs. bone with soft tissues or others)

If important effects are found in subgroups, the interaction analyses between the treatment and the subgroups may be performed.

Safety analysis (i.e. AE) will be conducted based on age subgroup, body weight and ideal body weight normalized dose quartiles.

5.6 Pharmacokinetics / pharmacodynamics

Pharmacokinetics of radium-223 dichloride will not be evaluated in this study; however, the PK of abiraterone will be investigated in 2 subgroups. In each subgroup, approximately 40 subjects (approximately 20 per treatment arm) will be enrolled.

Statistical analysis will be done only for Subgroup 1(Non-compartmental analysis). The Subgroup 2 population PK analysis is not part of the analyses described in this SAP. Subgroup 2 population PK analysis results will be presented in a separate report . For Non-compartmental PK analysis see section 5.6.2.

5.6.1 Biomarkers assessment

Summary statistics for biomarkers and their changes from baseline will be presented by visits and treatment.

5.6.2 Pharmacokinetics

The concentration-time courses of abiraterone will be tabulated separated by treatment. The following statistics will be calculated for each of the sampling points: geometric mean, geometric standard deviation (re-transformed standard deviation of the logarithms), and geometric coefficient of variation, as well as arithmetic mean, standard deviation, and arithmetic coefficient of variation, minimum, median, maximum value and the number of measurements. Means at any time will only be calculated if at least 2/3 of the individual data were measured and were above the lower limit of quantification (LLOQ). For the calculation of the mean value a data point below LLOQ will be substituted by 1 half of this limit. In tables showing mean values, where values below LLOQ are included in the calculation of mean values, these means will be marked.

Individual and geometric mean concentration vs time curves of abiraterone (using the actual sampling times for individual plots and the planned sampling times for mean plots) will be plotted by treatment using both linear and semilogarithmic scale.

Pharmacokinetic characteristics (t_{max} excluded) will be summarized by the statistics mentioned above. T_{max} will be described utilizing minimum, maximum and median as well as frequency counts.

6. Document history and changes in the planned statistical analysis

SAP dated 20 Sep 2013 v1.0 without attachments submitted to FDA.

SAP Amendment 1 dated 22 September 2014 v2.0 without attachments submitted to FDA.

SAP Amendment 2 dated 21 April 2016 v3.0 without attachments submitted to FDA.

SAP Amendment 3 dated 20 April 2017 v4.0 without attachments submitted to FDA.

SAP Amendment 4 dated 10 July 2017 v5.0 without attachments submitted to FDA. The main change in this amendment is to address FDA's comment on statistical testing procedure in Amendment 3.

7. References

1 Lan K.K.G. and DeMets D.L. (1983). Discrete sequential boundaries for clinical trials. Biometrika 70;3:659-63

2 O'Brien P.C. and Fleming T.R. (1979). A multiple testing procedure for clinical trials. Biometrics 35:459-56

3 Scher HI, Halabi S, Tannock I, et al. Design and End Points of Clinical Trials for Patients With Progressive Prostate Cancer and Castrate Levels of Testosterone: Recommendations of the Prostate Cancer Clinical Trials Working Group. J Clin Oncol 2008;26:114-1159

8. SAP amendments - amended

Amendment 1 - Table of Changes in the SAP Amendment

Note: This SAP amendment is prepared based on the protocol amendment 1. Only changes with potential impact on the statistical analysis are included in the comparison table below. All other changes made according to the protocol amendment were considered as cosmetic changes. Details for these changes can be found in the amended protocol Section 13.

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
6	1.3	• Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measuredAssessment of pharmacokinetics (PK) of abiraterone Resource utilization	• Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured **during period between start of treatment and end of active follow-up with clinic visits** • **Assessment of pharmacokinetics (PK) of abiraterone** • **Resource utilization**	Per protocol amendment; adjust format
8	2.1.1	If based on the investigator assessment, the subject does continue to receive clinical benefit, administration of study drug can be continued.	~~If based on the investigator assessment, the subject does continue to receive clinical benefit, administration of study drug can be continued.~~ **If a SSE occurs during the first 6 cycles of treatment, subjects can complete the full 6 administrations of radium-223 dichloride / placebo and can continue abiraterone acetate plus prednisone / prednisolone as long as the subject continues to receive clinical benefit based on the investigator assessment.**	Per protocol amendment

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
			If a SSE occurs after the first 6 cycles of treatment, subjects can continue abiraterone acetate plus prednisone / prednisolone as long as the subject continues to receive clinical benefit based on the investigator assessment. **After an on-study SSE, abiraterone acetate will be considered standard of care (i.e., non-investigational medicinal product [non-IMP]) and will not be supplied as study medication.** **Thus, after an on-study SSE, abiraterone acetate should be prescribed locally and recorded on the concomitant medications electronic case report form (eCRF) page.**	
21	4.1 Pharmacokinetic analysis set (PKS)	All subjects with a valid pharmacokinetic profile for non-compartmental analysis (NCA) will be included in the pharmacokinetic analysis set (PKS).	All subjects **of Subgroup 1 defined in the protocol** with a valid pharmacokinetic profile for non-compartmental analysis (NCA) will be included in the pharmacokinetic analysis set (PKS).	Per protocol amendment
23	5.1.4	Duration of treatment will be calculated in days as the date of the last dose of study treatment – date of the first dose of study treatment + 1. This will be summarized using descriptive statistics. Number of radium 223/placebo injections received will be summarized by treatment group. Total dose of radium-223 dichloride and abiraterone as IMP will be summarized separately.	Duration of treatment will be calculated in days as the date of the last dose of study treatment **(when abiraterone as IMP)** – date of the first dose of study treatment + 1. This will be summarized using descriptive statistics. Number of radium 223/placebo injections received will be summarized by treatment group. Total dose of radium-223 dichloride and abiraterone **as IMP** will be summarized separately.	Per protocol amendment
23	5.1.5	Non-study medications taken during the study will be categorized as prior medications, concomitant medications,	Non-study medications taken during the study will be categorized as prior medications, concomitant medications **during the treatment period**, and	Per protocol amendment

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
		and post treatment medications.	post treatment medications **during the active follow-up and long-term follow-up.**	
24	5.2.1		**In addition the impact of baseline total body weight and ideal body weight on SSE-FS will be explored.**	Per protocol amendment
27	5.2.2.1 Table 5	Situation: Subject discontinued from study due to PD, but no documented PD date Censored: No*	Situation: Subject discontinued from study due to PD, but no documented PD date Censored: **Yes**	correction
32	5.2.3.1	Alkaline phosphatase response is defined as ≥30% reduction of the blood level, compared to the baseline value. Confirmed ALP response is defined as a ≥30% reduction of the blood level, compared to the baseline value, confirmed by a second ALP value 4 or more weeks later. Prostate specific antigen response is defined as ≥30% reduction of the blood level, compared to the baseline value. A confirmed PSA response is defined as ≥30% reduction of the blood level, compared to the baseline value, confirmed by a second PSA value 4 or more weeks later.	Alkaline phosphatase response is defined as ≥30% reduction of the blood level, compared to the baseline value. Confirmed ALP response is defined as a ≥30% reduction of the blood level, compared to the baseline value, confirmed by a second ALP value 4 or more weeks later. **ALP responses will be evaluated at week 12 and week 24 after treatment is started.** Prostate specific antigen response is defined as ≥30% reduction of the blood level, compared to the baseline value. A confirmed PSA response is defined as ≥30% reduction of the blood level, compared to the baseline value, confirmed by a second PSA value 4 or more weeks later. **PSA responses will be evaluated at week 12 and week 24 after treatment is started.**	Per protocol amendment
35	5.3.1	The treatment period for this study extends from the initiation of treatment until 4 weeks after the last administration of abiraterone and	The treatment period for this study extends from the initiation of treatment until 4 weeks after the last administration of abiraterone **as IMP** and prednisone/prednisolone or 6 months after last	Per protocol amendment

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
		prednisone/prednisolone or 6 months after last administration of radium-223/placebo, whichever occurs later.	administration of radium-223/placebo, whichever occurs later.	
38	5.4.2.2	The items are aggregated into two dimensions, (1) Pain severity index, using the sum of the four items on the pain intensity, and (2) Function interference index, using the sum of the seven pain interference items. All four severity items must be completed for aggregating the pain severity index. The function interference index is scored as the mean of the item scores multiplied by seven, given that more than 50% or four of seven, of the items have been completed.	The items are aggregated into two dimensions, (1) Pain severity index, using the **mean** of the four items on the pain intensity, and (2) Function interference index, using the **mean** of the seven pain interference items. All four severity items must be completed for aggregating the pain severity index. The function interference index is scored as the mean of the item scores multiplied by seven, given that more than 50% or four of seven, of the items have been completed.	correction
39	5.6	The population PK analysis is not part of the analyses described in this SAP. For Non-compartmental PK analysis see section 5.6.2.	**Pharmacokinetics of radium-223 dichloride will not be evaluated in this study; however, the PK of abiraterone will be investigated in 2 subgroups. In each subgroup, 40 subjects (approximately 20 per treatment arm) will be enrolled.** **Statistical analysis will be done only for Subgroup 1(Non-compartmental analysis).** The **Subgroup 2** population PK analysis is not part of the analyses described in this SAP. **Subgroup 2 population PK analysis results will be presented in a separate report .** For Non-compartmental PK analysis see section 5.6.2..	To clarify

Amendment 2 - Table of Changes in the SAP Amendment 2

Note: This SAP amendment 2 is mainly for clarifications and some minor corrections. Major changes in this SAP amendment have been summarized as follows.

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
20	3.2	• Baseline albumin value; • Baseline hemoglobin value; • Baseline LDH value; • Baseline ECOG performance status (0, 1); • Baseline PSA value; • Baseline total ALP value; • Age.	• Baseline albumin value (< or ≥ median); • Baseline hemoglobin value (< or ≥ median); • Baseline LDH value (< or ≥ median); • Baseline ECOG performance status (0, 1); • Baseline PSA value (< or ≥ median); • Baseline total ALP value (< or ≥ 90 U/L); • Age (<65, ≥65).	Pre-specify the cutoff values to be used for the baseline covariates in exploratory analysis
23	5.1.2		• PSA and total ALP at randomization	Added a summary of baseline characteristics at randomization
24	5.1.1	In addition, the number of subjects screened, and included in each analysis population will be displayed by region, country and center.	In addition, the number of subjects screened, screen failures, and included in each analysis population will be displayed by region, country and center. The screen failure reasons will be summarized by treatment group	Added a table for screen failure reasons.
24	5.1.4	Extent of exposure will be summarized for the ITT and safety populations by treatment group.	Extent of exposure will be summarized for the safety populations by treatment group.	ITT includes randomized patients who are not treated, and extent of exposure does not apply to these patients
25	5.1.5	Prior, concomitant, and post treatment medications will be summarized using frequencies of subjects reporting each drug category and generic drug name.	Prior, concomitant, and post treatment medications including anti-cancer therapies will be summarized using frequencies of subjects reporting each drug category and generic drug	Clarify

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
			name.	
25	5.2.1	The primary efficacy endpoint is symptomatic skeletal event-free survival (SSE-FS). It is defined as the time from randomization to the occurrence of one of the following:	The primary efficacy endpoint is symptomatic skeletal event-free survival (SSE-FS). It is defined as the time from randomization to the earliest occurrence of the following:	Clarify that the earliest date will be used
26	5.2.1 Table 2	No baseline or post-baseline SSE assessment	No post-baseline SSE assessment and no death.	Patients do not need baseline SSE assessment to have an SSE event
26	5.2.1 Table 2	Death before 1st SSE assessment during the study	Death without prior SSE (<13 weeks last SSE assessment and death)	Clarify
26	5.2.1 Table 2	Missing two or more SSE assessments (defined as ≥ 13 weeks between last SSE assessment and death) before death	Death without prior SSE (≥ 13 weeks between last SSE assessment and death)	Clarify
26	5.2.1 Table 2	No SSE or Death at data cutoff or end of follow up	Neither SSE nor Death at data cutoff	Clarify
27	5.2.1 Table 3		Same changes as Table 2	Clarify
27	5.2.1		In order to assess the impact of death censoring rule on the analysis of SSE-FS, a sensitivity analysis will be performed for SSE-FS by counting all deaths without prior SSE as SSE-FS events.	Added a sensitivity analysis to evaluate the impact of censoring rules on SSE-FS

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
29	5.2.2.1	• analgesic diary which will be completed by subjects on each day that the BPI-SF "rate your worst pain experienced in the last 24 hours" question is asked. In the analgesic diary, the subject will be asked to record the name, dosage and number of pills consumed in the last 24 hours for pain relief.	• Opiate use case report form, in which the information of opiate pain medication since last assessment was collected.	Added one data source from Opiate Use page; ePRO device data can not be used for analysis due to high proportion of unkwown analgesic drug types
29	5.2.2.1	For the time to first opiate use for cancer pain secondary endpoint, the date of the first opiate use recorded via any of the following three methods: the 24 hour analgesic consumption case report form, analgesic concomitant medication case report form or analgesic diary will be used.	For the time to first opiate use for cancer pain secondary endpoint, the earliest date of the first opiate use recorded via any of the following three methods: the 24 hour analgesic consumption case report form, analgesic concomitant medication case report form or opiate use case report form will be used. Patients who had opiate analgesicsat baseline are not eligible for the analysis of this endpoint.	Same rationale as the previous one; Added a clarification sentence.
30	5.2.2.1		For both asymptomatic and mildly symptomatic patients, the opioid use component of this endpoint will be based on same source data for time to opiate use. For patients who took a weak opiate at baseline, the initiation of any strong opioid would be counted as pain progression; For patients who took a strong opioid at baseline, the initiation of an additional strong opioid would be counted as pain progression.	Added a clarification paragraph
30	5.2.2.1		Subjects who have not started cytotoxic chemotherapy during the study will be censored at last assessment date.	Clarify

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
32	5.2.2.1 Table 5	N/A	Subjects discontinued from study due to PD, but no documented date of PD	Clarified a few censoring rules
33	5.2.2.2	N/A	Sensitivity analysis will be performed for time to opiate use by only using the analgesic use data on the Opiate Use eCRF page. Similary sensivitiy analysis will be performed for time to pain progression using pain sore data and analgesic use data only based on the Opidate Use eCRF page. The pain score change at week 12 will also be analyzed using an ANCOVA model with baseline pain socre and study treatment as covariates.	Added some sensitivity analyses
34	5.2.2.2.1		Added a figure for testing procedure for US submission to clarify the testing sequency	Clarify
34	5.2.2.2.1	OS will be considered for inclusion in the US label if the result is statistically significant at the 0.025 level (2-sided).	OS will be considered for inclusion in the US label if the result is statistically significant at the 0.0015 level (2-sided) at the interim OS analysis with 275 deaths or at the 0.0245 level (2-sided) at the final OS analysis with 500 deaths. To maintain an overall 2-sided significance level of 0.025, these levels will be adjusted based on the number of deaths observed at the interim OS analysis.	Clarified the significance levels at each analysis for overall survival.
35	5.2.2.2.2		Added a figure for testing procedure for EU submission to clarify the testing sequency	Clarify

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
38	5.2.3.2		Sensitivity analysis of Time to bone ALP progression will be performed by droping the 12 week restriction in the definition.	Added a sensitivity analysis
38	5.2.3.2		To assess the potential bias due to missing data , FPSI-DRS-P data will also be analyzed using a pattern mixture model. Patients will be ranked based on the percentage of completed HRQoL assessments (where the numerator for a patient is the number of assessments and the denominator is the number of assessments that the patient should have had if the completed the questionnaire at every scheduled visit prior to their death or prior to leaving the study due to administrative censoring). The median value of the percentage of completed HRQoL assessment will be used to categorize patients into "patients with higher degree of missing data" or "patients with lower degree of missing data". An index variable for the two groups (0= patient with a "high degree of missing data" and 1 = patient with a "lower degree of missing data") will be created. A stratified Cox regression model will be used which contains the treatment variable, the index variable, and a term for the interaction between treatment and the index variable. The stratification factors for the Cox model will be the stratification factors used at randomization. If the interaction term is statistically significant, the analysis will indicate that there might be evidence of informative censoring.	Moved from previous section

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
41	5.3.1	The same summaries will be repeated for related TEAES, serious TEAEs, serious related TEAEs, CTC Grade ≥3 TEAEs, TEAEs leading to discontinuation and TEAEs leading to death.	The same summaries will be repeated for related TEAES, serious TEAEs, serious related TEAEs, CTC Grade ≥3 TEAEs, CTC Grade 5 TEAEs, TEAEs leading to discontinuation and TEAEs leading to death.	Added CTC Grade 5 TEAEs as a separate table
42	5.3.1	A data listing will be produced for all AEs. Verbatim descriptions and coded terms will be listed for all AEs.	A data listing will be produced for all AEs, TEAEs leading to discontinuation and drug related TEAEs. Verbatim descriptions and coded terms will be listed for all AEs.	Added some extra AE listings
42	5.3.2		A data listing of all deaths will be provided.	Added a death listing
42	5.5	Subgroup analyses will be conducted for the primary efficacy endpoint SSE-FS based on the ITT population. Descriptive statistics and hazard ratio estimates with 95% CI will be provided at least for the subgroups listed below, provided there is a sufficient number of events in total within the subgroup across the treatment arms.	Subgroup analyses will be conducted for the primary efficacy endpoint SSE-FS, rPFS and OS based on the ITT population. Descriptive statistics and hazard ratio estimates with 95% CI will be provided at least for the subgroups listed below, provided there is a sufficient number of events in total within the subgroup across the treatment arms.	Added rPFS and OS for subgroup analysis

Amendment 3 - Table of Changes in the SAP Amendment 3

Note: This SAP amendment 3 is mainly focused on statistical testing procedure changes. Some minor corrections and clarifications are also included. Major changes in this SAP amendment have been summarized as follows.

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
	2.1.1 and 2.2		Deleted	The two sections were copied from protocol amendment. No need to repeat it in the SAP.
8	2.2	This study is also powered (~60%) for the analysis of OS	This study is also powered (~70%) for the analysis of OS.	Power increased due to more alpha is assigned to OS testing
8	2.2	If the final analysis of OS after 500 deaths reveals that the experimental treatment is statistically significantly better than treatment with control ($p \leq 0.0245$), then the OS endpoint will be declared positive for the final analysis.	If the final analysis of OS after 500 deaths reveals that the experimental treatment is statistically significantly better than treatment with control ($p \leq 0.048$), then the OS endpoint will be declared positive for the final analysis.	p-value threshold is updated due to more alpha assigned to OS testing
	3.2	Baseline total ALP value (< or >=90 U/L)	Deleted	It was already mentioned earlier in the same section
	3.4	If values are missing at the Baseline (Week 0) visit, data recorded at screening visit will be considered as baseline value. If screening record is also missing, the baseline value will be left as missing.	Deleted	Baseline values will be defined in section 3.6 (data rules)
9-10	3.5.1	If the analysis of OS after 275 deaths following treatment with radium-223 dichloride plus abiraterone plus prednisone/prednisolone is statistically significantly better compared to control ($p \leq 0.0015$, based on O'Brien-Fleming alpha spending function (1, 2)), then OS will be declared positive for the interim analysis, assuming the final event number for OS is	If the analysis of OS after 275 deaths following treatment with radium-223 dichloride plus abiraterone plus prednisone/prednisolone is statistically significantly better compared to control ($p \leq 0.005$, based on O'Brien-Fleming alpha spending function (1, 2)), then OS will be declared positive for the interim analysis, assuming the final event number for OS is 500.	p-value threshold is updated due to more alpha assigned to OS testing

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
		500.		
10	3.5.1	If the interim OS analysis is not statistically significant, the final analysis for OS will be performed when approximately 500 deaths have occurred, corresponding to an approximately 60% power to detect a 25% improvement (in OS with Ra-223 Cl_2 compared with placebo) with a two-sided alpha of 0.025.	If the interim OS analysis is not statistically significant, the final analysis for OS will be performed when approximately 500 deaths have occurred, corresponding to an approximately 70% power to detect a 25% improvement (in OS with Ra-223 Cl_2 compared with placebo) with a two-sided alpha of 0.05.	Power increased due to more alpha is assigned to OS testing
11	3.6		The baseline for all efficacy endpoints is defined as the last non-missing assessment taken on or before the randomization date. Unless otherwise specified, the baseline for all safety data, is the last non-missing assessment taken on or before first treatment date.	New texts added to clarify the definition of baseline for efficacy and safety endpoints.
	3.7	The results of validity review meetings will be documented in the Validity Review Reports and may comprise decisions and details relevant for statistical evaluation. Any changes to the statistical analysis prompted by the results of the validity review meeting will be documented in an amendment and, if applicable, in a supplement to this SAP.	Deleted	Validity review meetings will not impact the patients in each analysis set defined in this SAP because there is no per-protocol set defined for this study.
	4.1	Final decisions regarding the assignment of subjects to analysis sets will be made during the Validity Review Meetings and documented in the Validity Review Reports	Deleted	Validity review meetings will not impact the patients in each analysis set

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
		(see SAP Section 3.7).		defined in this SAP because there is no per-protocol set defined for this study.
15	5.2.1	SSE-FS analysis will be performed for the ITT population using a stratified log-rank test stratified by the three randomization strata	SSE-FS analysis will be performed for the ITT population using a stratified log-rank test stratified by the three randomization strata from IxRS	Clarified that the stratification data are from IxRS instead of eCRF pages.
16	5.2.1	Another sensitivity analysis will be done considering the initiation of systemic anticancer therapy	Another sensitivity analysis will be done by considering the initiation of systemic anticancer therapy or initiation/change of bone targeted treatment as a SSE-FS event.	Added bone targeted treatment to align with the wording in the censoring rule in Table 2
	5.2.2.1	For each visit, a patient will be assigned an AQA scale rating based on 24 hour analgesic consumption case report form. Table 4 **Analgesic Quantification Algorithm**	deleted	AQA is not required for analysis in this study because patients are not allowed to take opiate analgesics at baseline
18	5.2.2.1	For patients who took a weak opiate at baseline, the initiation of any strong opioid would be counted as pain progression; For patients who took a strong opioid at baseline, the initiation of an additional strong opioid would be counted as pain progression.	Patients who had opiate analgesics at baseline are not eligible for the analysis of this endpoint.	Patients who had opiate analgesics at baseline will be excluded from the analysis of Time to Pain Progression in order to be consistent with Time to Opiate use. The original method of handling these patients will be used as a sensitivity

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
				analysis.
20	5.2.2.2	A two-sided type I error rate of 0.025 will be will be used for the analysis of each secondary endpoint.	A hierarchical testing procedure will be followed for the analysis of each secondary endpoint.	Statistical testing procedure has been updated.
21	5.2.2.2		Although patients who had opiate analgesics at baseline are excluded from the primary analysis of time to pain progression, they will be included in the analysis as a sensitivity analysis. For patients who took a weak opiate at baseline, the initiation of any strong opioid would be counted as pain progression; For patients who took a strong opioid at baseline, the initiation of an additional strong opioid would be counted as pain progression. Sensitivity analysis will be conducted for rPFS without considering new systemic anti-cancer treatment received in the censoring rules.	Sensitivity analyses were added for the analysis of rPFS and Time to Pain Progression
21-22	5.2.2.2.1	All texts in sections 5.2.2.2.1 and 5.2.2.2.2	The overall type I error rate for the primary endpoint and selected secondary endpoints will be controlled at 2-sided 0.05 or less using a hierarchical testing procedure as below: If the primary endpoint SSE-FS is statistically significant at the 0.05 level (2-sided), then a claim on the secondary endpoints will be made in the following order at the same 0.05 level (2-sided): - rPFS - Time to pain progression	Given the importance of the rPFS endpoint and higher hurdle for OS testing, it was considered that split alpha is not the most efficient testing procedure, hence a sequential testing is utilized for OS and other secondary endpoints.

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
			- Time to cytotoxic chemotherapy - OS (interim) If SSE-FS, rPFS, time to pain progression and time to cytotoxic chemotherapy are all statistically significant at the 0.05 level (2-sided), the OS will be formally tested. OS will be considered for inclusion in the label if the result is statistically significant at the 0.005 level (2-sided) at the interim OS analysis with 275 deaths or at the 0.048 level (2-sided) at the final OS analysis with 500 deaths. To maintain an overall 2-sided significance level of 0.0-5, these levels will be adjusted based on the number of deaths observed at the interim OS analysis. Time to opiate use will not be formally tested but will be analyzed using the same approach as other secondary endpoints.	Time to opiate use was removed from the testing hierarchy because it is contained in TTPP and will be analyzed separately as a secondary endpoint.
25	5.2.3.2.1	For labeling purposes, analysis for time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during the treatment period will be included in the multiplicity adjustment for this study. This exploratory endpoint will be tested at a two-sided alpha level of 0.05 if all primary and secondary efficacy endpoints have statistically significant	For labeling purposes, analysis for time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during the treatment period will be included in the multiplicity adjustment for this study. This exploratory endpoint will be tested at a two-sided alpha level of 0.05 if the primary endpoint and the selected secondary efficacy endpoints specified in section 5.2.2.2.1 all have statistically significant results based on the	The testing sequence was updated based on the new statistical testing procedure

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
		results based on the preplanned method for alpha adjustment	preplanned method for alpha adjustment.	
26	5.3.1	The treatment period for this study extends from the initiation of treatment until 4 weeks after the last administration of abiraterone as IMP and prednisone/prednisolone or 6 months after last administration of radium-223/placebo, whichever occurs later.	The treatment period for this study extends from the initiation of treatment until 4 weeks after the last administration of abiraterone as IMP and prednisone/prednisolone or 6 months after last administration of radium-223/placebo, whichever occurs later unless a new systemic anti-cancer therapy is initiated. If a new anti-cancer therapy is initiated before the end of the period defined above, the treatment period extends from the initiation of treatment until 4 weeks after the start date of the new anti-cancer therapy.	Treatment period was updated based on the latest protocol amendment
30	5.5		• Baseline LDH by median • Baseline tumor location (bone only vs. others)	New subgroup variables were added

Amendment 4 - Table of Changes in the SAP Amendment 4

Note: This SAP amendment 4 is primarily focused on addressing FDA's comment on the statistical testing procedure in the submitted SAP Amendment 3. The changes have been summarized as follows.

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
9	2.2	This study is also powered (~70%) for the analysis of OS. For the concluding analysis of OS, 500 deaths are projected to occur by approximately 71.4 months after the first subject is randomized, assuming the median	This study is also powered (~70%) for the analysis of OS if OS is tested at a 0.05 level (2-sided) For the concluding analysis of OS, 500 deaths are projected to occur by approximately 71.4 months after the first subject is	Due to the changes in the statistical testing procedure, the power and p-value boundary for OS analysis are

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
		OS for the control arm is 35.3 months and a 25% improvement for the radium-223 dichloride arm. If the final analysis of OS after 500 deaths reveals that the experimental treatment is statistically significantly better than treatment with control (two-sided p≤0.048), then the OS endpoint will be declared positive for the final analysis.	randomized, assuming the median OS for the control arm is 35.3 months and a 25% improvement for the radium-223 dichloride arm. If the final analysis of OS after 500 deaths reveals that the experimental treatment is statistically significantly better than treatment with control , then the OS endpoint will be declared positive for the final analysis.	dependent on the analysis results of the secondary endpoints.
10	3.2		A step-wise selection method will be used to choose the final statistical model with entry alpha level 0.1 and exit alpha level 0.1. The stratification factors are always included as covariates in the final model. If deemed necessary, additional variables may also be added in the model selection process.	To clarify how to select the final statistical model in this exploratory analysis.
11	3.5.1	If the analysis of OS after 275 deaths following treatment with radium-223 dichloride plus abiraterone plus prednisone/prednisolone is statistically significantly better compared to control (two-sided p ≤ 0.005, based on O'Brien-Fleming alpha spending function (1, 2)), then OS will be declared positive for the interim analysis, assuming the final event number for OS is 500. The actual nominal alpha levels will be calculated based on the actual number of events accrued at the OS interim analysis. If the interim OS analysis is not statistically significant, the final analysis for OS will be performed when	If SSE-FS is statistically significant at the 0.05 level (2-sided) and rPFS, time to pain progression and time to cytotoxic chemoterhapy are all statistically significant at the 0.049 level (2-sided), then the OS will be formally tested at an overall alpha=0.05 (2-sided). If the analysis of OS after 275 deaths following treatment with radium-223 dichloride plus abiraterone plus prednisone/prednisolone is statistically significantly better compared to control (two-sided p ≤ 0.005, based on O'Brien-Fleming alpha spending function (1, 2)), then OS will be declared positive for the interim analysis, assuming the final event number for OS is 500. The actual nominal alpha levels will be	The testing strategy of OS is changed based on the new testing procedure described in section 5.2.2.2.1

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
		approximately 500 deaths have occurred, corresponding to an approximately 70% power to detect a 25% improvement (in OS with Ra-223 Cl_2 compared with placebo) with a two-sided alpha of 0.05.	calculated based on the actual number of events accrued at the OS interim analysis. If the interim OS analysis is not statistically significant, the final analysis for OS will be performed when approximately 500 deaths have occurred, corresponding to an approximately 70% power to detect a 25% improvement (in OS with Ra-223 Cl_2 compared with placebo) with a two-sided alpha of 0.05. If at least one of the secondary endpoints, rPFS, time to pain progression or time to cytotoxic chemotherapy, is not statistically significant at the 0.049 level (2-sided), then the OS will be formally tested at an overall alpha=0.001 (2-sided). Refer to scenario 2 in section 5.2.2.2.1 for details.	
22	5.2.2.2.1			To address FDA's comment on the statistical testing procedure in SAP amendment 3, the alpha level will be split between OS (alpha=0.001) and the secondary endpoints rPFS, TtPP and TtCC (alpha=0.049) when the SSE-FS is statistically significant at the 0.05 level (2-sided).

SAP PAGE	SAP SECTION	ORIGINAL TEXT	AMENDED TEXT	RATIONALE
31	5.5	• Received prior Antiandrogens including enzalutamide (Yes vs. No)	• Received prior Antiandrogens (enzalutamide vs. first generation Antiandrogens vs. none)	Since enzalutamide is expected to have more impact on efficacy analysis than other antiandrogens, it was separated from other antiandrogens in subgroup analysis

Title page

A phase III randomized, double-blind, placebo-controlled trial of radium-223 dichloride in combination with abiraterone acetate and prednisone/prednisolone in the treatment of asymptomatic or mildly symptomatic chemotherapy-naïve subjects with bone predominant metastatic castration-resistant prostate cancer (CRPC)

Short title: E**valuation of** R**adium-223 dichloride in combination with** A**biraterone in CRPC - ERA 223**

Bayer study drug	BAY no. 88-8223 / Radium-223 dichloride / Xofigo		
[Study purpose:]	To compare the symptomatic skeletal event-free survival of subjects with asymptomatic or mildly symptomatic chemotherapy-naïve bone predominant metastatic castration-resistant prostate cancer (CRPC) treated with radium-223 dichloride or placebo, in combination with abiraterone and prednisone/prednisolone		
Clinical study phase:	III	**Date:**	15 October 2018
Study No.:	BAY 88-8223 / 15396	**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

Abbreviations

All abbreviations that are used in the SAP must be defined in this section.
The following is an example of commonly used abbreviations:

AE	Adverse Event
CI	Confidence Interval
CSR	Clinical Study Report
OS	Overall Survival
rPFS	Radiological Progression-free Survival
SSE-FS	Symptomatic Skeletal Event-free Survival
SAP	Statistical Analysis Plan

1. Introduction

This Supplemental Statistical Analysis Plan (SAP) describes analyses that were not included in the main SAP but may be used for CSR. Changes to planned analyses in the SAP v5.0 are also described in this Supplemental SAP.

This Supplemental SAP version 1.0 is a supplement of SAP version 5.0 dated 10th July 2017.

2. Study Objectives

The primary objective is to compare, in subjects with asymptomatic or mildly symptomatic chemotherapy-naïve bone predominant metastatic CRPC, the clinical benefit of radium-223 dichloride versus placebo in combination with abiraterone and prednisone/prednisolone, with the primary efficacy endpoint being:

- Symptomatic skeletal event-free survival (SSE-FS)

.

The secondary objectives of this study are:

- OS
- Time to opiate use for cancer pain
- Time to pain progression
- Time to cytotoxic chemotherapy
- rPFS
- Safety, acute and long term, including new primary malignancies and hematopoietic reserve for tolerability of subsequent chemotherapy

The exploratory objectives of this study are:

- Time to first on-study SSE
- Percentage change in total ALP from baseline
- Time to ALP progression
- Time to PSA progression
- ALP response
- PSA response
- Bone scan-specific rPFS
- Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during the treatment period
- Time to increase in physical symptoms of disease based on the NCCN-FACT FPSI-17 physical disease related symptoms (FPSI-DRS-P) subscale score measured during period between start of treatment and end of active follow-up with clinic visits

- Assessment of pharmacokinetics (PK) of abiraterone
- Resource utilization
- Biomarker assessments
- An additional explorative objective is to evaluate the impact of baseline total body weight (TBW) and ideal body weight (IBW) on SSE-FS and adverse events.

3. Study Design

Refer to the main SAP v5.0 dated 10th July 2017.

4. General Statistical Considerations

Refer to the main SAP v5.0 dated 10th July 2017.

5. Analysis Sets

Refer to the main SAP v5.0 dated 10th July 2017.

6. Statistical Methodology

Refer to the main SAP v5.0 dated 10th July 2017.

6.1 Changes in the Definition of Prior and Concomitant Medications

Non-study medications taken before and/or during the study will be categorized as prior medications, concomitant medications, and post treatment medications. Two modifications will be made for the definition.

The definition of concomitant medications will be updated for the scenarios below:

Old definition:

- Non-study medications that started prior to the first dose of study treatment and are ongoing during the treatment period;

New Definition:

- Non-study medications that started prior to the first dose of study treatment and are ongoing after the start of study treatment;

The definition of post-treatment medications will be updated as below:

Old definition: Post treatment medications are defined as non-study medications taken after the treatment period.

New definition: Post treatment medications are defined as non-study medications taken after the last dose of study treatment.

Detailed classifications of prior/concomitant/post treatment medication are also illustrated in **Table 6–1**:

Table 6–1 Medication Classification

	Prior to study drug	Study drug started	Treatment with study drug	Study drug stopped	Follow up	**Prior Medication?**	**Concomitant Medication?**	**Post-treatment Medication?**
C1	■	■				**Yes**	**No**	**No**
C2	■	■	■	■		**Yes**	**Yes**	**No**
C3		■	■	■	■	**No**	**Yes**	**Yes**
C4	■	■	■	■	■	**Yes**	**Yes**	**Yes**
C5				■	■	**No**	**No**	**Yes**
C6		■	■	■		**No**	**Yes**	**No**
C7		■				**No**	**Yes**	**No**
C8				■		**No**	**Yes**	**No**

C1= medication started before study drug administration and ended on or before study drug administration
C2= medication started before study drug administration and ended during study drug administration
C3= medication started on or after study drug administration and ended after study drug administration
C4= medication started before study drug administration and ended after study drug administration
C5= medication started on or after study drug administration and ended after study drug administration
C6= medication started on or after study drug administration and ended before or on the same date as end of study drug administration
Note: C7and C8 are covered under C6.

6.2 Efficacy

6.2.1 Change in Sensitivity Analyses and Exploratory Analyses for Primary Efficacy Endpoint

The planned sensitivity analyses and exploratory analyses for SSE-FS in the main SAP v5.0 will not be performed because the primary endpoint was not met.

6.2.2 Changes in Sensitivity Analyses for Secondary Efficacy Endpoints

The sensitivity analysis for time to pain progression by including the subjects with opiate use at baseline will not be performed.

6.2.3 Changes in Sensitivity Analyses for Exploratory Efficacy Endpoint

The pattern mixture model for analyzing FPSI-DRS-P data will not be performed. CMH test for comparing treatment difference will not be performed for the ALP and PSA response.

6.2.4 Changes in Examination of Subgroups

The list used for subgroup analyses for SSE-FS, rPFS and OS were updated and provided as below:

- ECOG performance status at baseline (0 vs 1)
- Extent of Disease (number of bone lesions <6 vs. 6-20 vs. >20 but not superscan vs. superscan at baseline)
- Asymptomatic vs mildly symptomatic
- Race (collected as Caucasian, Hispanic, Black, Asian and Other)
- Ethnicity(collected as Hispanic or Latino, not Hispanic or Latino, not reported)
- Randomization stratification factors(IxRS data):
    - geographical regions (Western Europe/North America/Australia vs. Asia vs. rest of world)
    - concurrent use of denosumab or bisphosphonates (combined strata) (Yes vs. No)
    - total ALP < 90 U/L versus total ALP ≥ 90 U/L.
- US vs. EU vs. rest of the world
- CRF data: concurrent use of denosumab or bisphosphonates(Y/N)
- CRF data: concurrent use of biophosonates (Y/N)
- CRF data: concurrent use of denosumab (Y/N)
- BMI < 30 and >=30
- Gleason Score at the time of diagnosis (<8, >=8, Missing)
- Baseline PSA values by median
- Received prior Antiandrogens including enzalutamide(Y/N)
- Baseline LDH by median
- Baseline tumor location (bone only vs. bone with soft tissues or others)
- Prior use of enzalutamide(Y/N)
- Piro chemotherapy(<=1 regimen vs. >1 regimen)
- Baseline hemoglobin by median
- Prior skeletal related event(SRE)(Y/N)
- Baseline renal impaired(Normal GFR vs mild decrease in GFR vs. moderate decrease in GFR)

6.3 Exploratory variables and analyses

Refer to the main SAP v5.0 dated 17 NOV 2017.

6.4 Pharmacokinetics

6.4.1 Changes in Planned PK Analyses

The planned summary statistics for concentration and PK parameters of abiraterone by treatment will not be performed due to very small number of samples collected.

6.5 Safety

Refer to the main SAP v5.0 dated 10th July 2017.

6.5.1 Adverse Events

Additional analyses of Adverse Events of Special Interest

Time to fracture is defined as the time (months) from the first dose date to the date of first fracture. Subjects without fractures are censored at last visit date during the treatment period. Time to fracture will be summarized using Kaplan-Meier estimates. Median survival time together with the 25th and 75th percentiles and associated 95% Brookmeyer-Crowley confidence intervals (CI) will be presented by treatment arm. Corresponding Kaplan-Meier curves will be generated by treatment arm.

The summary statistics will be also be provided for first fracture occurred after treatment started by including:

- Timing of first fracture(<6 months, 6-12 months, 12-24 months and >=24 months)
- Number of subjects with at least one fracture
- Number of subjects with first fracture after last study treatment date
- Number of subjects with first fracture before or on last study treatment date
- Number of subjects with first fracture before or on last xofigo dose date
- Patients who had a fall or dizziness within 7 days before or on the date of fracture
- Number of fractures considered as serious AE

7. Document history and changes in the planned statistical analysis

- Main SAP v5.0 dated 10th July 2017.

8. References

Refer to the main SAP v5.0 dated 10th July 2017.