

Document Type:	Statistical Analysis Plan
Official Title:	A three arm randomized, open-label Phase II study of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of National Institute of Standards and Technology [NIST] update) versus 80 kBq/kg (88 kBq/kg after implementation of new NIST update and versus 50 kBq/kg (55 kBq/kg after implementation of NIST update) in an extended dosing schedule in subjects with castration-resistant prostate cancer metastatic to the bone
NCT Number:	NCT02023697
Document Date:	08 OCT 2018

A three arm randomized, open-label Phase II study of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) versus 80 kBq/kg (88 kBq/kg after implementation of NIST update) , and versus 50 kBq/kg (55 kBq/kg after implementation of NIST update) in an extended dosing schedule in subjects with castration-resistant prostate cancer metastatic to the bone

[Standard dose versus high dose and versus extended standard dose radium-223 dichloride in castration-resistant prostate cancer metastatic to the bone]

BSP study drug	BAY 88-8223 / Radium-223 dichloride / Xofigo		
[Study purpose:]	Efficacy, dose-Finding		
Clinical study phase:	II	Date:	17MAY2017
Study No.:	16507	Version:	Final 2.0
Author:	PPD [REDACTED]		
	PPD [REDACTED]		
	PPD [REDACTED]		

Confidential

The information provided in this document is strictly confidential and is intended solely for the guidance of the clinical investigation. Reproduction or disclosure of this document, whether in part or in full, to parties not associated with the clinical investigation or its use for any other purpose without the prior written consent of the sponsor is not permitted.

Throughout this document, symbols indicating proprietary names (®, TM) are not displayed. Hence, the appearance of product names without these symbols does not imply that these names are not protected.

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

1.	Introduction.....	4
2.	Study Objectives	5
3.	Study Design.....	7
3.1	Study periods and duration	9
3.2	Schedule of procedures.....	9
3.3	Planned Analyses.....	9
3.4	Determination of sample size	9
4.	General Statistical Considerations	11
4.1	General Principles.....	11
4.2	Handling of Dropouts	11
4.3	Handling of Missing Data.....	11
4.3.1	Imputation rules	11
4.4	Interim Analyses and Data Monitoring	11
4.5	Data Rules.....	12
4.5.1	Determining Baseline Values	12
4.5.2	Pooling cut-off date for Arm C subjects in Comparison 1	12
4.6	Validity Review	12
4.6.1	Protocol Deviations	13
4.7	Additional Statistical Considerations	13
4.7.1	Delayed Effect Efficacy Assumptions.....	13
5.	Analysis Sets.....	13
6.	Statistical Methodology	14
6.1	Population characteristics	14
6.1.1	Disposition of Subjects.....	14
6.1.2	Demographic and Baseline Characteristics	14
6.1.3	Medical history	16
6.1.4	Extent of Exposure	16
6.1.5	Prior and Concomitant Medications	17
6.2	Efficacy.....	18
6.2.1	Primary Efficacy Analysis.....	18
6.2.2	Secondary Efficacy Analysis.....	20
6.2.3	Exploratory Analysis	22
6.3	Pharmacokinetics / pharmacodynamics.....	29
6.4	Safety.....	29
6.4.1	Adverse events.....	29
6.4.2	Deaths	31
6.4.3	Clinical Laboratory Data	31
6.4.4	New Primary Malignancies	33
6.5	Analysis of Other Endpoints.....	33
6.5.1	ECOG Performance Status	33

6.5.2	Quality of Life	33
6.5.3	Long Term Safety	33
6.6	Examination of Subgroups	33
7.	Document history and changes in the planned statistical analysis	34
8.	References.....	35
9.	Appendices.....	37
9.1	Censoring Rules.....	37
9.1.1	Censoring rules for Symptomatic Skeletal Event-Free Survival (SSE-FS) ...	37
9.1.2	Censoring rules for Overall Survival (OS).....	38
9.1.3	Censoring rules for other time-to-event endpoints	39
9.1.4	Censoring rules for time to radiological progression	40
9.1.5	Censoring rules for radiological progression-free survival (rPFS)	42

Abbreviations

AD	Associated Document
AE	Adverse event
ALP	Alkaline phosphatase
APAC	Asia-Pacific
ATC	Anatomical Therapeutic Class
BPI-SF	Brief Pain Inventory – Short Form
BSP	Bayer Schering Pharma
CBC	Complete blood count
CRPC	Castration-resistant prostate cancer
CI	Confidence interval
CSR	Clinical Study Report
CT	Computer tomography
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee
DRS-E	Disease Related Symptoms – Emotional
DRS-P	Disease Related Symptoms – Physical
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
ePRO	Electronic Patient Reported Outcomes
F/WB	Function/Well Being
FAS	Full Analysis Set
FACT	Functional Assessment of Cancer Therapy
FPSI	Functional assessment of cancer therapy Prostate Symptom Index
FPSI-17	Functional assessment of cancer therapy Prostate Symptom Index – 17 items
FU	Follow-Up
GCP	Good Clinical Practices
GMS	Global Medical Standards
HEOR	Health Economics, Outcomes & Reimbursement
IBW	Ideal Body Weight
ICH	International Committee on Harmonization
ID	Identifier
INN	International Non-proprietary Name
ITT	Intent-to-treat
IV	Intravenous

IVRS/IWRS	Interactive voice/web response system
IXRS	Interactive response system (independent of type)
JSRC	Joint Safety Review Committee
kBq	Kilobecquerel
LCL	Lower Confidence Limit
LTFU	Long-Term Follow-Up
LTFUAE	Long-Term Follow-Up Adverse Event
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
NCI	National Cancer Institute
NCCN	National Comprehensive Cancer Network
NIS	Non-interventional Study
NIST	National Institute of Standards and Technology
nMiss	Number of missing values
OS	Overall survival
PC	Prostate cancer
mPCWG2	Modified Prostate Cancer Working Group 2
PRO	Patient Reported Outcomes
PSA	Prostate specific antigen
PT	Preferred Term
QoL	Quality of Life
Ra-223	Radium-223 dichloride
rPFS	Radiological progression free survival
SAE	Serious Adverse Event
SAF	Safety
SAP	Statistical Analysis Plan
SD	Standard Deviation
SMQ	Standardized MedDRA Queries
SOC	System Organ Class
SOP	Standard Operating Procedure
SSE	Symptomatic skeletal event
SSE-FS	Symptomatic skeletal event free survival
Std	Standard Deviation
TAS	(Bayer) Therapeutic Area (Oncology) Standards
TBW	Total Body Weight
TEAE	Treatment-Emergent Adverse Event
TTP	Time to progression
uCi	Microcurie
UCL	Upper Confidence Limit
W12	Week 12
W24	Week 24
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary

1. Introduction

Protocol 16507 is a 3-arm, randomized, open-label Phase II study to compare the efficacy and safety of a “standard dose” regimen of radium-223 dichloride with, a “high dose” regimen and an “extended dose” regimen, in subjects with castration-resistant prostate cancer metastatic to bone .

The study's 3 arms are:

- Arm A: 6 doses of 50 kBq/kg (55 kBq/kg after implementation of NIST update) radium-223 dichloride every 4 weeks ("standard" dose regimen)
- Arm B: 6 doses of 80 kBq/kg (88 kBq/kg after implementation of NIST update) radium-223 dichloride every 4 weeks ("high dose" regimen)
- Arm C: 12 doses of 50 kBq/kg (55 kBq/kg after implementation of NIST update) radium-223 dichloride every 4 weeks ("extended dose" regimen)

The study's co-primary endpoints are:

- Symptomatic skeletal event free survival from randomization (SSE-FS1) for the "standard dose" vs. "high dose" comparison (Comparison 1).
- Symptomatic skeletal event free survival from the 6th dose (SSE-FS2) for the "standard dose" vs. "extended dose" comparison (Comparison 2)

Two analyses of efficacy and safety are planned in this study:

- A final analysis to occur after the latest of the following maturation criteria are met:
 - 135 SSE-FS (SSE-FS1) events in subjects included in Comparison 1
 - 75 SSE-FS (SSE-FS2) events following the 6th dose in subjects included in Comparison 2
 - The last subject has been followed for 30 days from last treatment.
- An updated analysis to occur after the last subject has completed the active follow-up period in this study for two years following last dose of Ra-223 dichloride (if after primary endpoint cut-off).

Additional details on the timing of the analyses are described in Section 3.3 below.

This Statistical Analysis Plan (SAP) specifies the analyses and data presentations planned for both the final and the updated analysis.

This SAP was written based on the following documentation:

Document	Date	Version
Protocol	9 August 2013	1.0
Protocol Amendment no. 1	20 November 2013	2.0
Protocol Amendment no. 2	16 August 2015	3.0
Protocol Amendment no. 3	16 May 2017	4.0

2. Study Objectives

Co-primary objectives:

- To evaluate efficacy as measured by SSE-FS of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 doses

compared to radium-223 dichloride 80 kBq/kg (88 kBq/kg after implementation of NIST update); and

- To evaluate efficacy as measured by SSE-FS of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 additional doses compared to no further radium-223 dichloride treatment in subjects with CRPC metastatic to the bone who previously received radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 doses, and survived SSE free and are eligible for further radium-223 dichloride treatment.

Symptomatic skeletal events (SSEs) are defined as:

- The use of external beam radiotherapy (EBRT) to relieve skeletal symptoms
- New symptomatic pathological bone fractures (vertebral and non-vertebral)
- Tumor-related orthopedic surgical intervention
- Spinal cord compression

Secondary objectives:

- To evaluate safety and tolerability
- To evaluate overall survival (OS)
- To evaluate pain improvement rate
- To evaluate time to pain progression
- To evaluate time to first SSE
- To evaluate time to radiological progression
- To evaluate radiological progression-free survival (rPFS)

Exploratory objectives:

- To evaluate quantitated whole body technetium-99 bone scan tumor burden area and index, and determine bone tumor response
- To explore the impact of patient body size on the efficacy and safety of radium-223 dichloride
- Time to increase in physical symptoms of disease based on the NCCN FACT-P Symptom Index-17 (NCCN-FACT FPSI-17) physical disease related symptoms (FPSI-DRS-P) subscale score measured up to Week 48 after the start of treatment
- To evaluate laboratory indicators of efficacy, including:
 - PSA response
 - Time to PSA progression
 - ALP response
 - Time to ALP progression

- Percentage change in ALP from baseline
- To evaluate change in analgesic use.

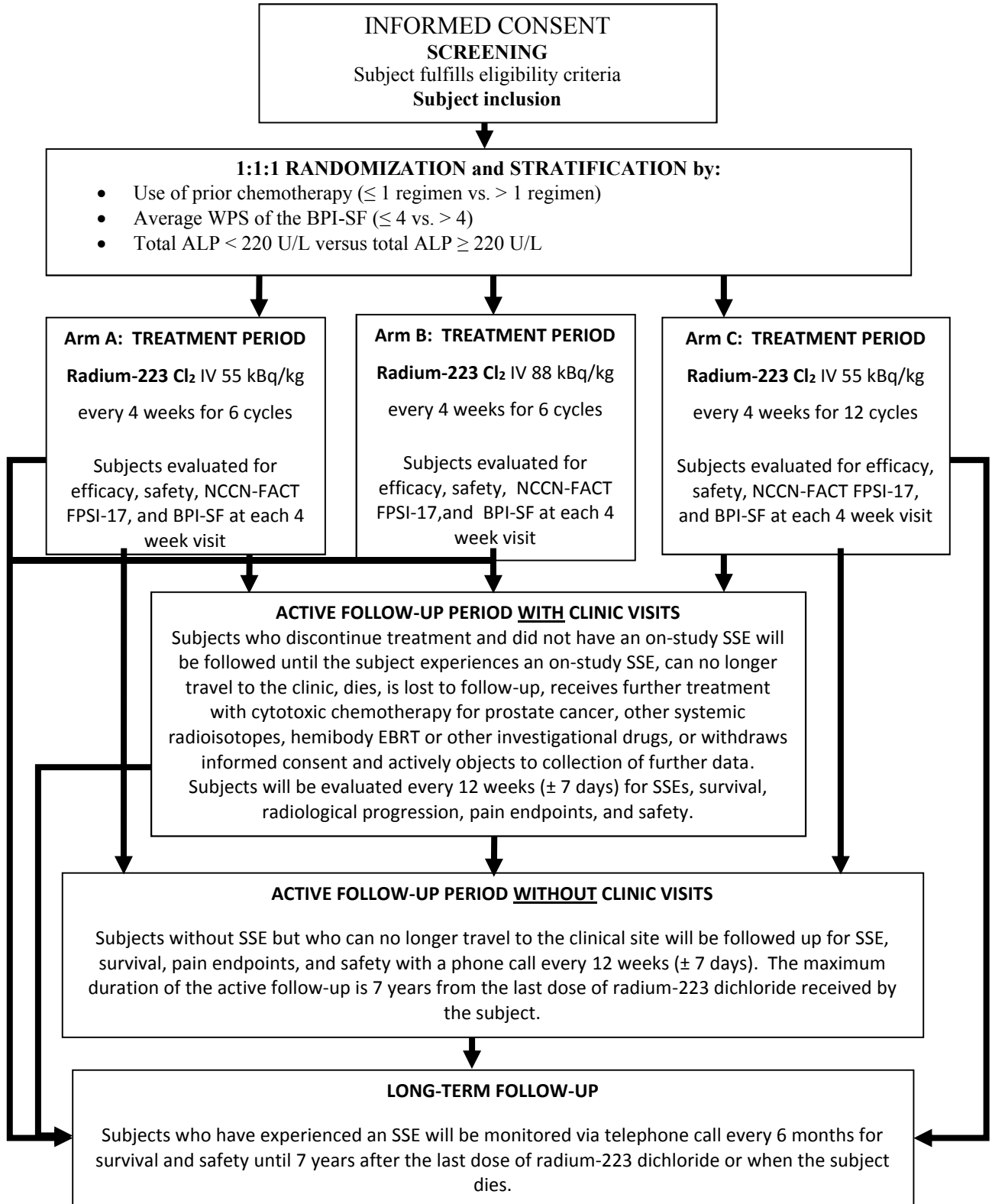
3. Study Design

This is a three arm, randomized, open-label Phase II study of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) versus 80 kBq/kg (88 kBq/kg after implementation of NIST update) , and versus 50 kBq/kg (55 kBq/kg after implementation of NIST update) in an extended dose schedule in subjects with castration-resistant prostate cancer metastatic to the bone. Approximately 360 subjects will be randomized in this study.

The randomization will be stratified by use of prior chemotherapy (≤ 1 regimen versus > 1 regimen), by total ALP (< 220 U/L versus ≥ 220 U/L), and by average worst pain score (WPS) of the Brief Pain Inventory – Short Form (BPI-SF) (WPS ≤ 4 versus WPS > 4).

A schematic of the study design is presented in [Figure 3–1](#).

Figure 3–1: Study Design Schematic



3.1 Study periods and duration

The study periods will consist of screening/randomization, treatment, active follow-up with clinic visits, active follow-up without clinic visits, and long-term follow-up.

Long-term follow-up:

All study subjects will be eligible for further follow-up either in this study or in a separate long-term follow-up study for up to 7 years.

Transition of all subjects into the separate long-term follow-up study will begin following the later of primary endpoint completion or 2 years after the last radium-223 dichloride treatment is administered to the last treated subject.

Until the transition to the long term follow-up study begins, some subjects may begin long-term follow-up in this study. This study will end when all subjects have transitioned into the long-term follow-up study or discontinued from this study for another reason (e.g., Consent Withdrawn, Lost to Follow up).

Data collected in the separate long-term follow-up study is outside the scope of this SAP.

3.2 Schedule of procedures

See Protocol Table 7-1.

3.3 Planned Analyses

The final analysis of this study will take place after the primary endpoint cut-off is reached.

The primary endpoint cut-off will be reached at the later of

- 135 SSE-FS events in subjects included in Comparison 1, and
- 75 SSE-FS events following the 6th dose in subjects included in Comparison 2, and
- the last subject has been followed for 30 days from last treatment.

3.4 Determination of sample size

The sample size is based on two separate comparisons with no multiple-testing adjustment:

Comparison 1:

H01: $SSE-FS_{\text{high dose}} = SSE-FS_{\text{standard dose}}$, versus

HA1: $SSE-FS_{\text{high dose}} > SSE-FS_{\text{standard dose}}$.

This comparison will consider SSE-FS following randomization. Arm C subjects will be pooled with Arm A for the standard dose group. Refer to pooling algorithm in Section 4.5.2 and Censoring rules in Table 9–1.

Comparison 2:

H02: SSE-FS_{extended dose} = SSE-FS_{Standard dose}, versus

HA2: SSE-FS_{extended dose} > SSE-FS_{Standard dose}.

This comparison will include Arm A and Arm C subjects who received 6 or more doses.

The sample size calculations assume 30 subjects/month enrollment with 7 months ramp-up and a 3% loss to follow-up per month. The accrual period is anticipated to last approximately 16 months. Overall, 360 subjects are expected to be randomized at a 1:1:1 ratio. The primary endpoint cut-off, as defined in Section 3.3, is estimated to be reached in an average of 39.1 months.

- Comparison 1 assumes:
 - Arm A (standard dose) subjects have constant median SSE-FS of 9 months throughout the study
 - Arm C (extended dose) subjects have median SSE-FS of 9 months during the first 24 weeks of the study
 - Arm B (high dose) subjects have constant 50% improvement in SSE-FS (median 13.5, hazard ratio 0.667) .

Simulations indicated an average of 186.61 events would occur for this comparison at primary endpoint cut-off, and a 1-sided log-rank test with 0.10 significance gave approximately 90.5% power to test H01 versus HA1.

- Comparison 2 assumes:
 - Arm A (standard dose) subjects have constant median SSE-FS of 9 months
 - Arm C Extended dose subjects have
 - Median SSE-FS of 9 months for the first 24 weeks
 - A 65% improvement (median 14.85, hazard ratio 0.606) thereafter.

With 240 of 360 subjects randomized to Treatment Arm A and Treatment Arm C, simulations indicated that 44.65% of randomized subjects (107.2 of 240) would have a SSE-FS event or loss to follow-up during the first 24 weeks, leaving an estimated 55.35% (132.8) surviving SSE-free and eligible for inclusion at 24 weeks. An average of 74.91 simulated events occurred for this comparison, and a 1-sided log-rank test with 0.10 significance (0.2 for two-sided test) gave approximately 80.6% power to test H02 versus HA2.

Simulations assume continuous event and censoring times. Last treatment is assumed to occur 44 weeks (10.12 months) after last subject accrued.

Primary endpoint power calculations were based on simulations programmed using SAS version 9.2 and used 10,000 replicates.

The simulations, taking into account the correlated nature of the data supporting the two comparisons in this study design, indicated that the proportion of replicates rejecting at least one null hypothesis was 16.4%, and the proportion of replicates rejecting both was 2.3%.

4. General Statistical Considerations

4.1 General Principles

The statistical evaluation will be performed by using the software package SAS release 9.2 or higher (SAS Institute Inc., Cary, NC, USA). Descriptive statistics will summarize the number of data available and missing data, mean, standard deviation, minimum, quartiles, median, and maximum for continuous data, and frequency tables (number of data available and missing, and percent of available data) for categorical data.

4.2 Handling of Dropouts

A “dropout” is defined as a randomized subject who discontinues study participation prior to start of study treatment for any reason.

Dropouts and subjects withdrawn from study treatment will not be replaced. Refer to Section 5.2.1 in the study protocol for withdrawal of subjects from study.

4.3 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.

4.3.1 Imputation rules

Unless otherwise specified data will not be imputed, and rules for imputation are specified in the dataset specification documents.

4.4 Interim Analyses and Data Monitoring

No interim analysis for efficacy is planned.

A Joint Safety Review Committee (JSRC) will perform monitoring for safety and protocol compliance. The JSRC will meet at regular intervals as specified by the JSRC charter, approximately every 6 months, and may decide to meet more frequently. The JSRC will be supported by an independent Statistical Analysis Center (SAC) as specified in the Charter. Reports to the JSRC will be specified in a separate document.

The conduct of the JSRC is generally similar to a Data Monitoring Committee, except that membership may include sponsor, investigator, and/or independent members, and interim results of this open-label trial may be disclosed to the sponsor. Further details are described in the JSRC charter.

4.5 Data Rules

4.5.1 Determining Baseline Values

This study has two distinct baselines for efficacy analyses, the randomization baseline and the Week 24 baseline.

Randomization baseline value for efficacy parameters: is defined as the last non-missing value on or before the date of randomization.

Week 24 baseline: Unless otherwise specified, the ‘Week 24 baseline’ is defined as the date of the 6th injection. This baseline is defined only for those subjects who are eligible for Comparison 2 analyses. That is, subjects who had received at least 6 injections.

Week 24 baseline value for efficacy parameters: Unless otherwise specified, the ‘Week 24 baseline value’ is defined as the last non-missing value on or before the date of the 6th injection.

Ideally, this visit would be the date of the 7th injection for Arm C subjects and expected 7th injection date for Arm A subjects. However, the end of treatment visit (30±7 days from last dose) for Arm A subjects could be missing due to consent withdrawal or happen out of visit window. Hence, there is no comparable visit for Arm A subjects that could be used as the expected 7th injection date.

To account for this and achieve comparability between the two arms, the 6th injection date is used as the reference date for Comparison 2 subjects. It should be noted that this could lead to a somewhat dilution of the treatment effect in Arm C due to an early reference date.

Baseline value for safety parameters: is defined as the last non-missing value on or before the date of first dose.

4.5.2 Pooling cut-off date for Arm C subjects in Comparison 1

For Comparison 1, Arm C subjects are planned to be pooled with Arm A subjects. This pooling will be done such a way that only comparable data from Arm C subject pooled. The following algorithm will be used to identify the cut-off date for pooling:

Arm C: Pooling cut-off date = $7*28 + \text{randomization date}$; for subjects that were randomized but not treated,
= $(7-x)*28 + \text{last dose date}$; for $1 \leq x \leq 6$,
= 7th dose date; for $x \geq 7$,
where x be the number of injections received.

4.6 Validity Review

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.

4.6.1 Protocol Deviations

Assessment criteria, deviation definitions, are described in a separate Protocol Deviations document.

4.7 Additional Statistical Considerations

4.7.1 Delayed Effect Efficacy Assumptions

Since the comparison between the “standard” and “extended” regimens involves identical treatment for the first 6 doses, the following assumptions were made:

- Both regimens have identical efficacy for the first 6 doses, and
- The efficacy would diverge at the time that extended-dosing subjects received their 7th dose.

This feature is consistent with a delayed effect as described in Fine (2007) [1] and Zhang and Quan (2009) [2]. To account for this delayed-effect feature, hazards were modeled using a 2-compartment approach:

- With the standard (Arm A) and high-dose (Arm B) regimens having constant hazards throughout the study, and
- The extended regimen (Arm C) initially having the same hazards as the standard regimen (Arm A) but having a fixed change point at Week 24, approximately the point at which the 7th dose would begin, with constant, reduced hazards thereafter.

A sensitivity analysis using two-compartment Cox model will be performed by taking into account of this delayed-effect feature for Comparison 2. The details are specified in Section 6.2.1.

5. Analysis Sets

- **Intent-to-treat (ITT):** All randomized subjects. The ITT population will be used in the analysis of efficacy endpoints for Comparison 1. Subjects will be included in ITT analyses according to the treatment to which they were randomized as based on IXRS. The randomization baseline will be used in the ITT analyses.
- **Week 24(W24):** All ITT subjects in Arm A (standard dose) and Arm C (extended dosing) treated with radium-223 dichloride and eligible for further treatment at W24 (i.e., 7th injection). Week 24 applies to subjects who received the 6th injection without any dose delays. As dose delays are allowed per protocol, 6th injection can occur after Week 24 for subjects who had dose delays. Hence, this population is defined based on the number of injections not based on the timing. The W24 dataset

will be used for the analysis of efficacy endpoints related to Comparison 2 and associated evaluations. Subjects will be included in W24 analyses according to the treatment to which they were randomized.

- **Safety (SAF):** All subjects who have received at least one study drug administration. 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.

6. Statistical Methodology

6.1 Population characteristics

6.1.1 Disposition of Subjects

The number and percentage of subjects screened, randomized, and treated will be presented by treatment group and overall. In addition, the number of subjects discontinued prior to receiving the 6th injection in Arm A and C will be tabulated. The reasons for subjects discontinuing from treatment will be summarized by treatment group. In addition, the number of subjects screened and included in each analysis population will be displayed overall and by country and investigator.

6.1.2 Demographic and Baseline Characteristics

Descriptive summaries of demographics and baseline characteristics will be presented by treatment group and overall for the ITT, Safety and W24 populations whenever applicable. Comparability of the treatment groups with respect to demographics and baseline characteristics will be assessed using descriptive summaries.

For these tables, the summaries refer to the randomization baseline by default, using the ITT or Safety population. Additional summaries for the W24 population, using week 24 baseline, are identified below as “(W24)”. ITT population tables will be reported both by treatment arm, and by pooled arms (A & C).

The following demographic data will be summarized:

- Age (years)
- Age category (< 65, 65 – 74, 75-84, ≥85 years)
- Race and ethnicity
- Height (cm)
- Weight (kg) (W24)
- Vital signs: blood pressure (mm Hg), heart rate (bpm), respiratory rate (rpm), and temperature (°C)

The following baseline characteristics will be summarized:

- Stage of prostate cancer at initial diagnosis (TNM)
- Histology Stage of prostate cancer at study entry (TNM)

- Status of primary tumor at study entry
- Gleason score at initial diagnosis of prostate cancer
- PSA and total ALP(W24)
- ECOG Performance Status (W24)
- Number of bone lesions (1-5, 6-20, >20 but not a superscan vs. superscan at baseline)
- Bone scan lesion area
- BPI-SF average WPS (W24)
- Time from initial diagnosis to randomization for prostate cancer and bone metastases (months)
- Time from first progression to randomization for prostate cancer and bone metastases (months)
- Time from most recent progression to randomization for prostate cancer and bone metastases (months)
- Time from first progression to most recent progression for prostate cancer and bone metastases (months)

Categorical summaries of each of three randomization stratification factors per IXRS:

- Use of prior chemotherapy (≤ 1 regimen versus > 1 regimen)
- Total ALP (< 220 U/L versus ≥ 220 U/L) and
- Average worst pain score (WPS) of the BPI-SF at randomization ($WPS \leq 4$ versus $WPS > 4$), will also be presented for the ITT and W24 populations.

The discordance for each stratification factor between CRF and IXRS data will be presented in a shift table.

In addition, average WPS at randomization, as used to calculate the average WPS randomization stratification factor, will be summarized.

The following body descriptors will be summarized as described in Green and Duffull (2004) [3]:

- Total body weight(TBW), defined as the weight (kg) used for calculating the first dose
- Total body weight (W24), defined as the Week 24 baseline weight (kg)
- Body Mass Index, and Body Mass Index (W24), calculated as:

$$BMI = \frac{Weight (kg)}{[Height (m)]^2}$$

- Ideal Body Weight(IBW) will be calculated as:
 $IBW = 50.0 + [0.89 \times (HT(cm)-152.4)]$ for male

Subjects will be classified by BMI as low weight ($BMI < 18.5$), normal weight ($18.5 \leq BMI < 30$), or high weight ($BMI \geq 30$) at each baseline [4]. Baseline BMI will also be categorized as low versus normal/high.

6.1.3 Medical history

Medical history will be summarized by MedDRA body system organ class (SOC) and preferred term (PT) for the ITT population overall, by treatment group, and for pooled arms A & C.

6.1.4 Extent of Exposure

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

- Duration of study treatment will be calculated in days as the date of the last dose of study – date of the first dose of study treatment + 1
- Number of radium-223 dichloride injections received. Summary statistics will include both descriptive statistics of number of doses received, and number and percent of subjects by number of doses received
- Total activity of radium-223 dichloride (sum of activity in all doses) in kBq and kBq/kg
- Dose intensity of radium-223 dichloride calculated as total activity of radium-223 dichloride/(duration of study treatment + 27), in kBq/day and KBq/kg/day
- Number and percent of subjects with dose modifications (none, at least one interruption/delay)
- Normalized dose for total body weight (TBW) and ideal body weight (IBW) will be summarized by quartile:

Total body weight normalized dose=Activity in the first dose administered to subject /body weight used for first dose

Ideal body weight normalized dose= Activity in the first dose administered to subject /ideal body weight

All exposure summaries will be based on the old standard (50 or 80 kBq/kg). Hence, doses administered with the new standard (55 or 88 kBq/kg) will be converted to the old standard prior to any analyses as specified below. A country specific 2016 NIST implementation date will be utilized for this conversion.

Radioactivity will be calculated as follows:

- Activity per injection (kBq) = Activity in syringe before injection (kBq) - activity in syringe after injection (kBq)
- Activity adjusted
 1. If the dose date is prior to 2016 global NIST implementation date in the country, then Activity per injection = Activity per injection

2. If the dose date is after the 2016 global NIST implementation date in the country, then Activity per injection = Activity per injection/1.1

- Total activity injected (kBq) = Sum of ('Activity per injection')

6.1.5 Prior and Concomitant Medications

All investigator-reported non-study medications taken before and/or during the study will be coded using the Bayer modified version of 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 before and/or during the study will be categorized as prior medications, concomitant medications, and post treatment medications.

Classifications of prior and concomitant medication as based on [Table 6–1](#).

If a start date is missing, the medication will be assumed to start prior to first dose of study drug. If the end date is unknown and 'ongoing' was not checked in CRF, the medication will be assumed to end at the last visit date, death date or withdrawal from study date, whichever is the latest.

All concomitant medications will be listed, including verbatim descriptions and coded terms, and flags for prior/concomitant/post treatment medications. Prior, concomitant, and post treatment medications 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.

Summaries will also be produced for anti-cancer therapies (systemic therapy, radiotherapy as well as diagnostic and therapeutic procedures). These summaries will be created for ITT population and Week 24 population.

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: C7 and C8 are covered under C6.

6.2 Efficacy

For Comparison 1, Arm A will be pooled with Arm C in the time-to-event analyses. Refer to censoring rules in Section 4.5 for the pooling algorithm. No pooling of Arm C subjects will occur in Comparison 1 for non-time-to-event endpoints (e.g. pain improvement rate).

As sensitivity analyses, the analyses described for Comparison 1 will also be performed between Arm A and Arm B only (i.e., without pooling Arm C data).

Each efficacy endpoint will be defined separately for Comparison 1 and Comparison 2. For Comparison 1 analyses, endpoints will be defined following randomization baseline. For Comparison 2 analyses, endpoints will be defined following Week 24 baseline (the 6th dose date) unless otherwise specified.

Imaging endpoints based on tumor assessments, will not follow the Week-24 baseline for Comparison 2 analyses. The study did not plan to do Week 24 re-baseline for the site investigators. The site investigators would determine PD only based on the randomization baseline. Once a subject had a PD (or a confirmed PD for bone progression) based on randomization baseline, the subject will not be further followed for a second PD using Week-24 assessment as the new baseline. Therefore, there will be no assessments after an observation of PD to perform re-reads of tumor assessments based on the Week-24 baseline by the central reviewer. This incomplete tumor assessments can lead to a biased analysis favoring one arm vs the other (i.e., the arm with more PDs based on randomization baseline could have more censored data for the analyses based on Week-24 baseline). Hence, the Comparison 2 analysis on the Week-24 population will only be performed based on the randomization baseline, not the Week-24 baseline for both investigator and centrally reviewed data.

6.2.1 Primary Efficacy Analysis

The primary efficacy endpoint, symptomatic skeletal event-free survival (SSE-FS), will be defined separately for each of the two efficacy evaluations described above, as follows:

For Comparison 1, SSE-FS following randomization is defined in ITT subjects as the time from randomization to an SSE or death, whichever occurs first. Refer to Appendix 9.1.1 for censoring rules.

For Comparison 2, SSE-FS from 6th dose is defined in W24 subjects as the time from Week 24 baseline (the 6th dose date) to an SSE or death, whichever occurs first. Refer to Appendix 9.1.1 for censoring rules.

For both evaluations, a symptomatic skeletal event (SSE) is defined as follows:

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

The censoring rules for SSE-FS for each comparison are summarized in [Table 9–1](#) and [Table 9–2](#) of [Appendix 9.1.1](#). Stratified statistical tests described in this section refers to two of the three randomization strata as determined by the IXRS: use of prior chemotherapy (≤ 1 regimen versus > 1 regimen), and total ALP (< 220 U/L versus ≥ 220 U/L). Note that Average worst pain score (WPS) was included as one of stratification factor for randomization in order to balance the baseline pain across three arms. However, WPS will not be included in the analysis to avoid sparse cells due to relatively small sample size within each arm.

Each analysis will use a stratified log-rank test. A one-sided significance level of 0.10 will be used for interpretation of the results. Additionally, the hazard ratio (respectively high dose/standard dose and extended dose/standard dose) will be computed together with the two sided -80% and -95% CIs using a stratified Cox regression model. There will be no adjustment for multiple comparisons.

SSE-FS will also be summarized using Kaplan-Meier [5] estimates. Kaplan-Meier [5] curves will be generated, and median survival time together with the 25th and 75th percentiles and associated Brookmeyer-Crowley [6] 80% and 95% CIs will be presented separately for the two comparisons. Kaplan-Meier plots will be presented for the time from randomization for both Comparison 1 and 2. Additionally, similar descriptive summaries will be created for Week 24 population using the randomization baseline.

For each comparison, the contribution of each component of the SSE will be evaluated. Descriptive statistics will be presented.

A Kaplan-Meier [5] curves will be generated for each BMI category (< 18.5 , $\geq 18.5 - < 30$), and ≥ 30) by treatment group within the 2 comparisons. The effect of low baseline BMI status on efficacy will be addressed by including baseline BMI category as a binary covariate (low versus normal/high) in a stratified Cox proportional hazards model.

A 2-compartment Cox model will be fitted as a sensitivity analysis to explore differences between arms A and C for the two time periods before and after the 6th dose, with period as a time-varying covariate. Hazard ratio (and CIs) will be reported for each time period and treatment arm.

Supportive analysis for SSE-FS will be performed for each comparison using a non-stratified log-rank test. The treatment effect (hazard ratio) will be estimated using an unstratified Cox model.

The comparison 2 for SSE-FS will be also performed using randomization baseline as supportive analysis.

As an exploratory analysis, stratified Cox proportional hazards regression models may be fitted for each comparison including other applicable baseline covariates considered to be of prognostic importance. Time-dependent covariates that may be investigated individually in the model include: non-study systemic anti-cancer therapy, increase in analgesics use, radiological progression, pain progression, and pain improvement.

6.2.2 Secondary Efficacy Analysis

6.2.2.1 Secondary efficacy endpoints

The secondary efficacy variables are specified below.

- Overall survival
- Time to first SSE
- Radiological progression-free Survival (rPFS)
- Time to radiological progression (TTP)
- Pain improvement rate
- Time to pain progression

Overall survival is defined for each comparison as the time (days) from the applicable start date to the date of death due to any cause. See [Table 9–3](#) in Appendix 9.1.2 for detailed censoring rules.

Comparison 1:

Arm C subjects who have died or are known to be alive after ‘pooling cut-off date’ will be censored on the ‘pooling cut-off’ date. Refer to Section 4.5 for pooling cut-off date.

All other subjects in arms A, B and C, refer to [Table 9–3](#) for censoring rules.

Overall Survival (OS) following randomization = End Date – Date of Randomization +1

Comparison 2:

For Comparison 2 eligible subjects, refer to [Table 9–3](#) for censoring rules.

Overall Survival (OS) following Week 24 baseline = End Date – Date of Week 24 baseline(the 6th dose date) +1

Time to first symptomatic skeletal event is defined for each comparison as the time (days) from the applicable start date to the date of the first SSE. See

[Table 9–4](#) in Appendix 9.1.3 for detailed censoring.

Time to radiological progression is defined for each comparison as the time (days) from the randomization date to the date of the first radiological progression. Comparison 2 will not follow the Week-24 baseline and randomization baseline will be used instead. See [Table 9–5](#) and [Table 9–6](#) in Appendix 9.1.4 for detailed censoring rules.

The central reviewer will evaluate the radiological imaging produced by the investigator. Primary radiological progression endpoints will be based on the central radiological reviewer assessments. Investigator assessments will be provided as a sensitivity analysis.

Radiological progression of soft-tissue disease is determined according to modified RECIST criteria, version 1.1 (See Protocol Section 14.1) based on MRI / CT scans of the chest, abdomen, and pelvis.

Radiological progression of bone disease is determined according to adapted PCWG2 criteria (see Protocol Section 14.2) based on whole body technetium-99 bone scans.

Radiological progression of bone disease requires confirmation. 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).

The date of progression of bone disease will be the date of the first observed bone progression, not the date that confirmed the bone progression.

The date of radiological progression will be determined following the rules in [Table 9–6](#).

Radiological progression-free survival is defined for each comparison as the time (days) from the randomization baseline to the date of radiological disease progression or death from any cause (if death occurs without prior progression). See [Table 9–7](#) in Appendix 9.1.5 for detailed censoring rules for both Comparison 1 and Comparison 2.

If data permits, unconfirmed radiological progression-free survival and time to unconfirmed radiological progression will be analyzed as sensitivity analyses based on central reviewer. Bone progression in this case does not require confirmation.

Pain endpoints

The pain data collected from ePRO will be used for these analyses.

Worst pain score (WPS) at a given visit is defined for each subject as the mean of the 7 day “worst pain score” assessments taken on or before each applicable visit. If less than 4 non-

missing WPS values are available for a given visit, then the weekly worst pain score will be left missing for that visit.

Timepoint pain improvement rate is defined with respect to each applicable baseline as the proportion of subjects with a 30% and 2-point decrease in WPS with respect to the applicable baseline assessment over 2 consecutive assessment periods conducted at least 3 weeks apart among subjects with a WPS score ≥ 4 at the applicable baseline. The first of the 2 consecutive periods is taken to be the date of improvement.

Time to pain progression is defined for each comparison in subjects with a WPS of ≤ 7 at the applicable baseline as the time from the applicable start date to an increase of 2 or more points in WPS (with respect to the applicable baseline). Refer to

Table 9–4 in Appendix 9.1.3 for detailed censoring rules.

6.2.2.2 Analysis of secondary efficacy endpoints

The following secondary efficacy time-to-event endpoints will have hypothesis testing performed for both Comparison 1 and Comparison 2: OS, time to first SSE, time to radiological progression, and rPFS.

A one-sided type I error rate of 0.10 will be used for secondary efficacy endpoint hypothesis testing.

Hypothesis tests for these endpoints will be performed using a stratified log-rank test accounting for the same two randomization stratification factors as for the primary endpoint analysis (See Section 6.2.1). The treatment effect (hazard ratio) will be estimated using the Cox proportional hazards regression model stratified by the above-mentioned two randomization stratification factors. For rPFS and time to radiological progression, analyses will be done based on both the investigator's and the independent assessments. Analysis based on the independent central review assessment will be considered the primary analysis for rPFS.

Secondary efficacy time-to-event endpoints will be summarized using Kaplan-Meier [5] estimates. Median survival time together with the 25th and 75th percentiles and associated 80% and 95% Brookmeyer-Crowley [6] confidence intervals (CI) will be presented by group. Corresponding Kaplan-Meier [5] curves will be generated by group.

Pain improvement rate will be calculated for each post-baseline timepoint, with exact binomial 80% and 95% Clopper-Pearson [7] confidence intervals. Pain improvement rate with exact binomial 80% and 95% CIs will also be plotted by timepoint.

6.2.3 Exploratory Analysis

For exploratory radiological endpoints, only central reviewer assessments will be analyzed.

6.2.3.1 Exploratory endpoints

The exploratory variables are specified below:

- Time to radiological bone progression
- Time to radiological soft-tissue progression
- Patient bone scan lesion area (BSLA)
- Patient bone scan time point response rate (week 8, week 16 and week 24) and best overall response rate, based on central review
- Total alkaline phosphatase (ALP) response rate
- Time to total alkaline phosphatase (ALP) progression
- Time to total alkaline phosphatase (ALP) progression without 12 week restriction
- Percentage change in total alkaline phosphatase (ALP) at 12 and 24 weeks
- Prostate specific antigen (PSA) response rate
- Time to PSA progression
- Time to PSA progression without 12 week restriction
- Pain improvement rate without increase in analgesic use (week 12, week 24, EOT and overall)
- Time to pain progression or increase in analgesic use
- Change in analgesic use
- Time to increase in analgesic use
- NCCN-FACT FPSI-17 subscale scores
 - FPSI Disease Related Symptoms – Physical (DRS-P)
 - FPSI Disease Related Symptoms – Emotional (DRS-E)
 - FPSI Treatment Side Effects (TSE)
 - FPSI Function/Well Being (F/WB)
- Time to increase in physical symptoms of disease based on FPSI-DRS-P
- BPI-SF subscale scores and aggregate indices
 - BPI-SF Pain Severity Index
 - BPI-SF Function Interference Index
- Average Pain Score (APS)

6.2.3.1.1 Exploratory endpoints

For all exploratory time-to-event endpoints, randomization baseline will be used for all the analyses unless otherwise specified. Week 24 baseline will not be defined for exploratory endpoints.

Radiological endpoints

Only central reviewer assessments will be analyzed for exploratory radiological endpoints.

Time to radiological bone progression is defined for each comparison as the time (days) from randomization date to the date of confirmed radiological bone progression. See [Table 9–5](#) in Appendix [9.1.4](#) for detailed censoring rules.

Time to radiological soft-tissue progression is defined as the time (days) from randomization date to the date of radiological soft tissue progression. See [Table 9–5](#) in Appendix [9.1.4](#) for detailed censoring rules.

Quantitated bone scan endpoints

Digitized images of whole body quantified technetium-99 bone scans will be evaluated by the central imaging reviewer using CAD system software using the method of Brown et al. (2012) [8], as described in the imaging charter. The central reviewer will use the CAD system software to evaluate each subject’s digitized whole body quantified technetium-99 scans at each timepoint. The reviewer will identify the bone pixels, and determine the area (cm²) and disease status (bone lesion, not bone lesion) associated with each bone pixel [8].

The analysis timepoints for the analysis of quantitative bone scan endpoints are randomization baseline, Week 8, 16, 24, and every 12 weeks thereafter.

For analyses reported by visit, nominal visit per CRF will be used.

Patient bone scan lesion area (BSLA) is defined for each subject at each assessment as the sum of the pixel areas (cm²) of the set of the whole body quantified technetium-99 bone scan imaging pixels identified as bone lesion, as determined by the central imaging reviewer using the CAD system software.

Patient bone scan timepoint response rate (week 8, 16 and week 24) For each comparison, patient bone scan timepoint response with respect to the randomization baseline will be assessed according to the following criteria in [Table 6–2](#):

Table 6–2 Patient Bone Scan Timepoint Response Criteria Following Randomization

Responder (R)	30% or greater resolution of the BSLA compared to baseline
Stable Disease (SD)	Not meeting the criteria for R, PD, or UE
Progressive Disease (PD)	Two or more new areas of radiotracer uptake attributable to metastatic disease in regions of bone that had not previously shown radiotracer uptake or greater than 30% increase from baseline in bone scan lesion area in areas attributable to metastatic disease.
Unable to Evaluate (UE)	Assigned if bone scan results cannot be interpreted due to inconsistent image acquisition parameters compared to the reference scan, incomplete imaging, or other similar technical deficiencies.

Subjects with no randomization baseline assessment or a baseline BSLA value of 0 will be assigned ‘UE’ values for all timepoints for the comparison. The response rate is the proportion of subjects with a response at the timepoint among the subjects evaluable for bone scan response at the randomization baseline.

Patient bone scan best overall response rate is defined for each comparison as the proportion of subjects having the best bone scan timepoint overall response as R, SD, PD, or UE among subjects with a non-zero BSLA measurement at the randomization baseline. Overall response is judged on all responses observed during the post-baseline period, including unscheduled visits. Responder and SD responses will only be reported if observed prior to PD. No confirmation of R or PD status is required.

Total ALP and PSA Endpoints

Total ALP and PSA endpoints are determined based on the total ALP and PSA laboratory assessments collected as described in the schedule of procedures.

Total ALP response rate is defined as the proportion of subjects with a $\geq 30\%$ reduction of the blood total-ALP level compared to the randomization baseline, confirmed by a second consecutive ALP value 4 or more weeks later but within 9 weeks, among evaluable subjects.

PSA response rate is defined as the proportion of subjects with a $\geq 30\%$ reduction of the blood PSA level, compared to the randomization baseline value, confirmed by a second consecutive PSA value 4 or more weeks later but within 9 weeks, among evaluable subjects.

In addition, **percentage change from randomization baseline in Total ALP** at weeks 12 and 24 will be summarized for Comparison 1.

Time to total ALP progression is defined as the time (days) from randomization date to the date of first total ALP progression. A progression event is defined as $\geq 25\%$ increase from randomization baseline value, at least 12 weeks from baseline in subjects with no ALP decline from baseline; or $\geq 25\%$ increase above the nadir value, which is confirmed by a second consecutive value obtained 3 or more weeks later but within 9 weeks in subjects with an initial ALP decline from the baseline.

Time to total ALP progression without 12-week restriction is similar to the time to total ALP progression defined above except that 12-week rule is removed in subjects with no ALP decline from baseline.

See [Table 9–4](#) in Appendix [9.1.3](#) for detailed censoring rules.

Time to PSA progression is defined as the time (days) from randomization baseline to the date of first PSA progression. A progression event is defined as $\geq 25\%$ increase from randomization baseline value and an increase in absolute value of ≥ 2 ng/mL, at least 12 weeks from the applicable baseline in subjects with no PSA decline from baseline; or $\geq 25\%$ increase and an absolute increase of ≥ 2 ng/mL above the nadir value, which is confirmed by a second consecutive value obtained 3 or more weeks later but within 9 weeks in subjects with an initial PSA decline from the baseline.

Time to PSA progression without 12 week restriction is similar to the time to PSA progression defined above except that 12-week rule is removed in subjects with no PSA decline from baseline.

See

Table 9–4 in Appendix 9.1.3 for detailed censoring rules .

Analgesic Use and Related Endpoints

Analgesic use in this study will be captured via two 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.

Data from both sources will be used to identify strength of each analgesic used by each subject at baseline and post-baseline visits. The amount of medication a subject actually took cannot be calculated due to only prescribed information is collected on the eCRF. Hence, only drug name will be used in identifying non-, weak- or strong- opioids for analgesics related analyses.

Analgesic names will be coded into standard terms using the World Health Organization Drug Dictionary (WHO-DD), Version 2005 Q3.

Each standardized analgesic name will be classified by strength type as non-opioid, weak opioid, or strong opioid.

An increase in analgesic use following randomization baseline is defined in Table 6–3 below. If a subject is using more than one type of analgesic, the strongest type will be used in the analysis.

Table 6–3 Increase in Analgesic Use Algorithm

Baseline intake	Post baseline
No analgesics or a non-opioid	Initiation of any opioid
Weak opioid	Initiation of any strong opioid or adding an additional weak opioid
Strong opioid	Initiation of an additional strong opioid

Pain improvement rate without increase in analgesic use following baseline is defined at each post-baseline assessment time point as the proportion of subjects with a 30% and 2-point decrease in average WPS over 2 consecutive assessment periods conducted at least 4 weeks apart, without a corresponding increase in analgesic use over those 2 periods. The denominator is the number of subjects with a non-missing WPS of at least 4 at randomization.

Time to increase in analgesic use is defined as the time in days from randomization baseline until the first increase in analgesic use for subjects with baseline average WPS of ≤ 7 and non-missing baseline assessment of analgesic use. See Table 9–5 in Appendix 9.1.4 for detailed censoring rules.

Pain progression or increase in analgesic use following baseline is defined as the time in days from randomization baseline until the earlier of pain progression or a first increase in analgesic use for subjects with a non-missing baseline average WPS of ≤ 7 . See [Table 9–4](#) in [Appendix 9.1.3](#) for detailed censoring rules.

Change in analgesic use following baseline : A shift table will be presented to summarize the change in strength of analgesic use from randomization baseline to the strongest use at post-baseline. The strength type will include: no analgesics or non-opioid, weak opioid, and strong opioid.

NCCN-FACT FPSI-17 Endpoints

The NCCN-FACT FPSI-17 is a validated instrument that was developed to assess symptoms of prostate cancer, symptoms caused by the treatment of prostate cancer, and the health related quality of life of prostate cancer patients. [12] The instrument contains 17 items, each of which utilizes a Likert scale with 5 possible responses. Ten items reflect disease related physical symptoms of disease and the responses on the items are to be summed to calculate a disease related physical symptom subscale score (FPSI-DRS-P, Disease Related Symptoms – Physical), One item represents emotional symptoms of disease and the response to that item is used to calculate a disease related emotional symptom subscale score (FPSI-DRS-E, Disease Related Symptoms – Emotional). Four items represent treatment related symptoms and the responses to these items are summed to calculate a treatment side effect subscale score (FPSI-TSE, Treatment Side Effects). Finally, two items represent functional well-being and the responses to those items are summed to calculate a functional/well-being subscale score (FPSI-F/WB, Function/Well Being).

The NCCN-FACT FPSI-17 total score and each of its subscale scores, FPSI-DRS-P, FPSI-DRS-E, FPSI-TSE, and FPSI-F/WB, will be summarized by visit.

Time to increase in physical symptoms of disease based on FPSI-DRS-P will be evaluated based on the FPSI-DRS-P subscale score of the NCCN-FACT, as the time from treatment start to the first increase in physical symptoms of disease on-study. NCCN-FACT FPSI-17 was not collected prior to randomization, so baseline is the assessment collected prior to first dosing. An increase 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. See [Table 9–4](#) in [Appendix 9.1.3](#) for detailed censoring rules.

BPI Subscales

Each subject will fill out the complete 11-question BPI-SF using an ePRO device at the Treatment Day 1 visit and at each clinic visit thereafter through Week 48. For subjects in Arm A and Arm B, the modified complete BPI-SF will also be filled out on the day of telephone contacts scheduled for Weeks 28, 32, 40, and 44. The data will be collected for one week (6 days prior to the applicable visit and the morning of the visit) prior to each clinic visit or phone call follow-up.

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 sum of the four items on the pain intensity, and (2) **Function interference index**, using the sum of the seven pain interference items (at least 4 out of 7).

Each BPI-SF subscale, and the two indices for pain severity and function interference, will be summarized using descriptive statistics.

Weekly Average pain score (APS) is defined for each subject at baseline and at each post-baseline date, as the mean of the “average pain score” values in the last 24 hours from the preceding 7 days before each applicable visit or telephone contact per the BPI-SF.

Analysis of Exploratory Endpoints

Identical statistical methods for analyzing time-to-event secondary endpoints, as elucidated in section 6.2.2.2, will be applied to the analysis of time-to-event exploratory endpoints:

- Time to radiological bone progression
- Time to radiological soft-tissue progression
- Time to total alkaline phosphatase (ALP) progression
- Time to ALP progression without 12-week restriction
- Time to PSA progression
- Time to PSA progression without 12-week restriction
- Time to pain progression or increase in analgesic use following baseline
- Time to increase in analgesic use
- Time to increase in physical symptoms of disease based on FPSI-DRS-P

Descriptive statistics (n, nmiss, mean, std, median, interquartile range, range) will apply to other exploratory endpoints by timepoint:

- Bone scan lesion area
- Percent change in ALP from baseline
- NCCN-FACT FPSI-17 subscale scores and total score
- BPI-SF subscales and 2 indices (pain severity and function interference)

The following rates will be summarized using (n, %) with exact binomial 80% and 95% CIs using the method of Clopper and Pearson [7]

- Timepoint and Best overall bone scan response rate

- ALP response rate
- PSA response rate
- Pain improvement rate without increase in analgesic use
 - Plots with 80% and 95% confidence intervals at each timepoint will also be provided.

A shift table will be presented to summarize the change in the strength of analgesic use from randomization baseline to the strongest use at post-baseline.

Summary statistics on weekly worst pain score at baseline and at each post-baseline timepoint (n, nmiss, mean, std, median interquartile range, range) will be presented.

6.3 Pharmacokinetics / pharmacodynamics

Not applicable.

6.4 Safety

No formal statistical tests will be done for the safety endpoints.

A safety analysis will be performed at the same time as the final primary endpoint analysis. This analysis will include treatment-emergent safety events as of database closure for primary endpoint completion.

An updated analysis of safety events will be performed after the last followed patient completes active follow-up 2 years following the last dose of radium-223 dichloride treatment, which will be reported in a separate document from the Clinical Study report. In addition, safety events emerging during long-term follow-up conducted as part of this study will be reported in separate document(s) from the study Clinical Study Report.

6.4.1 Adverse events

All adverse events (AE) will be coded according to the latest Medical Dictionary for Regulatory Activities (MedDRA) Version as of database lock for the applicable analysis. The MedDRA version used will be documented in the CSR. The intensity of an AE will be documented using the NCI-CTCAE v4.03.

Adverse Events of Interest

A summary table and a listing will be provided for subjects who experienced following adverse events: pancytopenia and myelotoxicity including myelodysplastic syndrome, aplastic anemia and myelofibrosis. The events will be identified by medical reviewer based on verbatim recorded in database.

Treatment Period Analyses

The treatment period for this study, for purposes of safety analyses, extends from the initiation of study treatment until 30 days after the last administration of radium-223 dichloride.

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 CTCAE toxicity grade or became serious) during the treatment period.

Treatment-emergent AEs

All AEs starting or worsening within the treatment period will be considered TEAEs,

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 treatment-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 treatment related TEAEs
- any serious TEAEs,
- any serious treatment related TEAEs
- any CTCAE Grade ≥ 3 TEAEs
- any TEAEs leading to drug modification or discontinuation
- any TEAEs leading to death
- any AEs leading to drug modification or discontinuation

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. For post-standard treatment AE tables, only treatment-related AEs, serious treatment-related AEs will be reported. Subjects who received 6 or less doses from Arm A and C versus subjects received 7 or more doses in Arm C will also be summarized for evaluating potential cumulative toxicity of extended regimen.

The same summaries will be repeated for related TEAEs, serious TEAEs, serious related TEAEs, CTCAE Grade ≥ 3 TEAEs, TEAEs leading to drug modification or discontinuation, TEAEs leading to death and any AEs leading to drug modification or discontinuation.

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 subjects experiencing TEAEs will be presented.

The analysis of TEAEs will be reported by treatment group as treated (Arm A, Arm B, and Arm C).

Long-Term Follow-Up AEs

Long-term follow-up AEs are AEs arising during the long-term follow-up period. LTFU AEs will not be included in safety analysis summary tables for this study, but will be included in listings.

AE Listings

Data listings will be produced for all AE recorded in the study. Verbatim descriptions and coded terms will be listed for all AEs.

Serious adverse events (SAEs), deaths, AEs leading to discontinuation and AEs of interest will each have a separate listing.

6.4.2 Deaths

Deaths reported during the study period will be tabulated by treatment group.

6.4.3 Clinical Laboratory Data

The following laboratory parameters will be summarized:

- Complete blood count (CBC), including hematocrit, hemoglobin, red blood cell (RBC) count, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), white blood cell (WBC) count, platelet count, and differential blood count reported as absolute count or percentage for neutrophils, lymphocytes, monocytes, eosinophils, and basophils
- Chemistry panel, including serum creatinine, sodium, potassium, chloride, calcium, phosphorous, magnesium, total bilirubin, total ALP, AST, ALT, lactate dehydrogenase (LDH) (screening only), and albumin (screening only)
- Other parameters: PSA, serum testosterone

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

Hematological and biochemical laboratory toxicities assigned by the investigator that include clinical assessments are available in Adverse Events database and are summarized in Adverse Event tables.

Any additional specific handling of the CTCAE v4.03 toxicity grading assignments will be noted in the footnotes of the corresponding tables.

Laboratory parameters will be evaluated for subjects in the safety (SAF) population by treatment group (Arm A, B, C).

For each analysis, 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 change from baseline by group. Box plots will be generated for hematology parameters over all visits, by group, to investigate trends over time and outliers in the data. Graphs of hematological toxicity parameters (boxplots of mean and quartile values over time) will be produced. Similar graphs of chemistry and/or other lab parameters may also be produced.

In addition, for each analysis, change from baseline will be summarized in shift tables according to CTCAE 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 not be included in the by visit summary tables, however will be included in the overall CTCAE grade tables.

6.4.4 New Primary Malignancies

The new primary malignancies will be summarized overall and by cancer type (n, percent), and reported through listings.

6.5 Analysis of Other Endpoints

6.5.1 ECOG Performance Status

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

6.5.2 Quality of Life

Summaries of NCCN-FACT FPSI-17 questionnaire 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. Analyses of NCCN-FACT data are described in the sections on efficacy and exploratory endpoints.

6.5.3 Long Term Safety

Safety data arising during the long-term safety follow-up period in this study will be presented through listings. In the event subjects in this study transfer into a long term safety follow-up study for all or a portion of their long-term follow-up period, any data not collected in this study is outside the scope of this SAP.

6.6 Examination of Subgroups

Subgroup analyses will be conducted for the primary efficacy endpoint SSE-FS following randomization based on the ITT population, and SSE-FS following Week 24 baseline on the W24 population, and for secondary endpoint OS. Subgroup analyses will be provided by subgroup within each respective pooled treatment grouping, as described for the primary analysis for each comparison. Descriptive statistics and hazard ratio estimates with 80% and 95% CIs will be provided at least for the subgroups listed below, provided there is a sufficient number of events (at least 10 events) in total within the subgroup across the applicable pooled treatment groups. Hazard ratios by subgroup with 80% and 95% CIs will be plotted using forest plots.

- ECOG performance status at baseline (0 vs 1)
- Extent of Disease (number of bone lesions: <6, 6-20, >20 or superscan at baseline)
- Race (White, Asian and Other)
- Age group (<65, 65-74, 75-84, >=85 years)
- Baseline Total Body Weight and Ideal Body Weight normalized dose quartiles
- Randomization stratification factors:

- use of prior chemotherapy (≤ 1 regimen versus > 1 regimen)
- total ALP (< 220 U/L versus ≥ 220 U/L)
- average worst pain score (WPS) of the BPI-SF at screening assessment (WPS ≤ 4 versus WPS > 4).
- BMI
 - Low, non-low (< 18.5 and ≥ 18.5)
 - Non-high, high (< 30 and ≥ 30)
- Gleason Score at the time of diagnosis (< 8 , ≥ 8 , Missing)
- Baseline PSA values by median ($< \text{median}$, $\geq \text{median}$)
- Received prior anti-androgens at baseline: abiraterone or enzalutamide (Yes vs. No)

For time-to-event endpoints, Kaplan-Meier [5] estimates will be presented with median, 25th, and 75th percentile survival time and associated 80% and 95% Brookmeyer-Crowley [6] confidence intervals (CI), and number and percentage of censored observations.

Corresponding Kaplan-Meier curves will also be plotted. A Cox proportional hazards model will be fitted for each time-to-event secondary efficacy endpoint, with baseline BMI category coded as a binary covariate (low versus non-low).

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

7. Document history and changes in the planned statistical analysis

- SAP 21 February 2016 version 1.0
- SAP 17 May 2017 version 2.0

Table 7–1 The Changes Between SAP V1.0 and SAP V2.0

Endpoints/Definitions	SAP V1.0	SAP V2.0
Week 24 baseline	The Week 24 baseline assessment was defined as the assessment within the window (Day 127, Day 211) (Week 24 target – 14 through Day 28 target + 14) closest to Day 169 (Week 24 target date), selecting the earlier date in the event of a tie.	‘Week 24 baseline’ is defined as the last non-missing value on or before the date of the 6 th injection for Arm A and Arm C.
Time to radiologic progression / rPFS/Bone scan time point & overall response	All imaging-based endpoints will be based on week-24 baseline for Comparison 2.	All imaging endpoints will be based on randomization baseline for Comparison 2.
PSA progression	Prostate specific antigen progression is defined with respect to each applicable baseline, as a $\geq 25\%$ increase	Will use the definition consistent with compound standard: Prostate specific antigen progression is defined as $\geq 25\%$ increase from the baseline value and an

	above the applicable nadir (lowest baseline or post-baseline) value, and an increase in absolute value of ≥ 2 ng/mL above the nadir.	increase in absolute value of ≥ 2 ng/mL, at least 12 weeks from baseline in subjects with no PSA decline from baseline; or $\geq 25\%$ increase and an absolute increase of ≥ 2 ng/mL above the nadir value, which is confirmed by a second consecutive value obtained 3 or more weeks later but within 9 weeks in subjects with an initial PSA decline from baseline.
Total ALP progression	Total ALP progression is defined with respect to each applicable baseline, as a $\geq 25\%$ increase above the applicable nadir (lowest respective baseline or post-baseline) value.	Will use the definition consistent with the compound standard: Alkaline phosphatase progression is defined as $\geq 25\%$ increase from the baseline value, at least 12 weeks from baseline in subjects with no ALP decline from baseline; or $\geq 25\%$ increase above the nadir value, which is confirmed by a second consecutive value obtained 3 or more weeks later but within 9 weeks in subjects with an initial ALP decline from baseline.
ALP response	Total ALP response is defined with respect to each applicable baseline, as a $\geq 30\%$ reduction of the blood total-ALP level compared to the baseline value.	Will use the definition consistent with the compound standard: Alkaline phosphatase response is defined as $\geq 30\%$ reduction of the blood level, compared to the baseline value, confirmed by a second consecutive ALP value 4 or more weeks later but within 9 weeks.
Increase in analgesic use algorithm	Use AQA score in the algorithm	Will be based on strength and type of analgesic use. No AQA score will be calculated.
Exploratory endpoint analyses	Both randomization baseline and week 24 baseline will be used in the analyses	Only randomization baseline will be used. Endpoints with respect to Week 24 baseline will not be derived as formal statistical comparisons will not be made.

8. References

1. Fine, D. Consequences of delayed treatment effects on analysis of time-to-event endpoints. *Drug Inform J.* 41:535–539 (2007)
2. Zhang, D., and Quan, H. Power and sample size calculation for log-rank test with a time lag in treatment effect. *Stat Med.* 28: 864–879 (2009)
3. Green B. and Duffull S. What is the best size descriptor to use for pharmacokinetic studies in the obese? *J Clin Pharmacol.* Aug 2004; 58(2): 119–133.
4. World Health Organization. “Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation.” *WHO Technical Report Series* 894. Geneva: WHO; 2000. 8-9.



5. Kaplan E and Meier P. Nonparametric estimation from incomplete observations. *J Amer Statist Assn* 53:457–481 (1958)
6. Brookmeyer R and Crowley J. A confidence interval for the median survival time. *Biometrics* 38:29-41 (1982)
7. Clopper C and Pearson E. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26:404–413. (1934)
8. Brown M, et al. Computer-aided quantitative bone scan assessment of prostate cancer treatment response, *Nucl Med Commun.* 33:384-94 (2012)
9. National Comprehensive Cancer Network. *NCCN Clinical Practice Guidelines in Oncology: Adult Cancer Pain.* Version 2.2014.
10. Butler, SF, Budman, SH, Fernandez, KC, et al. Development and validation of the current opioid misuse measure. *Pain* 142:144-156 (2007)
11. Pereira, J, Lawlor, P, Vigano, A, et al. Equianalgesic dose ratios for opioids: a critical review and proposals for long-term dosing *J Pain and Symptom Manage* 22:672-687 (2001).
12. <http://www.facit.org/FACITOrg/Questionnaires>.

9. Appendices

9.1 Censoring Rules

9.1.1 Censoring rules for Symptomatic Skeletal Event-Free Survival (SSE-FS)

Comparison 1:

- Arm C subjects who had SSE assessments or SSE event or death (in the absence of an SSE event) after ‘pooling cut-off date’ will be censored on last SSE assessment date prior to ‘pooling cut-off date’. Refer to Section 4.5 for pooling cut-off date.
- All other subjects in arms A, B and C, refer to [Table 9–1](#) for censoring rules.
- Symptomatic Skeletal Event Free Survival (SSE-FS) following randomization = End Date – Date of Randomization +1

Table 9–1 Comparison 1: Censoring for Symptomatic Skeletal Event Free Survival (SSE-FS)

Situation	End Date	Censored	Reason for Censoring
No post-randomization SSE assessment and no death	Date of Randomization	Yes	No post-baseline SSE assessments and no death
Subject had a 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 by database cutoff date	Last SSE assessment prior to database cutoff	Yes	Neither SSE nor death

*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-randomization SSE assessment.

Note: symptomatic skeletal events immediately after missing SSE assessments are still counted as events in the analysis of SSE-FS.

Comparison 2:

- All subjects eligible for Comparison 2, refer to [Table 9–2](#) for censoring rules.
- Symptomatic Skeletal Event Free Survival (SSE-FS) following Week 24 baseline= End Date – Date of Week 24 baseline(the 6th dose date) +1

Table 9–2 Comparison 2: Censoring for Symptomatic Skeletal Event Free Survival (SSE-FS)

Situation	End Date	Censored	Reason for Censoring
No SSE assessment or death following Week 24 baseline	Date of Week 24 baseline	Yes	No post-Week 24 baseline SSE assessments and no

			death.
Had an SSE event between date of randomization and Week 24 baseline	Date of Week 24 baseline	Yes	First event occurred between date of randomization and Week 24 baseline
Subject had a SSE event after Week 24 baseline	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 by database cutoff date	Last SSE assessment prior to database cutoff	Yes	Neither SSE nor death

*The earliest end date in the table is used in calculating the SSE-FS.

#: use Week 24 baseline date instead of last SSE assessment date if no post-Week 24 baseline SSE assessment.

Note: symptomatic skeletal events immediately after missing SSE assessments are still counted as events in the analysis of SSE-FS.

9.1.2 Censoring rules for Overall Survival (OS)

Comparison 1:

- Arm C subjects who have died or are known to be alive after ‘pooling cut-off date’ will be censored on the ‘pooling cut-off’ date. Refer to Section 4.5 for pooling cut-off date.
- All other subjects in arms A, B and C, refer to [Table 9–3](#) for censoring rules.
- Overall Survival (OS) following randomization = End Date – Date of Randomization +1

Comparison 2

- For Comparison 2 eligible subjects, refer to [Table 9-3](#) for censoring rules.
- Overall Survival (OS) following Week 24 baseline = End Date – Date of Week 24 baseline(the 6th dose date) +1

Table 9–3 Overall Survival censoring rules

Situation	End Date	Censored	Reason for Censoring
Death on or prior to database cut off date	Date of Death	No	N/A
Not known to have died as of database cut off date	Last known alive date (LKAD) on or prior to database cutoff	Yes	No Death
Lost to follow-up prior to database cutoff	Last known alive date (LKAD)	Yes	Lost to follow up

9.1.3 Censoring rules for other time-to-event endpoints

Comparison 1:

- Arm C subjects who had relevant event assessment or event after ‘pooling cut-off date’ will be censored on last assessment date prior to ‘pooling cut-off date’. Refer to Section 4.5 for pooling cut-off date.
- All other subjects in arms A, B and C, refer to
- [Table 9–4](#) for censoring rules.
- Randomization baseline is used for this comparison
- Time-to-event following randomization = End date - date of randomization +1

Comparison 2:

- Subjects who had a relevant event between date of randomization and Week 24 baseline will be censored at date of Week 24 baseline, with the reason ‘First event occurred between date of randomization and Week 24 baseline’.
- All other subjects refer to
- [Table 9–4](#).
- Week 24 baseline is used for this comparison for following endpoints: Time to first SSE and Time to Pain progression
- Randomization baseline is used for this comparison for exploratory endpoints :Time to ALP progression, time to PSA progression, time to increase in analgesic use and time to pain progression or increase in analgesic use.
- The assessment prior to treatment start is used as baseline for time to increase in physical symptoms of disease.
- Time-to-event following Week 24 baseline = End date - date of week 24 baseline +1.

[Table 9–4](#) applies to the following endpoints:

- **Time to First Symptomatic Skeletal Event**
- **Time to Pain Progression**
- **Time to ALP Progression**
- **Time to ALP Progression without 12-week Restriction**
- **Time to PSA Progression**
- **Time to PSA Progression without 12-week Restriction**
- **Time to Increase in Analgesic Use**
 - Requires Baseline WPS ≤ 7 for inclusion in analysis
- **Time to Pain progression or increase in analgesic use**
 - Requires Baseline WPS ≤ 7 for inclusion in analysis
 - The last assessment for censoring purposes refers to the latest date Pain Progression assessment. The earlier of the date of pain progression or increase in analgesic use is taken as the endpoint.
- **Time to increase in physical symptoms of disease following start of treatment based on FPSI-DRS-P**

Table 9–4 Time-to Event Censoring Rules

Situation	End Date	Censored	Reason for Censoring
No baseline assessment	Relevant Baseline Date	Yes	No baseline assessment.
No post baseline assessment	Relevant Baseline Date	Yes	No post-baseline assessment.
Subject had an event	Date of first event	No	N/A
No event by database cutoff	Last assessment prior to	Yes	No Event

	database cutoff		
--	-----------------	--	--

9.1.4 Censoring rules for time to radiological progression

Note: ‘radiological’ in the tables refer to either ‘radiological soft tissue’ or ‘radiological bone’ as applicable to the endpoint.

Refer to Section 4.5.2 for pooling cut-off date.

Table 9–5 applies to:

- **Time to radiological progression following randomization**
- **Time to radiological bone progression following randomization and**
- **Time to radiological soft tissue progression following randomization.**

The following censoring rules will be used for both Comparison 1 and Comparison 2. Randomization baseline is used for all comparisons.

Table 9–5 Censoring for Time to Radiological Bone/Soft Tissue Progression following randomization

Situation	End Date	Censored	Reason for Censoring
No randomization baseline or post-randomization baseline radiological assessment	Date of randomization	Yes	No baseline or post-baseline assessment.
Subject died without prior radiological progression	Date of last radiological assessment before death	Yes	No progression before death or database cutoff
Subject survived without radiological progression as of database cutoff date	Date of last radiological assessment before data cutoff	Yes	Alive and no progression before database cutoff
Subject had a radiological assessment of PD (no two consecutive missed radiological assessments*)	Date of first PD	No	N/A
PD immediately after two or more consecutive missed radiological assessments*	Date of last radiological assessment before missed assessments	Yes	Missed two or more consecutive tumor assessments
Subject discontinuation from study for reasons other than PD	Last radiological assessment date	Yes	Subject discontinued from study due to a reason other than PD
Subjects discontinued from study due to PD, but no documented date of PD	Date of last radiological assessment	Yes	Subjects discontinued from study due to PD, but no documented date of PD
New non-protocol permitted systemic anti-cancer treatment started	Date of last radiological assessment before starting new systemic anti-cancer treatment	Yes	New anti-cancer treatment started

*Two consecutive missing radiological assessments are defined as the time interval between two consecutive radiological assessments is more than 28 weeks = $2 \times (12+2)$ week, where 12-week is scheduled frequency for tumor assessment and 2-week is protocol-allowed window.

Duration = End date – randomization date +1

There will be no Week 24 baseline censoring for Comparison 2.

In addition to the rules in [Table 9–5](#) the following rule will be applied to Comparison 1:

Situation	End Date	Censored	Reason for Censoring
Arm C subject with no radiological progression event on or after pooling cutoff date	Last radiological progression assessment on or before pooling cutoff date	Yes	Post Week 24 Arm C events censored at pooling cutoff date

When determining time to radiographic progression, the following rules will be applied by taking into consideration of both bone and non-bone disease status:

Table 9–6 Censoring for Time to Radiological Progression Following Randomization

Situation		Event/Censor	Event/censor date	Combined Non-bone and Bone outcome
1	Bone	Event (PD)	Bone scan date	Event at min(Bone scan date, Non-Bone assessment date)Event at min(Bone scan date, Non-Bone assessment)
	Non-Bone	Event (PD)	Non-Bone assessment date	
2	Bone	Censor	Bone scan date	Censor at min(Non-Bone assessment date, Bone scan date)
	Non-Bone	Censor	Non-Bone assessment date	
3	Bone	Censor	Bone scan date	Event at Non-Bone assessment date: if Non-Bone assessment date \leq Bone scan date ; Else if Non-Bone assessment date $>$ Bone scan date AND Difference $<$ 28weeks; Else censor at bone scan Date; Event at Non-Bone assessment Date: if Non-Bone assessment Date \leq Bone scan date Else if Non-Bone assessment Date $>$ Bone scan date AND Difference $<$ 28weeks Else censor at bscan Date
	Non-Bone	Event(PD)	Non-Bone assessment date	
4	Bone	Event(PD)	Bone scan date	Event at bone scan date: if Bone scan date \leq Non-Bone assessment date Else if Bone scan date $>$ Non-Bone assessment date AND Difference $<$ 28weeks; Else censor at Non-Bone assessment date
	Non-Bone	Censor	Non-Bone assessment date	

9.1.5 Censoring rules for radiological progression-free survival (rPFS)

The following censoring rules will be used for both Comparison 1 and Comparison 2.

Table 9–7 Censoring for radiological progression-free survival following randomization

Situation	End Date	Censored	Reason for Censoring
No randomization baseline or post-randomization baseline radiological progression assessments	Date of Randomization	Yes	No baseline or post-baseline assessment.
Subject had a radiological assessment of PD following randomization (no 2 consecutive missing radiological assessments*)	Date of first radiological event	No	N/A
Death or PD <u>immediately</u> after two or more consecutive missed radiological assessments*	Date of last radiological assessment before missed assessments	Yes	Missed two or more consecutive tumor assessments
Subject discontinued from study for other than PD or death	Last radiological assessment date	Yes	Subject discontinued from study due to a reason other than PD or death
Death during the study (no 2 consecutive missing radiological assessments*) or before first radiological PD assessment	Date of Death	No	N/A
Subjects discontinued from study due to PD, but no documented date of PD	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
New non-protocol permitted systemic anti-cancer treatment started	Date of last radiological assessment before starting new systemic anti-cancer treatment	Yes	New systemic anti-cancer treatment started

*Two consecutive missing radiological assessments are defined as the time interval between two consecutive radiological assessments is more than 28 weeks = $2 \times (12+2)$ week, where 12-week is scheduled frequency for tumor assessment and 2-week is protocol-allowed window

Radiological progression date and status mentioned in above table will be obtained following the rules specified in [Table 9–6](#).

rPFS following randomization = End date – randomization date +1.

There will be no Week 24 baseline censoring for Comparison 2. Randomization baseline will be used for Comparison 2.

In addition to the rules in [Table 9–7](#), the following rule will be applied to Comparison 1:



Situation	End Date	Censored	Reason for Censoring
Arm C subject with no death or radiological progression event on or after pooling cutoff date	Last radiological progression assessment on or before pooling cutoff date	Yes	Post Week 24 Arm C events censored at pooling cutoff date



Title page

A three arm randomized, open-label Phase II study of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) versus 80 kBq/kg (88 kBq/kg after implementation of NIST update) , and versus 50 kBq/kg (55 kBq/kg after implementation of NIST update) in an extended dosing schedule in subjects with castration-resistant prostate cancer metastatic to the bone

[Standard dose versus high dose and versus extended standard dose radium-223 dichloride in castration-resistant prostate cancer metastatic to the bone]

Bayer study drug BAY 88-8223 / Radium-223 dichloride / Xofigo

[Study purpose:] Efficacy, dose-finding

Clinical study phase: II **Date:** 02 July 2018

Study No.: BAY 88-8223 / 16507 **Version:** 1.0

Author: PPD

Confidential
The information provided in this document is strictly confidential and is intended solely for the guidance of the clinical investigation. Reproduction or disclosure of this document, whether in part or in full, to parties not associated with the clinical investigation or its use for any other purpose without the prior written consent of the sponsor is not permitted.

Throughout this document, symbols indicating proprietary names (®, TM) are not displayed. Hence, the appearance of product names without these symbols does not imply that these names are not protected.

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.
The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Title page.....	1
Abbreviations.....	2
1. Introduction.....	2
2. Study Objectives.....	2
3. Study Design.....	4
4. General Statistical Considerations.....	4
5. Analysis Sets.....	4
6. Statistical Methodology.....	4
6.1 Population characteristics.....	4
6.1.1 Additional Summary of Baseline Characteristics.....	4
6.1.2 Additional Summary of Prior/Concomitant Bone Health Agent (BHA) use.....	4
6.2 Efficacy.....	4
6.3 Exploratory variables and analyses.....	5
6.4 Pharmacokinetics.....	5
6.5 Safety.....	5
6.5.1 Adverse Events.....	5
7. Document history and changes in the planned statistical analysis.....	5
8. References.....	5

Abbreviations

AE	Adverse event
BHA	Bone health agent
CI	Confidence interval
CSR	Clinical Study Report
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 submission purposes or CSR.

This Supplemental SAP version 1.0 is a supplement of SAP version 2.0 dated 17 May 2017.

2. Study Objectives

Co-primary objectives:

Statistical Analysis Plan

Protocol No.: < BAY 88-8223/16507 >

Page: 3 of 5:

- To evaluate efficacy as measured by SSE-FS of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 doses compared to radium-223 dichloride 80 kBq/kg (88 kBq/kg after implementation of NIST update); and
- To evaluate efficacy as measured by SSE-FS of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 additional doses compared to no further radium-223 dichloride treatment in subjects with CRPC metastatic to the bone who previously received radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 doses, and survived SSE free and are eligible for further radium-223 dichloride treatment.

Secondary objectives:

- To evaluate safety and tolerability
- To evaluate overall survival (OS)
- To evaluate pain improvement rate
- To evaluate time to pain progression
- To evaluate time to first SSE
- To evaluate time to radiological progression
- To evaluate radiological progression-free survival (rPFS)

Exploratory objectives:

- To evaluate quantitated whole body technetium-99 bone scan tumor burden area and index, and determine bone tumor response
- To explore the impact of patient body size on the efficacy and safety of radium-223 dichloride
- Time to increase in physical symptoms of disease based on the NCCN FACT-P Symptom Index-17 (NCCN-FACT FPSI-17) physical disease related symptoms (FPSI-DRS-P) subscale score measured up to Week 48 after the start of treatment
- To evaluate laboratory indicators of efficacy, including:
 - PSA response
 - Time to PSA progression
 - ALP response
 - Time to ALP progression
 - Percentage change in ALP from baseline
- To evaluate change in analgesic use.

3. Study Design

Refer to the main SAP v2.0 dated 17 MAY 2017.

4. General Statistical Considerations

Refer to the main SAP v2.0 dated 17 MAY 2017.

5. Analysis Sets

Refer to the main SAP v2.0 dated 17 MAY 2017.

6. Statistical Methodology

Refer to the main SAP v2.0 dated 17 MAY 2017.

6.1 Population characteristics

Refer to the main SAP v2.0 dated 17 MAY 2017.

6.1.1 Additional Summary of Baseline Characteristics

The following baseline cancer characteristics will be summarized for ITT and week 24 analysis set:

- The summary of patients with vertebral bone metastasis at baseline
- The number of subjects with prior local radiotherapy on prostate and prostate gland by intent.

6.1.2 Additional Summary of Prior/Concomitant Bone Health Agent (BHA) use

The number of subjects who used BHA at baseline and during treatment period will be summarized for ITT and week 24 analysis set.

6.2 Efficacy

Refer to the main SAP v2.0 dated 17 MAY 2017.

Pain improvement rate at week 12, week 24, EOT, and overall will be summarized by including all subjects in ITT population as denominator. Same analyses will be repeated for week 24 analysis set.

6.3 Exploratory variables and analyses

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

6.4 Pharmacokinetics

Not Applicable.

6.5 Safety

Refer to the main SAP v2.0 dated 17 MAY 2017.

6.5.1 Adverse Events

Additional analyses of Adverse Events of 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 the last visit date during treatment period. Time to fracture will be summarized using Kaplan-Meier estimates. Median survival time together with the 25th and 75th percentiles and associated 80% and 95% Brookmeyer-Crowley confidence intervals (CI) will be presented by treatment arm. Corresponding Kaplan-Meier curves will be generated by treatment arm.

Time to fracture will also be summarized for subjects with/without use of concomitant anti-androgen, enzalutamide and abiraterone during the treatment period using same approach mentioned above, respectively.

A subject listing of fractures will be provided.

7. Document history and changes in the planned statistical analysis

- Main SAP v2.0 dated 17 MAY 2017.

8. References

Refer to the main SAP v2.0 dated 17 MAY 2017.



Title page

A three arm randomized, open-label Phase II study of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) versus 80 kBq/kg (88 kBq/kg after implementation of NIST update) , and versus 50 kBq/kg (55 kBq/kg after implementation of NIST update) in an extended dosing schedule in subjects with castration-resistant prostate cancer metastatic to the bone

[Standard dose versus high dose and versus extended standard dose radium-223 dichloride in castration-resistant prostate cancer metastatic to the bone]

Bayer study drug BAY 88-8223 / Radium-223 dichloride / Xofigo

[Study purpose:] Efficacy, dose-finding

Clinical study phase: II **Date:** 08 Oct 2018

Study No.: BAY 88-8223 / **Version:** 2.0
16507

Author: PPD

Confidential

The information provided in this document is strictly confidential and is intended solely for the guidance of the clinical investigation. Reproduction or disclosure of this document, whether in part or in full, to parties not associated with the clinical investigation or its use for any other purpose without the prior written consent of the sponsor is not permitted.

Throughout this document, symbols indicating proprietary names (®, TM) are not displayed. Hence, the appearance of product names without these symbols does not imply that these names are not protected.

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.
The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Title page.....	1
Abbreviations.....	2
1. Introduction.....	2
2. Study Objectives.....	3
3. Study Design.....	4
4. General Statistical Considerations.....	4
5. Analysis Sets.....	4
6. Statistical Methodology.....	4
6.1 Population characteristics.....	4
6.1.1 Additional Summary of Disposition of Subjects.....	4
6.1.2 Additional Summary of Baseline Characteristics.....	4
6.1.3 Additional Summary of Prior/Concomitant Bone Health Agent (BHA) use.....	5
6.1.4 Additional Summary of Post-treatment Bone Radiotherapy Use.....	5
6.2 Efficacy.....	5
6.3 Exploratory variables and analyses.....	5
6.4 Pharmacokinetics.....	5
6.5 Safety.....	5
6.5.1 Adverse Events.....	5
6.5.2 Long Term Adverse Events.....	6
6.5.3 Long Term Safety.....	6
7. Document history and changes in the planned statistical analysis.....	6
8. References.....	6

Abbreviations

AE	Adverse event
BHA	Bone health agent
CI	Confidence interval
CSR	Clinical Study Report
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 submission purposes or CSR.

This Supplemental SAP version 1.0 is a supplement of SAP version 2.0 dated 17 May 2017.

2. Study Objectives

Co-primary objectives:

- To evaluate efficacy as measured by SSE-FS of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 doses compared to radium-223 dichloride 80 kBq/kg (88 kBq/kg after implementation of NIST update); and
- To evaluate efficacy as measured by SSE-FS of radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 additional doses compared to no further radium-223 dichloride treatment in subjects with CRPC metastatic to the bone who previously received radium-223 dichloride 50 kBq/kg (55 kBq/kg after implementation of NIST update) every 28 days for up to 6 doses, and survived SSE free and are eligible for further radium-223 dichloride treatment.

Secondary objectives:

- To evaluate safety and tolerability
- To evaluate overall survival (OS)
- To evaluate pain improvement rate
- To evaluate time to pain progression
- To evaluate time to first SSE
- To evaluate time to radiological progression
- To evaluate radiological progression-free survival (rPFS)

Exploratory objectives:

- To evaluate quantitated whole body technetium-99 bone scan tumor burden area and index, and determine bone tumor response
- To explore the impact of patient body size on the efficacy and safety of radium-223 dichloride
- Time to increase in physical symptoms of disease based on the NCCN FACT-P Symptom Index-17 (NCCN-FACT FPSI-17) physical disease related symptoms (FPSI-DRS-P) subscale score measured up to Week 48 after the start of treatment
- To evaluate laboratory indicators of efficacy, including:
 - PSA response
 - Time to PSA progression
 - ALP response
 - Time to ALP progression
 - Percentage change in ALP from baseline

- To evaluate change in analgesic use.

3. Study Design

Refer to the main SAP v2.0 dated 17 MAY 2017.

4. General Statistical Considerations

Refer to the main SAP v2.0 dated 17 MAY 2017.

5. Analysis Sets

Refer to the main SAP v2.0 dated 17 MAY 2017.

6. Statistical Methodology

Refer to the main SAP v2.0 dated 17 MAY 2017.

6.1 Population characteristics

Refer to the main SAP v2.0 dated 17 MAY 2017.

6.1.1 Additional Summary of Disposition of Subjects

The number and percentage of subjects entered long-term follow up, completed and discontinued and reason for discontinuation will be presented by treatment group and overall for both ITT and safety analysis set.

6.1.2 Additional Summary of Baseline Characteristics

The following baseline cancer characteristics will be summarized for ITT and week 24 analysis set:

- The summary of patients with vertebral bone metastasis at baseline
- The number of subjects with prior local radiotherapy on prostate and prostate gland by intent.

The following summary tables will be provided for subjects who entered long-term follow up in safety analysis set.

- Demographics
- Baseline prostate cancer characteristics

6.1.3 Additional Summary of Prior/Concomitant Bone Health Agent (BHA) use

The number of subjects who used BHA at baseline and during treatment period will be summarized for ITT and week 24 analysis set. The same summary tables will also be provided for subjects who entered long-term follow up period in safety analysis set.

6.1.4 Additional Summary of Post-treatment Bone Radiotherapy Use

The number of subjects who received bone radiotherapy after last dose of Radium-223 will be summarized for subjects who entered long-term follow up period in safety analysis set.

6.2 Efficacy

Refer to the main SAP v2.0 dated 17 MAY 2017.

Pain improvement rate at week 12, week 24, EOT, and overall will be summarized by including all subjects in ITT population as denominator. Same analyses will be repeated for week 24 analysis set.

6.3 Exploratory variables and analyses

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

6.4 Pharmacokinetics

Not Applicable.

6.5 Safety

Refer to the main SAP v2.0 dated 17 MAY 2017.

6.5.1 Adverse Events

Additional analyses of Adverse Events of 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 either end of treatment (EOT) visit date or the last visit date before the start of active follow-up Period if no EOT visit was recorded. Time to fracture will be summarized using Kaplan-Meier estimates. Median survival time together with the 25th and 75th percentiles and associated 80% and 95% Brookmeyer-Crowley confidence intervals (CI) will be presented by treatment arm. Corresponding Kaplan-Meier curves will be generated by treatment arm.

Time to fracture will also be summarized for subjects with/without BHA use at baseline, with/without concomitant use of anti-androgen, enzalutamide, and abiraterone by same approach mentioned above, respectively.

A subject listing of bone fractures occurred after first dose will be provided.

6.5.2 Long Term Adverse Events

All Bone fracture and bone events occurred after patients entering long-term follow up period will be summarized by NCI-CTCAE toxicity criteria. Same summary will also be provided for serious bone fracture and bone events during long-term follow up period. A listing of bone fracture and bone events will be provided for entire study period and during follow-up period, respectively.

6.5.3 Long Term Safety

All Safety data arising during the long-term safety follow-up period in this study will be presented through listings.

7. Document history and changes in the planned statistical analysis

- Main SAP v2.0 dated 17 MAY 2017.
- SAP supplement v1.0 dated 02JUN 2018

8. References

Refer to the main SAP v2.0 dated 17 MAY 2017.