NOVARTIS

Clinical Development

AAA617

Clinical Trial Protocol CAAA617B12302 / NCT04689828

PSMAfore: A phase III, Open-label, Multi-Center, Randomized Study Comparing 177Lu-PSMA-617 vs. a Change of androgen receptor-directed therapy in the Treatment of Taxane Naïve Men with Progressive Metastatic Castrate Resistant Prostate Cancer

Statistical Analysis Plan (SAP)

Document type:	SAP Documentation
Document status:	Amendment v5.0
Release date:	17-May-2024
Number of pages:	86

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
07-Jul-2021	Prior to DB lock	Creation of final version	N/A - First version	NA
18-Oct-2022	Prior to DB lock	Creation of version 1.0	Clarification of definitions including on-treatment period, addition of subgroup, supplementary, sensitivity analyses, and analyses addressing the impact of production halt Leaning the SAP to have a focused CSR by removing redundant outputs	Multiple existing sections and a new section 2.15
17-Mar-2023	Prior to Second OS IA	Creation of version 2.0	This SAP amendment aims to add an additional interim analysis for OS at ~9 months after the cut-off date for the primary analysis for rPFS. This will: · bolster the median follow-up time for efficacy and safety as a high number of patients were enrolled shortly before the primary analysis cut-off date, resulting in short median follow-up; · allow all patients randomized to the 177Lu-PSMA-617 arm to complete the full course of treatment; · provide more mature OS data based on a higher information fraction compared to the one observed at the time of the rPFS primary analysis. Adding this additional OS interim results in a 4-look design where type-I error for OS is controlled with pre-specified alpha spending function. Some minor clarifications of definitions were made for duration of response and time to response calculations. [redacted]	Multiple sections including section 1.2, 2.2,2.5,2.6
21-Apr-2023	Prior to Second OS IA, [redacted]	Creation of version 3.0	[redacted] to clarify that the second interim OS analysis will be event-driven; the analysis will be conducted after approximately 125 OS events (~42% information fraction) have been reported.	Section 2.5.2, [redacted]

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
26-Feb-2024	Prior to Third OS IA	Creation of version 4.0	This amendment aims to provide clarification on derivations of subgroups and covariates, to add table summaries by crossover status and additional information on new antineoplastic therapies. It also provides clarification on planned OS analyses, and sensitivity analyses for PSA50, TTCH, and QoL endpoints)	2.4.2, 2.5.4, 2.5.6, 2.6.3, 2.7.1, 2.7.2, 2.11.1, [redacted]
17-May-2024	[redacted] to make ITT the primary analysis method for OS.	Creation of version 5.0	[redacted] of making the ITT as primary analysis for OS (Section 1.2.2 and section 2.6). Additional changes include the following clarification of the Safety set definition, SSE data collection, hazard ratio for PFS2	1.2.2, 2.2, 2.6, 2.7.1, 2.7.2

Table of contents

List of tables

List of figures

List of abbreviations

AE	Adverse Event
ARDT	Androgen receptor-directed therapy
CRF	Case Report Form
CSR	Clinical Study Report
DMS	Document Management System
FAS	Full Analysis Set
IA	Interim Analyses
MedDRA	Medical Dictionary for Drug Regulatory Affairs
PK	Pharmacokinetics
PPS	Per-Protocol Set
PRO	Patient-reported Outcomes
RAP	Reporting & Analysis Process
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
TFLs	Tables, Figures, Listings
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes all planned analyses for the Clinical Study Report (CSR) of study AAA617B12302, a phase III, Open-label, Multi-Center, Randomized Study Comparing [^{177}Lu]Lu-PSMA-617 vs. a Change of androgen receptor-directed therapy in the Treatment of Taxane Naïve Men with Progressive Metastatic Castrate Resistant Prostate Cancer.

The content of this SAP is based on protocol AAA617B12302 version 03, released 21-Feb-2023. All decisions regarding final analysis, as defined in the SAP document, have been made prior to database lock of the study data.

1.1 Study design

This is a randomized, Phase III , open-label, active-controlled, multi-center study comparing safety and efficacy of [^{177}Lu]Lu-PSMA-617 to a change in Androgen receptor-directed therapy (ARDT) in PSMA-positive mCRPC participants previously treated with an ARDT where it is considered appropriate to delay taxane-based chemotherapy. Approximately 450 patients will be randomized to one of the following treatment arms in 1:1 ratio:

- [^{177}Lu]Lu-PSMA-617
- ARDT

Randomization will be stratified by the following factors:

- prior ARDT use in castrate-resistant prostate cancer (CRPC) vs. hormone-sensitive prostate cancer (HSPC) setting
- symptomatology i.e. asymptomatic or mildly symptomatic (score of 0-3 on item 3 of the Brief Pain Inventory Short Form (BPI-SF) questionnaire) vs symptomatic (score >3 on item 3 of the BPI-SF questionnaire).

Radiographic Progression free survival (rPFS) as assessed by BICR and using PCWG3-modified RECIST 1.1 criteria is the primary endpoint in this study. rPFS assessed by local investigators review of tumor response is a secondary endpoint in this study. Overall survival (OS) is the key secondary endpoint.

No formal interim analysis is planned for rPFS. The primary rPFS analysis will only be carried out when approximately 156 events are observed.

A hierarchical testing procedure will be adopted and the statistical tests for OS will be performed only if the primary efficacy endpoint rPFS is statistically significant.

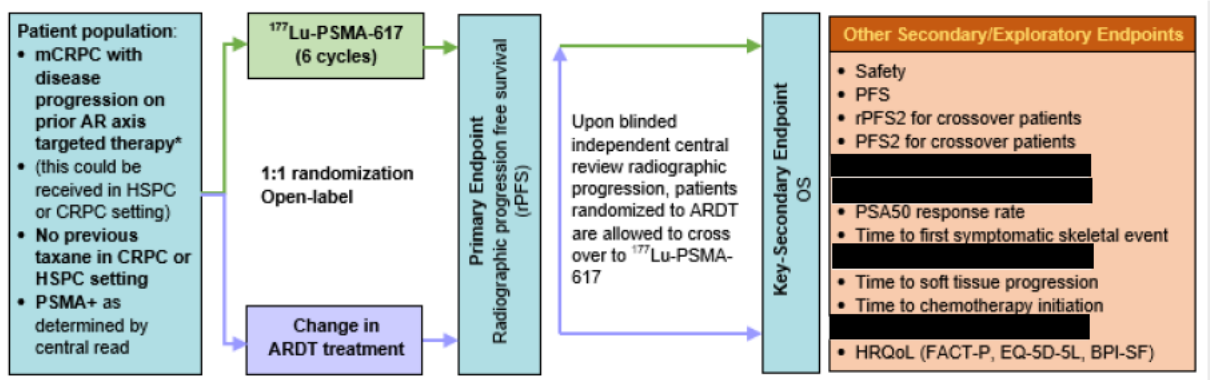
A maximum of three interim analyses for OS are planned. The first interim analysis was performed at the time of the primary rPFS analysis. The second interim analysis will be performed after approximately 125 of the 297 targeted OS events (42% information fraction) are observed. This interim analysis is expected to occur around 24 months from the date of the first participant randomized in the study. The third interim analysis is planned after approximately 223 of the approximately 297 targeted OS events (i.e., at approximately 75% information fraction) have been observed. This interim analysis is expected to occur around 38 months from the date of first participant randomized in the study. The primary intent of these

interim analyses is to stop early for superior efficacy. There is no intent to assess futility at these interim analyses. The first interim for OS will be carried out at the time of rPFS analysis.

An Independent Data Monitoring Committee (IDMC) will monitor unblinded safety data every 6 months during the conduct of the trial until the primary rPFS analysis is conducted.

Upon confirmation of rPD by BICR, participants randomized to the ARDT arm will be allowed to crossover to receive [^{177}Lu]Lu-PSMA-617. Figure 1-1 summarizes the overall study design.

Figure 1-1 Study Design



1.2 Study objectives, endpoints and estimands

Table 1-1 Objectives and related endpoints

Objective(s)	Endpoint(s)
Primary objective(s)	**Endpoint(s) for primary objective(s)**
• To evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the time to radiographic progression by PCWG3-modified RECIST v1.1 or death in participants with progressive PSMA-positive mCRPC compared to participants treated with ARDT	• Radiographic progression-free survival (rPFS) is defined as the time from the date of randomization to the date of first documented radiographic disease progression as assessed by blinded independent central review (BICR) and as outlined in Prostate Cancer Working Group 3 (PCWG3) Guidelines (Scher et al 2016) or death due to any cause.
Secondary objective(s)	**Endpoint(s) for secondary objective(s)**
• Key Secondary Objective: To evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the overall survival (OS) in participants with progressive PSMA-positive mCRPC compared to participants treated with ARDT treatment	• OS: Time from randomization to death due to any cause

Objective(s)	Endpoint(s)
• To estimate the time to radiographic progression or death in participants treated with ARDT who subsequently crossover to [^{177}Lu]Lu-PSMA-617 after radiographic progression (rPFS2)	• rPFS2 defined as time from the date of crossover (ARDT to ^{177}Lu-PSMA-617) to the date of radiographic disease progression by BICR or death from any cause [rPFS definition as outlined in PCWG3 guidelines]
• To evaluate Progression free survival (PFS) by investigator's assessment	• PFS defined as time from date of randomization to the first documented progression by investigator's assessment (radiographic, clinical, or PSA progression) or death from any cause, whichever occurs first
• To evaluate the second progression Free Survival (PFS2) by investigator's assessment	• PFS2 defined as time from date of randomization to the first documented progression by investigator's assessment (radiographic progression, clinical progression, PSA progression) or death from any cause, whichever occurs first, on next-line of therapy
• To evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the biochemical response as detected by Prostate specific antigen (PSA) halving compared to participants treated with ARDT	• PSA50 defined as proportion of participants who achieved a ≥ 50% decrease from baseline that is confirmed by a second PSA measurement ≥ 4 weeks. PSA50 will be evaluated at 3, 6 and 12 months.
• To evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the time to first symptomatic skeletal event (TTSE) compared to participants treated with ARDT	• Time to SSE (TTSSE) defined as date of randomization to the date of first new symptomatic pathological bone fracture, spinal cord compression, tumor-related orthopedic surgical intervention, requirement for radiation therapy to relieve bone pain or death from any cause, whichever occurs first
• To evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the time to radiographic soft tissue progression compared to participants treated with ARDT	• Time to soft tissue progression (TTSTP) defined as time from randomization to radiographic soft tissue progression per PCWG3-modified RECIST v1.1 (Soft Tissue Rules of Prostate Cancer Working Group modified Response Evaluation Criteria in Solid Tumors Version 1.1) as Assessed by Blinded Independent Central Review (BICR)
• To evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the time to chemotherapy compared to participants treated with ARDT	• Time to chemotherapy (TTCT) defined as time from randomization to initiation of the first subsequent chemotherapy or death, whichever occurs first
• To evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the health-related quality of life (HRQoL) compared to participants treated with ARDT	• HRQoL as assessed by EQ-5D-5L, FACT-P and BPI-SF
• To evaluate the safety and tolerability of [^{177}Lu]Lu-PSMA-617	• Frequency of adverse events, safety laboratory assessments and vital signs
Exploratory objective(s)	**Endpoint(s) for exploratory objective(s)**
[redacted]	[redacted]
[redacted]	[redacted]

Objective(s)	Endpoint(s)
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

1.2.1 Primary estimand(s)

Primary clinical question of interest: what is the effect of [^{177}Lu]Lu-PSMA-617 with best supportive care (BSC) versus change in ARDT with BSC with regard to time to radiographic progression or death in the treatment of taxane naïve men with Progressive PSMA-positive Metastatic Castrate Resistant Prostate Cancer as defined through inclusion/exclusion criteria, regardless of treatment discontinuations (TD) for any reasons and regardless of change in best supportive care or start of new antineoplastic therapy prior to rPFS event?

The justification for the primary estimand is that it will capture both the effect of the study drug and the effect of additional medications, mirroring the conditions in clinical practice. Further details can be found in Section 2.5.

The primary estimand is described by the following attributes:

1. Population: all randomized taxane naïve (taxanes in adjuvant or neoadjuvant setting is permitted if 12 months have elapsed since completion of taxane therapy) men with Progressive PSMA-positive Metastatic Castrate Resistant Prostate Cancer, treated with another ARDT as last treatment that are candidates for ARDT change and where it is considered appropriate to delay taxane-based chemotherapy as defined through inclusion/exclusion criteria. Further details about the population are provided in [Section 5 of the protocol].
2. Variable: rPFS defined as the time from date of randomization to the date of first documented radiographic progression-free survival as assessed by BICR and outlined in PCWG3 guidelines or death due to any cause. Further details on rPFS are provided in Section 2.5.1.
3. Treatment of interest: the investigational treatment is [^{177}Lu]Lu-PSMA-617 with best supportive care regardless of subsequent anti-neoplastic treatment. The control treatment

is ARDT with best supportive care regardless of subsequent anti-neoplastic treatment. Further details about the investigational treatment and control treatment are provided in [Section 6 of the protocol].

4. Handling of remaining intercurrent events (treatment policy):
   - Discontinuation of study treatment for any reason
   - Change in best supportive care
   - Start of anti-neoplastic therapy prior to radiographic progression

Details on how to handle intercurrent events are provided in Section 2.5.3.

5. Summary measure: rPFS hazard ratio [^{177}Lu]Lu-PSMA-617 with BSC versus ARDT with BSC) along with 95% confidence interval, estimated using a Cox proportional hazard model stratified by the randomization stratification factors. Further details on how the summary measure will be tested are provided in Section 2.5.2.

1.2.2 Secondary estimand

Key secondary clinical question of interest: what is the effect of [^{177}Lu]Lu-PSMA-617 with BSC versus change in ARDT with BSC with regard to overall survival in the treatment of taxane naïve men with Progressive PSMA-positive Metastatic Castrate Resistant Prostate Cancer as defined through inclusion/exclusion criteria, regardless of treatment discontinuations for any reasons, change in best supportive care, switching to AAA617 arm or start of new antineoplastic therapy?

The justification for the key secondary estimand is that it will capture both the effect of the study drug and the effect of additional medications while accounting for the impact of crossover. Further details can be found in Section 2.6.

The key secondary estimand is described by the following attributes:

1. Population: same as that of primary estimand
2. Variable: Overall Survival (OS) defined as the time from randomization to death due to any cause. Further details on OS provided in Section 2.6.1.
3. Treatment of interest: the investigational treatment is [^{177}Lu]Lu-PSMA-617 with best supportive care regardless of subsequent anti-neoplastic treatment and he control treatment was switch in ARPI regardless of subsequent anti-neoplastic treatment including crossing over to receive AAA617.
4. Handling of remaining intercurrent events (treatment policy):
   - Discontinuation of study treatment for any reason
   - Change in best supportive care
   - Start of new anti-neoplastic therapy
   - Crossover to [^{177}Lu]Lu-PSMA-617arm for participants in the ARDT arm

Details on how to handle the intercurrent events are provided in Section 2.6.3.

5. Summary measure: OS hazard ratio ([^{177}Lu]Lu-PSMA-617 with BSC versus ARDT with BSC) along with 95% confidence interval, estimated using a Cox proportional hazard

model stratified by the randomization stratification factors. Further details on how the summary measure will be tested are provided in Section 2.6.2.

2 Statistical methods

2.1 Data analysis general information

The primary analysis will be performed by Novartis. The primary analysis will be performed once study enrollment is complete and approximately 156 rPFS events occur. At this time, an interim efficacy analysis of OS will also be performed by Novartis. All subsequent interim and final OS analyses will also be performed by Novartis.

SAS version 9.4 or later will be used to perform all data analyses and to generate tables, figures and listings.

Data included in the analysis

There is one final analysis planned for the primary efficacy endpoint. Up to 3 interim and a final analysis may be performed for the key secondary endpoint. A unique cut-off date will be established after the targeted number of events for each of the planned interim and final analyses has been documented. For each of the analyses, all statistical analyses will be performed using all data collected in the database up to the data cut-off date. All data with an assessment date or event start date (e.g. vital sign assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.

All events with start date before or on the cut-off date and end date after the cut-off date will be reported as 'ongoing'. The same rule will be applied to events starting before or on the cut-off date and not having documented end date. This approach applies, in particular, to adverse event and concomitant medication reports. For these cases, the end date will not be imputed and therefore will not appear in the listings.

The analysis cutoff date for the final analysis of study data will be established when approximately 297 number of deaths have occurred. This will mark the end of the study.

General analysis conventions

Pooling of centers: Unless specified otherwise, data from all study centers will be pooled for the analysis. Due to expected small number of patients enrolled at centers, no center effect will be assessed.

Qualitative data (e.g., gender, race, etc.) will be summarized by means of contingency tables by treatment group; a missing category will be included as applicable. Percentages will be calculated using the number of patients in the relevant population or subgroup as the denominator.

Quantitative data (e.g., age, body weight, etc.) will be summarized by appropriate descriptive statistics (i.e. mean, standard deviation, median, minimum, and maximum) by treatment group.

2.1.1 General definitions

Investigational drug and study treatment

Investigational drug, will refer to the [^{177}Lu]Lu-PSMA-617 and [^{68}Ga]Ga-PSMA-11. Whereas, ***study treatment*** will refer to [^{177}Lu]Lu-PSMA-617 and control arm treatment (abiraterone or enzalutamide).

The term investigational treatment may also be referred to as ***study treatment*** which is used throughout this document.

Date of first administration of investigational drug

The date of first administration of investigational drug is defined as the first date when a non-zero dose of investigational drug is administered and recorded on the Dosage Administration Record (DAR) (e)CRF. The date of first administration of study drug will also be referred as start of investigational drug. There will be one start date for [^{177}Lu]Lu-PSMA-617 and, in most cases, one for [^{68}Ga]Ga-PSMA-11. In the event two doses of [^{68}Ga]Ga-PSMA-11 are administered for some patients at different timepoints prior to start of study treatment, there will be two start dates for [^{68}Ga]Ga-PSMA-11 unless the patient had two scans under the same subject ID.

Date of last administration of investigational drug

The date of last administration of investigational drug is defined as the last date when a non-zero dose of investigational drug is administered and recorded on DAR eCRF. The date of last administration of investigational drug will also be referred as end of investigational drug. There will be one last date for [^{177}Lu]Lu-PSMA-617 and, in most cases, one for [^{68}Ga]Ga-PSMA-11. In the event two doses of [^{68}Ga]Ga-PSMA-11 are administered for some patients at different timepoints prior to start of study treatment, there will be two last dates for [^{68}Ga]Ga-PSMA-11 unless the patient had two scans under the same subject ID.

Date of first administration of study treatment

The date of first administration of study treatment is derived as the first date when a non-zero dose of study treatment was administered as per the Dosage Administration CRF. The date of first administration of study treatment will also be referred as ***start of study treatment***.

The date of first administration of study treatment is the same as the date of first administration of investigational drug or control drug.

Date of last administration of study treatment

The date of last administration of study treatment is defined as the last date when a non-zero dose of study treatment was administered as per Dose Administration (e)CRF.

The date of last administration of study treatment is the same as the date of last administration of investigational drug or control drug.

Study day

The study day, describes the day of the event or assessment date, relative to the reference start date.

The study day is defined as:

- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date + 1 if event is on or after the reference start date;
- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date if event precedes the reference start date.

The reference start date for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption, pk etc.) is the start of study treatment, as well as for patient reported outcomes (PRO).

The reference start date for all other, non-safety assessments (i.e. survival time, disease progression, tumor response, and ECOG performance status) is the date of randomization.

The study day will be displayed in the data listings. If an event starts before the reference start date, the study day displayed on the listing will be negative.

Time unit

A year length is defined as 365.25 days. A month length is 30.4375 days (365.25/12). If duration is reported in months, duration in days will be divided by 30.4375. If duration is reported in years, duration in days will be divided by 365.25.

Baseline

For efficacy evaluations, the last non-missing assessment, including unscheduled assessments on or before the date of randomization is taken as “baseline” value or “baseline” assessment. In the context of baseline definition, the efficacy evaluations also include PRO and performance status.

For safety evaluations, the last available assessment on or before the date of start of study treatment is taken or ‘‘baseline” assessment. In case time of assessment and time of treatment start is captured (e.g. pre-dose ECG), the last available assessment before the treatment start date/time is used for baseline.

If patients have no value as defined above, the baseline result will be missing.

Note: For all analyses related [^{68}Ga]Ga-PSMA-11, baseline is defined as above but using [^{68}Ga]Ga-PSMA-11injection as a reference date for both safety and efficacy.

On-treatment assessment/event and observation periods

Randomized part of the protocol:

For adverse event reporting the overall observation period will be divided into three mutually exclusive segments:

1. ***pre-treatment period***: from day of patient's informed consent to the day before first administration of study treatment
2. ***on-treatment period***: from date of first administration of study treatment to 30 days after EOT or (last ^{177}Lu-PSMA-617 dose date + 41 days, last dose date of ARDT + 30 days), whichever is later, or the day before crossover start of treatment date, if applicable
3. ***post-treatment period***: starting at day 31 after EOT or (last ^{177}Lu-PSMA-617 dose date + 42 days, last dose date of ARDT + 31 days), whichever is later.

[^{68}Ga]-PSMA part of the protocol:

For adverse event reporting the overall observation period will be divided into three mutually exclusive segments:

1. ***pre-treatment period***: from day of patient's informed consent to the day before earliest [^{68}Ga]-PSMA administration.
2. ***on-treatment period***: from date earliest [^{68}Ga]-PSMA administration to 14 days after the date of latest [^{68}Ga]Ga-PSMA-11 administration, i.e. from Day 1 do Day 14, as long as prior to the first dose of study treatment in the randomized part of the trial (i.e. prior to first dose of [^{177}Lu]-PSMA, abiraterone or enzalutamide)
3. ***post-treatment period***: starting at day 15 after [^{68}Ga]Ga-PSMA-11administration or the day of start of study treatment in the randomized part of the trial.

Note: [^{68}Ga]-PSMA is an imaging agent and not a treatment, however for ease of use throughout the document, we use pre/on/post treatment period terminology.

[^{177}Lu]Lu-PSMA-617crossover part of the protocol:

For adverse event reporting the overall observation period will be divided into three mutually exclusive segments:

1. ***pre-treatment period***: from day of patient's informed consent to the day before first administration of [^{177}Lu]Lu-PSMA-617
2. ***on-treatment period***: from date of first administration of [^{177}Lu]Lu-PSMA-617 to 30 days after EOT2 or last [^{177}Lu]-PSMA-617 dose date + 41 days, whichever is later.
3. ***post-treatment period***: starting at day 31 after EOT2 or last [^{177}Lu]-PSMA-617 dose date + 42 days, whichever is later.

Safety summaries (tables, figures) include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). In addition, a separate summary for death including on treatment and post treatment deaths will be provided. In particular, summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date during the on-treatment period (treatment-emergent AEs).

However, all safety data (including those from the post-treatment period) will be listed and those collected during the pre-treatment and post-treatment period will be flagged.

Windows for multiple assessments

In order to summarize PRO measures collected over time (including unscheduled visits), the assessments will be time slotted. If multiple assessments on the same date then the average will be used. Data from all assessments (scheduled and unscheduled), including multiple assessments, will be listed.

Time windows will be defined for descriptive summary of PRO data by visit and longitudinal data analysis. If more than one assessment is available in the same time window, the assessment closest to the planned visit date will be considered. If two assessments are obtained with the same time difference compared to the scheduled visit day, the average will be considered. Data obtained at the end of treatment will be classified as other assessment in the corresponding time window.

Table 2-1 Time windows for PRO – randomized part of the trial

Time Window	Planned Visit Timing	Time Window Definition
On treatment		
Baseline (Cycle 1 Week 1)	On or before Study Day 1*	≤ Study Day 1
Cycle 2 Week 1	Study Day 43	Study Days 2 – 63
Cycle 3 Week 1	Study Day 85	Study Days 64 – 105
Cycle 4 Week 1	Study Day 127	Study Days 106 – 147
Cycle 5 Week 1	Study Day 169	Study Days 148 – 189
Cycle 6 Week 1	Study Day 211	Study Days 190 - 231
Cycle 7 Week 1	Study Day 253	Study Days 232 - 295
Every 12 weeks thereafter (only for ARDT arm)		
Cycle k Week 1 (with k = 8, 9, …)	Study Day 253+ (k-7)*84	Study Days 253+(k-7)*84-41 to 253+(k-7)*84 +42 Note: EOT data visit are included if obtained within 30 days of EOT visit.
End of treatment		
End of treatment	N.A.	N.A.
Post treatment		
Post treatment follow-up Week 24	Post treatment study day 169	Post treatment Study Days 140 – 196
Post treatment follow-up Week 48	Post treatment study day 337	Post treatment Study Days 308 – 364
Study Day 1 = treatment start date Post treatment study day 1=end of treatment date + 1 day		

Last contact date

The last contact date will be derived for patients not known to have died at the analysis cut-off using the last complete date among the following:

Table 2-2 Last contact date data sources

Source data	Conditions
Date of Randomization	No Condition
Last contact date/last date patient was known to be alive from Survival Follow-up page	Patient status is reported to be alive or unknown.
Start/End dates from further antineoplastic therapy	Non-missing medication/procedure term.
Start/End dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
End of treatment date from end of treatment page	No condition.
- Tumor response assessment date 1. Tumor response assessment per RECIST 1.1/PCWG3 based on CT/MRI/bone scan 2. Other tumor response assessments, by the methods other than CT/MRI/bone scan	Evaluation is marked as 'done'.
Laboratory collection dates	Sample collection marked as 'done'.
Vital signs date	At least one non-missing parameter value
Performance Status date	Non-missing performance status
Start/End dates of AE	Non-missing verbatim term

The last contact date is defined as the latest complete date from the above list on or before the data cut-off date. The cut-off date will not be used for last contact date, unless the patient was seen or contacted on that date. No date post cut-off date will be used. Completely imputed dates (e.g. the analysis cut-off date programmatically imputed to replace the missing end date of a dose administration record) will not be used to derive the last contact date. Partial date imputation is allowed for event (death)/censoring is coming from 'Survival information' eCRF.

The last contact date will be used for censoring of patients in the analysis of overall survival.

2.2 Analysis sets

The Full Analysis Set (FAS) comprises all participants to whom study treatment has been assigned by randomization. According to the intent to treat principle (ITT), participants will be analyzed according to the treatment and strata they have been assigned to during the randomization procedure. The FAS will be the primary population for all efficacy analyses.

The Safety Set includes all participants who received at least one dose of study treatment (i.e. at least one dose of [^{177}Lu]Lu-PSMA-617 or ARDTs) during the randomized part of the protocol. Participants will be analyzed according to the study treatment received, where treatment received is defined as the randomized treatment if the participant took at least one dose of that treatment.

The Crossover Analysis Set (CAS) consists of participants randomized to the ARDT arm who crossed over after BICR-confirmed radiographic progression to receive at least one dose of [^{177}Lu]Lu-PSMA-617. This analysis set will be used for all analyses (rPFS2) pertaining to rPFS evaluations collected after participants crossed over in to [^{177}Lu]Lu-PSMA-617.

The RECIST Analysis Set (RAS) consists of subset of participants in the FAS with evaluable soft tissue disease by PCWG3-modified RECIST v1.1 at baseline, defined as having at least one lesion (target or non-target) identified at baseline in soft tissue. Participants will be included in the treatment arm to which they were randomized. This analysis set will be used for the analysis of RECIST-based endpoints (e.g., TTSTP, [redacted], etc.).

The Ga-PSMA-11 Full Analysis set (Ga-FAS) includes all participants who received the administration of [^{68}Ga]Ga-PSMA-11. This analysis set will be the basis of the safety and imaging summaries for the [^{68}Ga]Ga-PSMA-11 specific analyses. This includes all screened participants who received [^{68}Ga]Ga-PSMA-11 and were randomized, and those who received [^{68}Ga]Ga-PSMA-11 but not randomized.

The Lu-PSMA-617 Safety Set includes all participants who received at least one dose of [^{177}Lu]Lu-PSMA-617, during the randomized part of the protocol or after crossover. This analysis set will be used to provide safety analyses.

Patient Classification:

Patients may be excluded from the analysis populations defined above based on the protocol deviations entered in the database but will be followed up as per protocol.

Table 2-3 Subject classification based on protocol deviations

Analysis set	Protocol deviations leading to exclusion	Protocol deviation codes
Full Analysis Set	No written informed consent	INCL01A
Safety set	No written informed consent	INCL01A
Crossover Analysis Set	No written informed consent	INCL01A

RECIST Analysis Set	No written informed consent	INCL01A
Ga-PSMA-11 Full Analysis set	No written informed consent	INCL01A
Lu-PSMA-617 Safety Set	No written informed consent	INCL01A

Patients or assessments will be excluded from the respective analyses when one or more of the assessments are in serious violation of GCP.

2.2.1 Subgroup of interest

The primary efficacy and key secondary endpoints will be summarized by the following subgroups to examine the homogeneity of treatment effect :

- Stratification factor(s) (based on randomization data from IRT)
- Race (White vs. Black or African American vs. Asian vs. Other)
- Age category (< 65 years, ≥ 65 years)
- Patients with Liver Metastases (Yes vs. No), defined by sites of disease identified by central imaging review from PCWG3-modified RECIST 1.1
- Baseline PSA level (< median, vs. ≥ median)
- Initial Gleason Score (Score < 8 vs. Score ≥ 8)
- Baseline LDH level (≤ 260 IU/L vs. > 260 IU/L)
- Region (North America vs. Europe)
- Prior ARDT medication (abiraterone vs. enzalutamide).
- Baseline ECOG performance status (0 vs 1)

No formal statistical test of hypotheses will be performed for the subgroups, only point estimate of the treatment effect and 95% confidence intervals will be provided. The objective of the efficacy subgroup analysis is to demonstrate homogeneity of treatment effect in the above subgroups.

Safety

Safety subgroup analyses will use the same method as for the analysis in the safety set. Key safety analyses will be repeated on safety set in the following subgroups:

- Age group (< 65 years, ≥ 65 years)
- Race (White vs. Black or African American vs. Asian vs. Other)
- Baseline eGFR level (normal vs. mild impairment vs. moderate impairment vs. severe impairment), defined by the last non-missing eGFR measurement on or before treatment start date

- Baseline liver parameters ((ALT or AST > ULN) and BILI > ULN vs. (ALT and AST ≤ ULN) or BILI ≤ ULN), defined by the last non-missing ALT, AST and bilirubin measurement on or before treatment start date

The objective for carrying out these subgroup analyses is to identify potential safety issues that may be limited to a subgroup of patients, or safety issues that are more commonly observed in a subgroup of patients. Note: subgroup analyses will be conducted for [^{177}Lu]Lu-PSMA-617.

Subgroups will be presented if at least 10 patients are present in the subgroup.

Safety subgroups analyses will include the following analyses.:

- AEs, regardless of study drug, by primary system organ class and preferred term
- AEs with suspected relationship to study drug by primary system organ class and preferred term
- Serious AEs, irrespective of causality, by primary system organ class and preferred term
- Serious AEs with suspected relationship to study treatment, by primary system organ class and preferred term
- Safety Topics of Interest, irrespective of causality, by grouping, preferred term, maximum grade and treatment

2.3 Patient disposition, demographics and other baseline characteristics

The Full Analysis Set (FAS) will be used for all baseline and demographic summaries and listings. Summaries will be reported by treatment arm and for all patients and listings will be reported by treatment arm to assess baseline comparability. No inferential statistics will be provided.

Basic demographic and background data

All demographic and baseline disease characteristics data will be summarized and listed by treatment arm and crossover status. Categorical data (e.g. age groups: < 65 and ≥ 65 years, race, ethnicity, ECOG performance status) will be summarized by frequency counts and percentages; the number and percentage of patients with missing data will be provided. Continuous data (e.g. age, weight, height) will be summarized by descriptive statistics (N, mean, median, standard deviation, minimum and maximum). Additionally, available baseline data at the time of crossover will be summarized.

Baseline stratification factors

The number (%) of patients in each stratum (prior ARDT use in castrate-resistant prostate cancer (CRPC) vs. HSPC setting and symptomatology i.e. asymptomatic or mildly symptomatic (score of 0-3 on item 3 of the Brief Pain Inventory Short Form (BPI-SF) questionnaire) vs. symptomatic (score >3 on item 3 of the BPI-SF questionnaire)) based on data obtained from the IRT system will be summarized overall and by treatment arm for the FAS. Discordances between the stratum recorded in IRT at the time of randomization and the actual stratum

recorded in the clinical database through the data collected on eCRF will be cross-tabulated and listed.

Diagnosis and extent of cancer

Summary statistics will be tabulated for diagnosis and extent of cancer.
This analysis will include the following:

- disease characteristics at initial diagnosis: clinical stage, adenocarcinoma, Gleason score
- disease characteristics at study entry: number of sites of metastatic spread, types of distant metastases (soft tissue, bone), presence/absence of soft tissue lesions (per PCWG3-modified RECIST 1.1), presence/absence of bone lesions (per PCWG3)

Sites of metastatic spread will be based on diagnosis page.

Medical history

Medical history and ongoing conditions, including cancer-related conditions and symptoms entered on (e) CRF will be summarized and listed by treatment arm. The summary will be presented by primary system organ class (SOC), preferred term (PT) and treatment arm. Medical history and current medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable tables/listings.

Other

All data collected at baseline will be listed.

2.3.1 Patient disposition

Enrollment by country and center will be summarized for all screened patients and also by treatment arm using the FAS. The number (%) of randomized patients included in the FAS will be presented overall and by treatment group. The number (%) of screened and not-randomized patients and the reasons for screening failure will also be displayed. The number (%) of patients in the FAS who are still on treatment, who discontinued the study phases and the reason for discontinuation will be presented overall and by treatment group.

The following summaries will be provided (with % based on the total number of FAS patients):

- Number (%) of patients who were randomized (based on data from IRT system)
- Number (%) of patients who were randomized but not treated (based on ‘DAR’ eCRF page not completed for any study treatment component)
- Primary reason for not being treated (based on “End of Treatment Phase Completion” eCRF page)
- Number (%) of patients who were treated (based on ‘DAR’ eCRF pages of each study treatment component completed with non-zero dose administered)
- Number (%) of patients who are still on-treatment (based on the ‘End of Treatment Phase’ page not completed)

- Number (%) of patients who discontinued the study treatment phase (based on the 'End of Treatment Phase' page)
- Primary reason for study treatment phase discontinuation (based on the 'End of Treatment Phase' page)
- Number (%) of patients who have entered the post-treatment follow-up (based on the 'End of Treatment Phase' page)
- Number (%) of patients who have discontinued from the post-treatment follow-up (based on the End of Post-treatment follow-up page)
- Reasons for discontinuation from the post-treatment follow-up (based on Study End of Post-treatment follow-up page)
- Number (%) of patients who have entered the survival follow-up (based on the 'End of Treatment Phase' or 'End of Post-treatment follow-up' page).

Protocol deviations

The number (%) of patients in the FAS with any protocol deviation will be tabulated by deviation category (as specified in the study Data Handling Plan) overall and by treatment group for the FAS. All protocol deviations will be listed. Specific protocol deviation categories will be assigned to important deviations related to COVID-19 (e.g. missing efficacy assessments and treatment interruptions) and these will be summarized separately and flagged in the listings.

For all important protocol deviations, the relationship to COVID19 will also be captured, except for a pre-defined list of protocol deviations where relationship to COVID-19 is not applicable.

The relationship to COVID-19 is defined using the following descriptions:

- COVID-19 Health Status: (i.e. participant's COVID-19 infection led to this PD)
- COVID-19 Site issue: (e.g. site closed, personnel not available)
- COVID-19 Lockdown: (e.g. site is active but patient not allowed to come)
- COVID-19 Subject/Patient concern: (e.g. site is active, subject/patient could come but refused to come / do assessment)
- COVID-19 Drug supply issue (e.g. drug was delivered to home)
- COVID-19 Other: (e.g. situation not already covered by the information above)

Analysis sets

The number (%) of patients in each analysis set (defined in Section 2.3) will be summarized by treatment group and stratum.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

The Safety Set will be used for the analyses of treatments received during the randomized part of the study. Analyses of [^{68}Ga]Ga-PSMA-11 will be performed on the Ga-FAS (see Section 2.2). Treatment received after the crossover will be analyzed in the CAS.

2.4.1 Study treatment / compliance

Duration of exposure, actual cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized by treatment arm. Duration of exposure will be categorized into time intervals; frequency counts and percentages will be presented for the number(%) of subjects in each interval. The number (%) of subjects who have dose reductions or interruptions, and the reasons, will be summarized by treatment group. The number of [^{177}Lu]Lu-PSMA-617 injections will also be summarized.

The same analyses will be conducted for patients who crossed over from ARDT to receive [^{177}Lu]Lu-PSMA-617 in the CAS. Analyses will also be presented in the Lu-PSMA-617 Safety Set.

Actual dose, DI and RDI will be presented for [^{68}Ga]Ga-PSMA-11 in the Ga-FAS.

Subject level listings of all doses administered on treatment along with dose change reasons will be produced.

Duration of exposure to study treatment

Duration of exposure to study treatment (days) = (last date of exposure to study treatment) – (date of first administration of study treatment) + 1.

Summary of duration of exposure of study treatment in appropriate time units will include categorical summaries and continuous summaries (i.e. mean, standard deviation etc.) using appropriate units of time.

Table 2-4 Definition of last date of exposure of study treatment

Drug	Definition of last date of exposure of study drug	Example
[^{177}Lu]-PSMA-617	The planned end date of the last cycle in which the last non-zero dose of the investigational drug was last administered. [^{177}Lu]Lu-PSMA-617 is given as a 42-day cycle with one infusion in the beginning of the cycle, then the last date of exposure is the date of infusion in the last cycle + 41 days. Note : If the patient died or was lost to follow-up before the derived last date, the last date of exposure to investigational drug is the date of death or the date of last contact, respectively.	The infusion of the last cycle was given on 01Apr2021, then the last date of exposure is the date of infusion in the last cycle + 41 days, i.e. 12May2021.

	If the derived last date of exposure goes beyond the data cutoff date, it should be truncated to the date of data cutoff.	
ARDT (abiraterone or enzalutamide)	Date of last administration of a non -zero dose of the study drug.	A patient had a permanent discontinuation of the study drug 06Jan2021 after being put on a temporary interruption since 01Jan2021. In this case the last date of exposure is-31Dec2020.

Number of [^{177}Lu]Lu-PSMA-617 injections

In the [^{177}Lu]Lu-PSMA-617 arm, the number of [^{177}Lu]Lu-PSMA-617 injections will be analyzed by descriptive continuous summaries. The frequency of patients per number of injections will also be presented.

Cumulative dose

Cumulative dose of a study treatment is defined as the total dose given during the study treatment exposure and will be summarized for each of the study treatment. In the control arm, abiraterone and enzalutamide will be presented separately.

The **planned cumulative dose** for a study treatment component refers to the total planned dose as per the protocol up to the last date of investigational drug administration.

- For [^{177}Lu]Lu-PSMA-617, the planned dose for each injection is 7.4 GBq
- For [^{68}Ga]Ga-PSMA-11, the planned dose is 150 MBq
- For abiraterone, the planned daily dose is 1000 mg
- For enzalutamide, the planned daily dose is 160 mg

The **actual cumulative dose** refers to the total actual dose administered, over the duration for which the subject is on the study treatment as documented in the Dose Administration eCRF.

For patients who did not take any drug the cumulative dose is by definition equal to zero.

For continuous dosing (abiraterone and enzalutamide), the actual cumulative dose is the sum of the non-zero doses recorded over the dosing period and the planned cumulative dose is the planned starting dose summed over the same dosing period.

Dose intensity and relative dose intensity

Dose intensity (DI) for patients with non-zero duration of exposure is defined as follows:

- For [^{177}Lu]Lu-PSMA-617: DI (GBq/month) = Actual Cumulative dose (GBq) / Duration of exposure to study treatment (months).

- For abiraterone/enzalutamide: DI (mg /day) = Actual Cumulative dose (mg) / Duration of exposure to study treatment (days).
- For [^{68}Ga]Ga-PSMA-11: DI (MBq) = Actual Cumulative dose (MBq). Note: [^{68}Ga]-PSMA-11 is given as a single injection

For patients who did not take any drug the DI is by definition equal to zero.

Planned dose intensity (PDI) is defined as follows:

- For [^{177}Lu]Lu-PSMA-617: PDI (GBq/month) = Planned Cumulative dose (GBq) / Duration of exposure (months).
- For abiraterone/enzalutamide: PDI (mg /day)= Planned Cumulative dose (mg) / Duration of exposure (days).
- For [^{68}Ga]Ga-PSMA-11: PDI (MBq)= Planned dose (MBq)

Relative dose intensity (RDI) is defined as follows:

RDI = DI / PDI

DI and RDI will be summarized for all study drugs. In the control arm, abiraterone and enzalutamide will be presented separately.

Dose reductions, interruptions or permanent discontinuations

The number of subjects who have dose reductions, permanent discontinuations or interruptions, and the reasons, will be summarized separately for each of the study treatments.

'Dose changed' and 'Dose interrupted' fields from the Dosage Administration CRF pages (DAR) will be used to determine the dose reductions and dose interruptions. Permanent discontinuations will be determined based on the treatment disposition CRF page.

The corresponding fields 'Reason for dose change/dose interrupted' and 'Reason for permanent discontinuation' will be used to summarize the reasons.

A dose change is either 'change in prescribed dose level' or 'dosing error' where actual dose administered/total daily dose is different from the prescribed dose.

For the purpose of summarizing interruptions and reasons, in case multiple entries for interruption that are entered on consecutive days with different reasons will be counted as separate interruptions. However, if the reason is the same in this mentioned multiple entries on consecutive days, then it will be counted as one interruption.

Reduction: A dose change where the prescribed dose level is lower than the previous prescribed dose level or where the actual dose administered/total daily dose is lower than the calculated dose amount based on the prescribed dose. Therefore, any dose change to correct a dosing error will not be considered a dose reduction. Only dose change is collected in the CRF, number of reductions will be derived programmatically based on the change and the direction of the change.

Table 2-5 Dose reductions examples for [^{177}Lu]Lu-PSMA-617

DAR record number[1]	Prescribed dose (GBq)	Administered dose (GBq)	Dose reduction Yes/No	Comment
1	7.4	7.4	No	
2	7.4	8.5	No	Dosing error
3	7.4	7.4	No	Correcting dosing error
4	5.9	5.9	Yes	Due to AE
5	5.9	2.5	Yes	Dosing error
6	5.9	5.9	No	Correcting dosing error

Note: It is assumed that if dose reduction is yes, that dose change was checked.

For ARDT, the prescribed doses are not collected on the eCRF. Therefore, the prescribed doses at the start of the trial will be assumed to be 1000 mg for abiraterone and 160 mg for enzalutamide. Reduction is defined as a dose change where the actual dose is lower than the previous actual dose or where the actual dose administered/total daily dose is lower than the calculated dose amount based on the prescribed dose defined above. In the same way as for [^{177}Lu]Lu-PSMA-617, a dose reduction following a dosing error for ARDT will not be counted as a dose reduction, provided it is not below the dose that was actually given before the dosing error.

Table 2-6 Dose reductions examples for abiraterone

DAR record number[1]	Prescribed dose (mg)	Administered dose (mg)	Dose reduction Yes/No	Comment
1	1000	1000	No	
2	1000	1250	No	Dosing error
3	1000	1000	No	Correcting dosing error
4	1000	500	Yes	Due to AE
5	1000	250	Yes	Dosing error
6	1000	500	No	Correcting dosing error
7	1000	1000	No	Dosing error
8	1000	250	Yes	This is below the 500 mg that was given before the dosing error

Note: It is assumed that if dose reduction is yes, that dose change was checked.

2.4.2 Prior, concomitant and post therapies

Prior anti-cancer therapy

The number and percentage of patients who received any prior anti-neoplastic medications, prior anti-neoplastic radiotherapy or prior anti-neoplastic surgery will be summarized by treatment arm. Prior anti-neoplastic medications will be summarized by therapy type (e.g. chemotherapy, hormonal therapy etc.), setting (e.g. adjuvant, metastatic, etc.) and also by lowest ATC class, preferred term and treatment. Summaries will include total number of regimens, number of progressions since diagnosis, best response and time from last treatment to progression for the last therapy. For radiotherapy, time since last radiotherapy, locations and setting of last therapy will be summarized. For prior surgery, procedure and time since last surgery will be summarized.

Separate listings will be produced for prior anti-neoplastic medications, radiotherapy, and surgery.

Anti-neoplastic medications will be coded using the WHO Drug Dictionary (WHO-DD); anti-neoplastic surgery will be coded using MedDRA. Details regarding MedDRA and WHO-DD version will be included in the footnote in the tables/listings.

The above analyses will be performed using the FAS and the listings will also be provided for Ga-FAS.

Post treatment anti-cancer therapy

Anti-neoplastic therapies since discontinuation of study treatment will be listed and summarized by ATC class, preferred term, overall and by treatment group by means of frequency counts and percentages using FAS. This will also be summarized by medication type, ATC class, preferred term, overall and by treatment group and crossover status. Similarly, number of anti-neoplastic regimens and therapy types (e.g. chemotherapy, hormonal therapy etc.) of subsequent medications for the first regimen (and overall) will be summarized by arm and crossover status. Reasons for discontinuation of the first regimen and reasons for not receiving subsequent anti-neoplastic therapy will also be summarized by arm and crossover status.

First anti-neoplastic therapies since discontinuation of study treatment will be summarized by medication type, ATC class, preferred term, overall and by treatment group and crossover status by means of frequency counts and percentages using FAS.

Time to first anti-neoplastic therapy will be defined as time from randomization to initiation of the first subsequent anti-neoplastic therapy or death, whichever occurs first. Crossover to [^{177}Lu]Lu-PSMA-617 arm will not constitute an event. Participants who did not have an event at the analysis data cut-off or are lost to follow-up at the time of analysis will be censored at the time of their last contact.

Time to first anti-neoplastic therapy will be analyzed in the FAS population according to the randomized treatment group and strata assigned at randomization. The distributions will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians and 95%

confidence intervals of the medians will be presented for each treatment group. The medians will also be presented separately for ARDT patients who crossed over and those who did not. The hazard ratio will be calculated, along with its 95% confidence interval, using a stratified Cox model.

Concomitant medications

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) other than the study treatment administered to a patient coinciding with the study treatment period. Concomitant therapy include medications (other than study drugs) starting on or after the start date of study treatment or medications starting prior to the start date of study treatment and continuing after the start date of study treatment.

Concomitant medications will be coded using the World Health Organization (WHO) Drug Reference Listing (DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (ATC) classification system and summarized by lowest ATC class and preferred term using frequency counts and percentages. Surgical and medical procedures will be coded using MedDRA and summarized by SOC and preferred term. These summaries will include:

1. Medications starting on or after the start of study drug but no later than the end of on-treatment period and
2. Medications starting prior to start of study drug and continuing after the start of study treatment.

All concomitant therapies will be listed. Any concomitant therapies starting and ending prior to the start of study treatment or starting after the end of the on-treatment period will be flagged in the listing. The safety set will be used for all concomitant medication tables and listings.

Antineoplastic radiotherapy administered during the randomized treatment phase will also be summarized by treatment group.

Above analyses will be performed on the Safety Set and Lu-PSMA Safety Set.

2.5 Analysis of the primary objective

The primary objective of the study is to determine whether treatment with [^{177}Lu]Lu-PSMA-617 prolongs radiographic progression-free survival (PFS) compared to treatment with ARDT (abiraterone or enzalutamide) arm in patients with progressive PSMA-positive mCRPC.

2.5.1 Primary endpoint

The primary estimand is defined in Section 1.2.1. The primary endpoint (variable attribute of the primary estimand) of the study is radiographic progression-free survival (rPFS), defined as the time from the date of randomization to the date of the first documented radiographic progression as outlined in PCWG3 or death due to any cause. rPFS will be assessed via blinded independent central review (BICR) of radiographic images provided by the treating physician.

Radiographic progression will be assessed:

- in soft tissue, using RECIST1.1 with modifications outlined in the PCWG3 guideline and defined in the study imaging charter
- in bone disease, using bone scan and the 2+2 rule defined in the PCWG3.

The 2+2 rule for bone progression is defined as follows:

1. Rule 1 (Progression at week 8 confirmed at week 16): If there are at least two new lesions on the first post-treatment scan, they must be confirmed with at least two additional lesions on the next scan (2+2 rule) obtained at least 6 weeks later and outside the 12 week flare period. The date of progression is the date of the first post-treatment scan.
2. Rule 2 (Progression at week 12 or later confirmed at next scan): For scans after the 12 week flare period, there must be at least two new lesions relative to the baseline or first post-treatment scan (if treated as a new baseline) that remain persistent (confirmed) on a subsequent scan obtained at least 6 week later. The date of progression is the date of the scan that first documents the second lesion compared to first post-treatment scan.

The patient will be considered progressing on the date of the earliest of the progressions in soft tissue or bone.

Participants who are alive without radiographic progression at the analysis data cut-off or are lost to follow-up at the time of analysis will be censored for rPFS at the time of their last radiographic assessment. Clinical deterioration without objective radiographic evidence will not be considered as documented radiographic progression.

Handling of missing data are provided in Section 2.5.4.

2.5.2 Statistical hypothesis, model, and method of analysis

Assuming proportional hazards for rPFS, the following statistical hypotheses will be tested to address the primary efficacy objective:

$$H_{01}: \theta_1 \geq 1 \quad \text{vs.} \quad H_{A1}: \theta_1 < 1$$

where θ_1 is the rPFS hazard ratio ([^{177}Lu]Lu-PSMA-617 arm versus ARDT arm).

The primary efficacy analysis to test these hypotheses and compare the two treatment groups will consist of a stratified log-rank test at an overall one-sided 2.5% level of significance. The stratification will be based on the randomization stratification factors, i.e. prior ARDT use in CRPC vs HSPC; asymptomatic and mildly symptomatic vs symptomatic participants (score of 0-3 on item 3 of the Brief Pain Inventory Short Form (BPI-SF) questionnaire) vs symptomatic (score >3 on item 3 of the BPI-SF questionnaire).

The primary efficacy variable, rPFS, was to be analyzed after approximately 156 rPFS events were observed. Analyses will be based on the FAS population according to the randomized treatment group and strata assigned at randomization (strata formed using the randomization factor as obtained via IRT). The rPFS distribution will be estimated using the Kaplan-Meier method, and Kaplan-Meier curves, medians and 95% confidence intervals of the medians will

be presented for each treatment group. The hazard ratio for rPFS will be calculated, along with its 95% confidence interval, from a stratified Cox model using the same stratification factors as for the log-rank test.

[redacted]

2.5.3 Handling of remaining intercurrent events of primary estimand

The primary analysis will account for different intercurrent events as explained in the following:

1. Discontinuation of study treatment for any reason: Tumor assessment data collected irrespective of discontinuation of study treatment will be used for the analysis (treatment policy strategy)
2. Change in supportive care: Tumor assessment data collected irrespective of change in supportive care will be used for the analysis (treatment policy strategy)
3. Start of a new anti-neoplastic therapy prior to radiographic progression or death: rPFS events documented irrespective of initiation of new anti-neoplastic therapy will be used for the primary analysis (treatment policy strategy).

2.5.4 Handling of missing values not related to intercurrent event

If rPFS event is observed after two or more missing or non-adequate response assessments, then rPFS will be censored at the last adequate response assessment before the rPFS event. If rPFS event is observed after a single missing or non-adequate tumor assessment, the actual date of event will be used.

The term "missing or non-adequate tumor assessment" is defined as a response assessment (TA) not performed or response assessment with overall lesion response of "UNK". In this study, progression can be declared by PCWG3-modified RECIST v1.1 for soft tissue lesions based on CT/MRI or bone scan per PCWG3 rule for bone lesions. A response assessment will be considered 'UNK' if soft-tissue lesions(s) response is missing/'UNK' or bone disease response is missing/'UNK', unless a progressive disease is reported by one of the two methods.
The rule to determine number of missing TAs is based on the time interval between the date of last adequate response assessment and the date of an event. If the interval is greater than twice the protocol-specified interval between the TAs and 2 times the protocol-allowed time window around assessments, then the number of missing assessments will be 2 or more. The scans in this study will be collected every 8 weeks after first dose of study treatment for the first 24 weeks (week 8, 16, 24) and then every 12 weeks (week 36, 48, etc) until confirmation of radiographic progression by BICR. The protocol allowed time window is ±7 days. Exact definitions for determining missed assessments will be provided in the Programming and Dataset Specifications (PDS) document.

Refer to Table 2-7 for censoring and event date options and outcomes for rPFS.

Table 2-7 Outcome and event/censor dates for rPFS analysis

Situation	Date	Outcome
No baseline assessment	Date of randomization	Censored
Radiographic progression or death at or before next scheduled Assessment	Date of radiographic progression (or death)	Event
Radiographic progression or death after exactly one missing assessment	Date of radiographic progression (or death)	Event
Progression or death after two or more missing assessments	Date of last adequate assessment prior to missed assessments	Censored
No radiographic progression (or death)	Date of last adequate assessment	Censored
New anticancer therapy given prior to protocol defined progression (*including patients who crossover from the control to the treatment arm*[1])	Ignore the new anticancer therapy and follow situations above	As per above situations
Death before first radiographic PD assessment	Date of death	Event

1. It is not allowed to crossover before documented radiographic progression confirmed by BICR, however this rule will be applied in case of protocol deviation

2.5.5 Sensitivity analyses

The same analysis conventions as the primary efficacy analysis will be used, with the exception of the specific rule for sensitivity mentioned in each analysis, and the treatment effect will be summarized by the hazard ratio with its 95% confidence interval. Kaplan-Meier curves, medians and 95% confidence intervals of the medians will be presented for each treatment group.

The following sensitivity analyses for rPFS will be conducted:

1. Hazard Ratio and 95% CI obtained from a stratified and covariate unadjusted Cox model with stratification factors derived from the clinical database (as collected on eCRF), in case at least 5% of the participants have discrepancies between strata at randomization (using IRT data) and strata derived from the eCRF data. In the summary tables, this approach is referred as ‘actual stratification rPFS sensitivity analysis’.
2. To account for the foreseen imbalanced delay between randomization and first treatment date (up to 14 days may be needed for site to administer [177Lu]Lu-PSMA-617), rPFS will be recalculated using treatment start date. Patients never treated will be excluded from this sensitivity analysis. In the summary tables, this approach is referred as ‘treatment start rPFS sensitivity analysis’. This is not aligned with treatment policy strategy.

3. rPFS will be recalculated by backdating tumor scan responses to the theoretical previous tumor assessment date determined by calculating the theoretical schedule from the randomization date. This approach is referred as 'Previous theoretical date rPFS sensitivity analysis'.
4. rPFS as per investigator review will be analyzed. rPFS as per investigator will be censored at the last adequate tumor assessment before the date of the start of crossover if no rPFS event is observed by investigator prior to crossover. In the summary tables, this approach is referred as 'investigator rPFS sensitivity analysis'.
5. including events whenever it occurs, even after two or more missing tumor assessments. In the summary tables, this approach is referred as 'actual event rPFS sensitivity analyses.
6. backdating of events occurring after missing one or more tumor assessments. In this analysis, the event will be assumed to be at the next scheduled assessment after the last adequate tumor assessment. In the summary tables, this approach is referred as 'backdating rPFS sensitivity analysis'.
7. censoring rPFS at the last adequate tumor assessment before the date of the start of new anticancer therapy if no rPFS event is observed prior to the start of new antineoplastic therapy. In the summary tables, this approach is referred as 'new anticancer therapy rPFS sensitivity analysis'.
8. censoring COVID-19 related deaths at the last adequate tumor assessment prior to the death. In the summary tables, this approach is referred as 'censoring COVID-19 related deaths rPFS sensitivity analysis a'.
9. censoring COVID-19 related deaths at the date of death. In the summary tables, this approach is referred as 'censoring COVID-19 related deaths rPFS sensitivity analysis b'.
10. censoring for withdrawal of consent at the time of withdrawal. In the summary tables, this approach is referred as 'censoring for withdrawal of consent at the time of withdrawal rPFS sensitivity analysis'.
11. treating withdrawal of consent as an event at the time of withdrawal. In the summary tables, this approach is referred as 'treating withdrawal of consent as an event at the time of withdrawal rPFS sensitivity analysis'.
12. censoring for more than 17 days delay to start treatment at the time of randomization. In the summary tables, this approach is referred as 'censoring for delayed treatment start at the time of randomization rPFS sensitivity analysis'.

2.5.6 Supplementary analyses

A multivariate Cox regression model stratified by randomization stratification factors will be fitted to evaluate the effect of other baseline demographic and disease characteristics on the estimated hazard ratio. The fitted model adjusting the treatment difference for key baseline and prognostic factors will include as covariates the following: Liver Metastases (Yes vs. No), Baseline PSA level (< vs ≥ median), Initial Gleason Score (Score < 8 vs. Score ≥8), ECOG (0 vs 1) and Baseline LDH level (< 260 vs ≥ 260 IU/L).
Further supportive analyses will include:

- Number of patients and number of events by treatment arm within each stratum will be presented along with the hazard ratio for treatment effect obtained using the Cox proportional hazards regression with corresponding confidence intervals. No p-values will be presented for this analysis.
- Timing of all tumor assessments will be depicted graphically for central radiology and displayed by treatment arm

Subgroup analyses for the primary endpoint

The primary endpoint of rPFS will be summarized for the subgroups specified in Section 2.2.1 and using the same conventions as for the primary analysis.

For each of the subgroups, the following analyses will be performed:

- Kaplan-Meier estimates of the survival distribution of rPFS
- Hazard ratio with 95% CI using stratified Cox proportional hazards model. For subgroup analyses by stratification factor, only the other stratification factor will be included in the model.

Efficacy analyses in subgroups are intended to explore the consistency (homogeneity) of treatment effect. Forest plot (including sample size/number of events and HR with 95% CI) will be produced to graphically depict the treatment effect estimates in different subgroups. No inferential statistics (p-values) will be produced for the subgroups.

Censoring pattern of rPFS

Number of patients with a rPFS event and number of patients censored for the rPFS analysis will be summarized. In addition, a summary of reasons for rPFS censoring will be provided by treatment arm based on the following reasons:

1: Ongoing without event
2: Lost to follow-up
3: Withdrew consent
4: Adequate response assessment no longer available
5: Event after >=2 missing response assessments

The rPFS censoring reasons are defined in the following way.

If the time interval between the last adequate TA date and the earliest of the following dates is smaller or equal to interval of 2 missing response assessments (see Section 2.5.1 for definition):

1. Analysis cut-off date,
2. Date of consent withdrawal,
3. Visit date of study treatment discontinuation or end of post-treatment follow-up discontinuation due to lost to follow-up.

Then the rPFS censoring reason will be:

1. ‘Ongoing’,

2. ‘Withdrew consent’,
3. ‘Lost to follow-up’,

If the time interval is larger than the interval of 2 missing tumor assessments with no event observed, then the rPFS censoring reason will always default to ‘Adequate assessment no longer available’. If the time interval between the last adequate tumor assessment date and the rPFS event date is larger than the interval of 2 missing tumor assessments then the patient will be censored and the censoring reason will be 'Event documented after two or more missing tumor assessments'.

These summaries on censoring reasons will be produced for rPFS by BICR. The censoring patterns will be compared between treatment arms within each of the two comparisons and also between investigator and BICR. Summary of the difference in days to radiographic progression as per BICR and investigator will also be generated.

Concordance analysis of rPFS

Cross-tabulation of ‘rPFS by central radiology’ vs. ‘rPFS by investigator’ by rPFS event type (i.e. ‘death’, ‘PD’, ‘censor’ for each of the two sources resulting in a 3-by-3 table) and by treatment will be constructed to investigate dis-concordance between the two sources on patient-by-patient basis. Discordance rate between central radiology and investigator will be calculated and presented as % as follows: $100 \times (n_{13} + n_{23} + n_{31} + n_{32})/ N$ by treatment arm.

Table 2-8 Comparison of rPFS between investigator and BICR

Investigator rPFS result	BICR rPFS result		
	Death	rPD	Censor
Death	n_{11}	n_{12}	n_{13}
rPD	n_{21}	n_{22}	n_{23}
Censor	n_{31}	n_{32}	n_{33}

A cross-tabulation will be produced displaying the rPFS timings for the local investigators’ assessment compared to the BICR assessment. For progression assessments, the frequency and percent of subjects with complete agreement [occurring on the same date plus or minus 14 days of each other], progression later, progression earlier, and cases where progression was called by one method and censored by the other will be displayed. Similarly, if censoring was recorded, the frequency and percent of subjects with complete agreement, censoring called later, censoring called earlier, and cases where censoring was called by one method and progression was called by the other method will be displayed.

Table 2-9 Comparison of rPFS event times between BICR and local assessments

		Treatment arm (N = XXX)			
Investigator	**BICR**	**Same time n (%)**	**BICR after Investigator n (%)**	**BICR before Investigator n (%)**	**Total**

rPD	rPD	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Death	Death	xx (xx.x)	0	0	xx (xx.x)
Censor	Censor	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
rPD	Censor	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
rPD	Death	0	xx (xx.x)	0	xx (xx.x)
Death	rPD	0	0	xx (xx.x)	xx (xx.x)
Censor	rPD	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Censor	Death	0	xx (xx.x)	0	xx (xx.x)
Total		xx (xx.x)	xx (xx.x)	xx (xx.x)	xxx (100.0)

2.6 Analysis of the key secondary objective

The key secondary objective of the study is to determine whether treatment with [^{177}Lu]Lu-PSMA-617 prolongs OS compared with ARDT.

A hierarchical testing strategy will be used to control the overall type I error rate, where OS will only be formally tested and interpreted if the primary analysis of rPFS is statistically significant.

2.6.1 Key secondary endpoint

The primary estimand is defined in Section 1.2.2. The key secondary endpoint (variable attribute of the key secondary estimand) of the study is OS, defined as the time from randomization to death due to any cause.

Participants who are alive at the analysis data cut-off or are lost to follow-up at the time of analysis will be censored for OS at the time of their last contact, as defined in Section 2.1.1.

2.6.2 Statistical hypothesis, model, and method of analysis

Assuming proportional hazards model for OS, the following statistical hypotheses will be tested only if rPFS is statistically significant:

$$H_{02}: \theta_2 \geq 1 \quad \text{vs.} \quad H_{A2}: \theta_2 < 1$$

where θ_2 is the OS hazard ratio ([^{177}Lu]Lu-PSMA-617arm versus ARDT arm).

The analysis to test these hypotheses will consist of a stratified log-rank test at an overall one-sided 2.5% level of significance.

The final OS analysis will not be performed at the time point of the final rPFS analysis, but after additional follow-up. Therefore, a four-look group sequential design is considered for OS.

OS will be hierarchically tested in the following way:

- The time point for the first OS interim analysis was at the same time as the primary rPFS analysis, when a very limited information fraction was observed
- OS was not statistically significant at the first interim analysis. The time point for the second OS interim analysis was after approximately 42% of deaths (125 deaths) had been recorded in the clinical database.
- OS was not statistically significant at the second interim analysis. The time point for the third OS interim analysis will be after approximately 75% of deaths (223 deaths) have been recorded in the clinical database.
- OS was not statistically significant at the third interim analysis. A final analysis is planned at the time approximately 297 deaths have been recorded.

The type I error probability will be controlled by using a Lan-DeMets (O'Brien-Fleming) alpha spending function for OS which is independent of the one used for rPFS. This guarantees the protection of the overall type I error ($\alpha = 2.5\%$) across all hypotheses and the repeated testing of the OS hypotheses at the interim and the final analyses (Glimm et al 2010).

The key secondary endpoint, OS will be analyzed following the ITT principle based on the FAS population according to the treatment group and strata assigned at randomization. The OS distribution will be estimated using the Kaplan-Meier method, and Kaplan-Meier curves, medians and 95% confidence intervals of the medians will be presented for each treatment group. The hazard ratio for OS will be calculated, along with its 95% confidence interval, from a stratified Cox model using the same stratification factors as for the log-rank test.

2.6.3 Handling of remaining intercurrent events of key secondary estimand

The OS analysis will account for different intercurrent events as explained in the following:

1. Discontinuation of study treatment for any reason: OS data collected irrespective of discontinuation of study treatment will be used for the analysis (treatment policy strategy)
2. Change in best supportive care: OS data collected irrespective of change in best supportive care will be used for the analysis (treatment policy strategy)
3. Start of a new anti-neoplastic therapy: OS data collected irrespective of initiation of new anti-neoplastic therapy or switching to [^{177}Lu]Lu-PSMA-617 will be used for the analysis (treatment policy strategy)

2.6.4 Handling of missing values not related to intercurrent event

If a patient is not known to have died at the time of analysis cut-off, then OS will be censored at the date of last known date patient was alive, i.e., last contact date (see Section 2.1.1).

2.6.5 Supplementary analyses

OS analysis will be performed without accounting for crossover, following the ITT principle based on the Safety Set. Unstratified OS analysis based on FAS will also be performed.

Since participants in the ARDT arm are allowed to crossover to [^{177}Lu]Lu-PSMA-617 arm upon confirmation of rPFS by BICR, adjustment for the effect of crossover on OS will be performed as supplementary analyses based on recognized methods.

The proportion and number of patients who crossed over from the ARDT arm to [^{177}Lu]Lu-PSMA-617 treatment will be provided. Time from randomization to crossover date will be summarized descriptively.

The proportion of total duration of exposure and follow-up time affected by crossover will be summarized descriptively.

A subset of demographics and baseline characteristics will be provided for the subgroup of patients who crossed over from ARPI arm to [^{177}Lu]Lu-PSMA-617 treatment.

The rank preserving structural failure time (RPSFT) model (Robins J and Tsiatis A 1991) is intended to be used to adjust for crossover for OS, however other methods, not limited to the ones below may also be used based on an examination of the appropriateness of the data to the assumptions required by the methods. Details for its implementation will be provided in a standalone analysis plan. As a sensitivity analysis, RPSFT may be conducted without applying re-censoring as proposed by (Latimer N 2019).

The following methods may also be used as appropriate:

- Inverse Probability of Censoring Weighting (IPCW) proposed by (Robins J and Finkelstein D 2000)
- A two-stage method proposed by (Latimer N 2014)

Restricted Mean Survival Time (RMST) method may be conducted for OS to account for a possible non-proportional hazards effect.
As a sensitivity analysis, censoring COVID-19 related deaths at the date of death may be performed if there are sufficient number of COVID-19 related deaths.

2.7 Analysis of secondary efficacy objective(s)

The other secondary efficacy objectives are to:

- estimate the time to radiographic progression or death in participants treated with ARDT who subsequently crossover to [^{177}Lu]Lu-PSMA-617 after radiographic progression (rPFS2)
- evaluate Progression Free Survival (PFS) by investigator's assessment
- evaluate the second progression Free Survival (PFS2) by investigator's assessment
- evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the biochemical response as detected by Prostate specific antigen (PSA) halving compared to participants treated with ARDT
- evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the time to first symptomatic skeletal event (TTSE) compared to participants treated with ARDT

- evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the time to radiographic soft tissue progression compared to participants treated with ARDT
- evaluate whether treatment with [^{177}Lu]Lu-PSMA-617 improves the time to chemotherapy compared to participants treated with ARDT

2.7.1 Secondary endpoints

rPFS2 defined as the time from the date of crossover (ARDT to [^{177}Lu]Lu-PSMA-617) to the date of radiographic disease progression assessed via blinded independent central review or death from any cause on the next line of therapy ([^{177}Lu]Lu-PSMA-617). The same analyses conventions as depicted in Sections 2.5.3 and 2.5.4 will be used to derive rPFS2.

PFS is defined as time from date of randomization to the first documented progression by investigator's assessment (radiographic, clinical, or PSA progression) or death from any cause, whichever occurs first. The same analyses conventions as depicted in Sections 2.5.3 and 2.5.4 will be used to derive PFS. PFS will be censored at the last adequate tumor assessment before the date of the start of new anticancer therapy (including crossover) if no PFS event is observed prior to new anticancer therapy.

PFS2 is defined as time from date of randomization to the first documented progression (radiographic progression, clinical progression, PSA progression) on next-line therapy or death from any cause, whichever occurs first. The first documented progression on next-line treatment is based on investigator's assessment of progression disease (PD) (i.e. as captured on the anti-neoplastic therapy after treatment discontinuation CRF page); it is not necessary to continue to collect tumor assessments data for subsequent anti-neoplastic therapies for the purpose of PFS2.

- Next-line therapy is defined as the first new (systemic) anti-neoplastic therapy initiated after discontinuation of study treatment regardless of EoT reason. Drugs given as part of the same regimen should be considered as first line (i.e. part of the next-line therapy).
- New anti-neoplastic therapies after EoT will be collected in the anti-neoplastic therapy after treatment discontinuation eCRF page including start/end date, reason for discontinuation, date and type of progression («clinical» vs «radiographic»).
- PFS2 will be censored if no PFS2 event (progression or death) is observed during next-line therapy before the analysis cut-off date; the censoring date will be the last contact date.
- However, in case a second new anti-neoplastic therapy is introduced without prior PFS2 event, then PFS2 will be censored at the end date of the first new anti-neoplastic therapy (i.e. next line therapy).
- Any death prior to initiation of next-line therapy will be considered as an event for PFS2. Any death occurring following end of next line therapy will be considered as an event if no second new anti-neoplastic therapy has been introduced.
- PFS and PFS2 may be identical if a participant did not experience an event (i.e. progression) prior to initiation of next-line therapy, and adequate tumor assessments continue until documented disease progression after initiation of next-line therapy.

- Note: for patients in ARDT arm who crossover, the [^{177}Lu]Lu-PSMA-617 treatment is considered the next line therapy, therefore data on progressive disease will be collected from the RECIST/PCWG3/PSA/Clinical progression respective CRF pages for the crossover.

PSA50 response is defined as the proportion of patients who have a ≥50% decrease in PSA from baseline (defined as the last value on or before treatment start date). It will be calculated at 12, 24 and 48 weeks from study treatment start and at any time during randomized treatment phase. According to the protocol, PSA will be measured every 6 weeks till discontinuation of treatment and then every 12 weeks during long term FUP. Any PSA measurement included in the above timepoints +/- 2 weeks will be included in the analyses, whether scheduled or unscheduled. If several measurements fall during the same time window, then the closest will be used. If equally distant then the first will be used. Measurements taken after any further antineoplastic therapy (including crossover) will be excluded from these analyses.

Time to SSE (TTSSE) is defined as the time from the date of randomization to the date of the first SSE or death from any cause. SSE date is date of first new symptomatic pathological bone fracture, spinal cord compression, tumor-related orthopedic surgical intervention, requirement for radiation therapy to relieve bone pain (when initiated), or death due to any cause, whichever occurs first. Censoring date is the end of on-treatment period as defined in Section 2.1.1. For non-treated patients, end of on-treatment period is EOT + 30 days. In addition, time to SSE will be analyzed without including death as an event.

Time to soft tissue progression (TTSTP) defined as time from randomization to radiographic soft tissue progression per PCWG3-modified RECIST v1.1 assessed by BICR. The same analyses conventions as depicted in Sections 2.5.3 and 2.5.4 will be used to derive TTSTP, except it will be based on RAS and deaths will not be counted as an event. If the progression leading to crossover was seen only in bone, then TTSTP will be censored at last adequate tumor assessment date before crossover.

Time to chemotherapy (TTCT) defined as time from randomization to initiation of the first subsequent chemotherapy or death, whichever occurs first. Crossover to [^{177}Lu]Lu-PSMA-617 arm will not constitute an event. Participants who did not have an event at the analysis data cut-off or are lost to follow-up at the time of analysis will be censored for TTCT at the time of their last contact.

2.7.2 Statistical hypothesis, model, and method of analysis

rPFS2

The rPFS2 distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curve, medians and 95% confidence intervals of the medians will be presented in the Crossover Analysis Set. The same analyses conventions as for rPFS will be used.

PFS2

PFS2 will be analyzed in the FAS population according to the randomized treatment group and strata assigned at randomization. The PFS2 distribution will be estimated using the Kaplan-

Meier method, and the Kaplan-Meier curves, medians, and 95% confidence intervals of the medians will be presented for each treatment group. The hazard ratio for PFS2 will be calculated, along with its 95% confidence interval estimated using Cox proportional hazard model stratified by the randomization stratification factors.

PFS, TTSSE, TTSTP and TTCT

PFS, TTSSE, TTSTP and TTCT will be analyzed in the FAS population according to the randomized treatment group and strata assigned at randomization. The distributions will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians and 95% confidence intervals of the medians will be presented for each treatment group. The hazard ratio will be calculated, along with its 95% confidence interval, using a stratified Cox model.

Since participants in the ARDT arm are allowed to crossover to [177Lu]Lu-PSMA-617 arm upon confirmation of rPFS by BICR, additional analysis adjusting for the effect of crossover on TTCT will also be performed based on RPSFT.

PSA50 response

PSA response rate at 12, 24 and 48 weeks and at any time during randomized treatment phase along with 95% confidence intervals will be presented by treatment group. Waterfall graphs, which display the best percent change from baseline in maximum decline in PSA for each participant will be used to depict the antitumor activity for each treatment group.

2.7.3 Handling of missing values/censoring/discontinuations

For PFS2, refer to Table 2-10 for censoring and event date options and outcomes.

Table 2-10 Outcome and event/censor dates for PFS2 analysis

Situation	Event/Censoring Date	Outcome
No baseline assessment	Date of randomization	Censored
Progression or death during the next-line therapy	Date of progression (or death)	Event
Death prior to initiation of next-line therapy	Date of death	Event
No progression (or death) during the next-line therapy and no second new anti-neoplastic therapy is initiated	Last contact date	Censored
No progression (or death) during the next-line anticancer therapy and a second new anti-neoplastic therapy is initiated	End date of the next-line therapy	Censored
No next-line therapy initiated with patient known to be alive	Last contact date	Censored

2.8 Safety analyses

All safety analyses related to [^{68}Ga]Ga-PSMA-11 will be based on the Ga-PSMA-11 Full Analysis set (Ga-FAS). Safety analyses of the randomized part of the protocol will be based on the Safety Set. Additionally, safety analyses of all patients that received one dose of [^{177}Lu]Lu-PSMA-617 will be performed in the Lu-PSMA-617 Safety Set.

2.8.1 Adverse events (AEs)

AE summaries will include all AEs occurring during on treatment period. All AEs collected in the AE (e)CRF page will be listed along with the information collected on those AEs e.g. AE relationship to study drug, AE outcome etc. AEs with start date outside of on-treatment period will be flagged in the listings. Three on-treatment periods are defined in this study: the [^{68}Ga]-PSMA part of the protocol, the randomized part of the study and the crossover part of the study (See Section 2.1.1 for definitions of on-treatment periods). The AEs will be flagged according to the on-treatment period it occurred. Note: Adverse events reported related to ^{68}Ga-PSMA-11 are also ^{68}Ga-PSMA-11 TEAEs, irrespective of time of onset or start of randomized treatment.

AEs will be summarized by number and percentage of subjects having at least one AE, having at least one AE in each primary system organ class (SOC) and for each preferred term (PT) using MedDRA coding. A subject with multiple occurrences of an AE will be counted only once in the respective AE category. A subject with multiple CTCAE grades for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event. AE with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

In AE summaries, the primary system organ class will be presented alphabetically, and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for the preferred term will be based on their frequency in the investigational arm.

The number (and percentage) of participants with treatment emergent adverse events (events started after the first dose of study medication or events present prior to start of treatment but increased in severity based on preferred term) will be summarized in the following ways:

- by treatment, primary system organ class and preferred term.
- by treatment, primary system organ class, preferred term and maximum severity.
- by treatment, Standardized MedDRA Query (SMQ) and preferred term.

Separate summaries will be provided for study treatment related adverse events, death, serious adverse events, other significant adverse events leading to discontinuation, and adverse events leading to dose adjustment. The study treatments for this study are [^{177}Lu]Lu-PSMA-617or ARDT.

To account for the possible difference in exposure observed between the two treatment arms, incidence rates of adverse events by preferred term will be presented as adjusted for number of Subject Treatment Years (STY). The adjusted rate for a given AE is calculated as number of events per 100 STY (=[n/STY]*100), where STY is defined as the sum of subject-years at risk

for all subjects within the treatment group. For each subject the exposure time at risk is calculated as follows:

- (Start date of the first event – treatment start date + 1)/365.25, if the subject experienced the event

Or

- (End of the treatment-emergent period – treatment start date + 1)/365.25, if the subject did not experience the event

This analysis will be done for the first occurence of any grade, the first occurence of grade ≥3, the first occurence of SAE, occurence of AE leading to discontinuation and first occurence of safety topic of interest (STI).

Adverse events which will be counted for a specific treatment period are those which are treatment-emergent. These events are those with an onset after the start of the treatment period, or which were present prior to the start of the treatment period but increased in severity, changed from being not suspected to being suspected of study drug relationship, or developed into SAEs after the start of the treatment period and before the start of next treatment period (See Section 2.1.1 for definitions of on-treatment periods).

The above analyses will be carried out in the Safety Set, and key safety analyses will be performed in the Lu-PSMA-617 Safety Set and Ga-FAS.

During long-term follow-up, serious adverse events related to [^{177}Lu]Lu-PSMA-617 will be collected. Summary tables will be provided in the Lu-PSMA-617 Safety Set.
To help evaluate the impact of the COVID-19 on the safety, the incidence of COVID-19 related adverse event preferred terms will be presented. All COVID-related AEs will be included in the listings.

For the legal requirements of ClinicalTrials.gov and EudraCT, two required tables on treatment emergent adverse events which are not serious adverse events with an incidence greater than 5% and on treatment emergent serious adverse events and SAE suspected to be related to study treatment will be provided by system organ class and preferred term on the Safety Set population.

If for a same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

1. a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE
2. more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment / non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.

2.8.1.1 Safety topic of interest / grouping of AEs

A safety topic of interest (STI) is a grouping of adverse events that are of scientific and medical concern specific to compound [^{177}Lu]Lu-PSMA-617. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HGLTs (high level group terms), HLT (high level terms) and PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad. STI will be defined at the project level and may be regularly updated. The grouping of AEs in STI according to project standards will be specified in the Case-Retrieval Sheet, For each specified STI, number and percentage of patients with at least one event of the STI occurring during on treatment period will be summarized.

Summaries of these STIs will be provided by treatment arm (when applicable), (specifying grade, SAE, relationship, leading to treatment discontinuation, leading to dose adjustment/interruption, hospitalization, death etc.).

Time to first STI occurrence

Time to first occurrence of an event is defined as time from start of study treatment to the date of first occurrence of this event (or first event within an AE grouping), i.e. time in days is calculated as (start date of first occurrence of event) – (start of study treatment) +1.

In the absence of an event during the on-treatment period, the censoring date applied will be **the earliest** of the following dates:

- death date
- new anticancer antineoplastic therapy start date (including start of Lu-PSMA for ARDT patients crossing over),
- end date of on-treatment period
- data cut-off date
- withdrawal of informed consent date.

The corresponding censoring reason will be used: death, new anti cancer therapy, treatment discontinuation, ongoing at cut-off date or consent withdrawal.

Failure curves (ascending Kaplan-Meier curves) will be constructed by treatment arm. Median together with 95% confidence interval as well as 25th percentile and 75th percentile will be presented for each treatment arm.

In addition, the median time to occurrence for the subset of subjects who experienced the event of interest will be calculated. Simple descriptive statistics, median, min and max as well as 25th percentile and 75th percentile, will be presented.

The same analysis will be repeated for events of grade ≥3.

The above analyses will be carried out in the Safety Set and Lu-PSMA-617 Safety Set.

A listing of all grouping levels down to the MedDRA preferred terms used to define each STI will be generated.

2.8.2 Deaths

Separate summaries for on-treatment and all deaths will be produced by treatment arm, system organ class and preferred term. On-treatment deaths analyses will be carried out in the Safety Set, Lu-PSMA-617 Safety Set and Ga-FAS. All deaths analyses will be carried out in the Safety Set only. In the control arm, analyses will make the distinction between long-term deaths that occurred in patients who crossed-over and patients who did not.

All deaths will be listed, post treatment deaths will be flagged. A separate listing of deaths prior to starting treatment will be provided for all screened subjects.

2.8.3 Laboratory data

On analyzing laboratory, data from all sources (central and local laboratories) will be combined. The summaries will include all assessments available for the lab parameter collected during the on-treatment period after the last study treatment administration date (see Section 2.1.1).

The analyses will be carried out in the Safety Set and key analyses will be carried out for the Lu-PSMA-617 Safety Set. No laboratory analyses are planned for the [^{68}Ga]Ga-PSMA-11on-treatment period.

Grading of laboratory values will be assigned programmatically as per NCI CTCAE version 5.0. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account.

CTCAE Grade 0 will be assigned for all non-missing values not graded as 1 or higher. Grade 5 will not be used.

The following summaries will be produced for hematology and biochemistry laboratory data (by laboratory parameter and treatment):

- Worst post-baseline CTC grade (regardless of the baseline status). Each subject will be counted only for the worst grade observed post-baseline.
- Shift tables using CTC grades to compare baseline to the worst on-treatment value
- For laboratory tests where CTC grades are not defined, shift tables using the low/normal/high/(low and high) classification to compare baseline to the worst on-treatment value.
- Trends of lab parameter values over time (baseline and selected on-treatment timepoints) will be displayed via boxplots based on visit number and corresponding tables displaying the statistics used for the box plots by the selected time points. The following parameters will be displayed: platelet count, ANC, lymphocytes, hemoglobin, eGFR, blood creatinine, AST, ALT.The following listings will be produced for the laboratory data:

- Listings of all laboratory data, with CTC grades and classification relative to the laboratory normal range. Lab data collected during the post-treatment period will be flagged.
- Listing of all CTC grade 3 or 4 laboratory toxicities

Analysis of time to first occurrence of eGFR worsening will be applied in the Safety Set. eGFR will be categorized as normal eGFR (≥ 90 mL/min), mild impairment (eGFR= 60-<90 mL/min), moderate impairment (eGFR 30-<60 ml/min) and severe impairment (eGFR <30 ml/min). A worsening is defined as the first time a patient changes to a worse category (e.g. from mild to moderate). eGFR is collected at screening and EOT only, therefore the above analyses will be performed using the MDRD formula (including at screening) as described below:

eGFR = 175 x $(S_{Cr})^{-1.154}$ x $(age)^{-0.203}$ x 0.742 [if Female] x 1.212 [if Black], where S_{Cr} is serum creatinine in mg/dL and age is expressed in years.

Time to first eGFR worsening occurrence

Time to first occurrence of an event is defined as time from start of study treatment to the date of first occurrence of this event, i.e. time in days is calculated as (start date of first occurrence of event) – (start of study treatment) +1.

In the absence of an event during the on-treatment period, the censoring date applied will be date of last laboratory assessment on treatment.

The corresponding censoring reason will be used: death, new anti cancer therapy, treatment discontinuation, ongoing at cut-off date or consent withdrawal.

Failure curves (ascending Kaplan-Meier curves) will be constructed by treatment arm. Median together with 95% confidence interval as well as 25th percentile and 75th percentile will be presented for each treatment arm.

In addition, the median time to occurrence for the subset of subjects who experienced the event of interest will be calculated. Simple descriptive statistics, median, min and max as well as 25th percentile and 75th percentile, will be presented.

Time to first occurrence of laboratory values

Time to first occurrence of grade 3/4 will be analyzed following the same method for the following parameters: platelet count, ANC, lymphocytes, hemoglobin, blood creatinine, AST, ALT. Theses analyses will include only patients that had a lower grade at baseline than the worsening being looked at. For time to first occurrence of grade 3/4, only patients with grade <3 or missing will be included.

Liver function parameters

Liver function parameters of interest are total bilirubin (TBL), ALT, AST and alkaline phosphatase (ALP). The number (%) of patients with worst post-baseline values as per Novartis Liver Toxicity guidelines will be summarized:

The following summaries will be produced:

- ALT or AST > 3xULN
- ALT or AST > 5xULN
- ALT or AST > 8xULN
- ALT or AST > 10xULN
- ALT or AST > 20xULN
- TBL > 2xULN
- TBL > 3xULN
- For patients with AST and ALT ≤ ULN at baseline
    - ALT or AST > 3xULN & TBL > 2xULN
    - ALT or AST > 3xULN & TBL > 2xULN & ALP ≥ 2xULN
    - ALT or AST > 3xULN & TBL > 2xULN & ALP < 2xULN
- For patients with AST or ALT > ULN at baseline
    - Elevated ALT or AST (*) & TBL (>2x Bsl and 2x ULN)
    - Elevated ALT or AST (*) & TBL (>2x Bsl and 2x ULN) & ALP >=2x ULN
    - Elevated ALT or AST (*) & TBL (>2x Bsl and 2x ULN) & ALP <2x

* Elevated AST or ALT defined as: >3x ULN if =< ULN at baseline, or (>3x Bsl or 8x ULN) if > ULN at baseline

2.8.4 Other safety data

2.8.4.1 ECG data

Data handling

In this study single ECGs will be collected. However, for patients with safety concerns, triplicate ECGs are required. In case of multiple replicates at any assessment, the average of the ECG parameters at that assessment will be used in the analyses. The scheduled timepoints for ECG collection are screening and EOT, and then as clinically indicated during the course of treatment.

Data analysis

ECG analyses will be performed in the Safety Set.

The number and percentage of subjects with notable ECG values will be presented by treatment arm.

- QT, QTcF, or QTcB (presented separately)
    - New value of > 450 and ≤ 480 ms

  - New value of > 480 and ≤ 500 ms
  - New value of > 500 ms
  - Increase from Baseline of > 30 ms to ≤ 60ms
  - Increase from Baseline of > 60 ms
- HR
  - Increase from baseline >25% and to a value > 100 bpm
  - Decrease from baseline >25% and to a value < 50 bpm
- PR
  - Increase from baseline >25% and to a value > 200 ms
  - New value of > 200 ms
- QRS
  - Increase from baseline >25% and to a value > 120 ms
  - New values of QRS > 120 ms

A listing of all ECG assessments will be produced by treatment arm and notable values will be flagged. In the listing, the assessments collected during the post-treatment period will be flagged.

2.8.4.2 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The following parameters were collected: height (cm), weight (kg), body temperature (°C), heart rate (beats per minute), systolic and diastolic blood pressure (mmHg).

Vital signs analyses will be performed in the Safety Set and Lu-PSMA-617 Safety Set.

Data handling

Vital signs collected on treatment will be summarized. Values measured outside of on treatment period will be flagged in the listings.

Data analysis

For analysis of vital signs the clinically notable vital sign criteria are provided in Table 2-11 below.

Table 2-11 Clinically notable changes in vital signs

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Weight (kg)	increase >= 10% from Baseline	decrease >=10% from Baseline
Systolic blood pressure (mmHg)	>=180 with increase from baseline of >=20	<=90 with decrease from baseline of >=20

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Diastolic blood pressure (mmHg)	>=105 with increase from baseline of >=15	<=50 with decrease from baseline of >=15
Pulse rate (bpm)	>=100 with increase from baseline of >25%	<=50 with decrease from baseline of > 25%
Body temperature	>= 39.1	

The number and percentage of subjects with notable vital sign values (high/low) will be presented by treatment arm.

A listing of all vital sign assessments will be produced by treatment arm and notable values will be flagged. In the listing, the assessments collected outside of on-treatment period will be flagged.

2.9 Pharmacokinetic endpoints

Not Applicable

2.10 PD and PK/PD analyses

Not Applicable

2.11 Patient-reported outcomes

Three participant-reported outcomes (PRO) questionnaires will be assessed in the study: EQ-5D-5L, FACT-P and BPI-SF. The FAS will be used for analyzing PRO data. Note that ePRO questionnaires are not provided during screening but at Cycle Day 1 before the treatment start. Therefore the baseline for ePRO analyses will be the last available assessment before treatment start, despite the analyses will use the FAS and randomization date as a reference date for time to event analyses.

Utilities derived from EQ-5D-5L together with FACT-P and BPI-SF are the PRO variables of interest.

Aspects of HRQoL will be self-reported by participants using the EQ-5D-5L and FACT-P questionnaires, and pain will be assessed by participants using the BPI-SF.

No formal statistical tests will be performed for PRO data and hence no multiplicity adjustment will be applied.

2.11.1 EQ-5 Dimension-5 Level (EQ-5D-5L) Questionnaire

The EQ-5D-5L is shown in [protocol Appendix 16.4]. The higher the EQ-VAS score, the better the QoL. The higher the EQ-5D items, the worse the QoL.

The EQ-5D-5L consists of the EQ-5D descriptive system and the EQ visual analogue scale (EQ-VAS). The EQ-5D-5L descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The EQ-VAS records the respondent's self-rated health on a vertical, visual analogue scale.

Each of the five dimension scales contain five levels, with level 1 indicating no problems, level 2 indicating slight problems, level 3 indicating moderate problems, level 4 indicating severe problems, and level 5 indicating unable to/extreme problems.

The EQ-VAS is scored by assigning an integer value, ranging from 0 (Worst imaginable health state) to 100 (Best imaginable health state), corresponding to the mark placed by the patient on the VAS. Ambiguous answers (e.g., two marks placed on the scale by a patient) should be treated as missing values.

Table 2-12 EQ-5D-5L scales and subscales

Scale name	Number of items	Item range*
Descriptive System Dimension		
Mobility	1	1-5
Self-Care	1	1-5
Usual Activities	1	1-5
Pain/Discomfort	1	1-5
Anxiety/Depression	1	1-5
Health State Evaluation		
EQ-VAS*	1	0-100

* EQ-VAS is a continuous visual analog scale, with integer scores ranging from 0 to 100.

Minimally important difference is 0.08 for both 1) change from baseline for within subject change and 2) between group differences for treatment comparisons Pickard et al 2007.

A utility score will be obtained by using a weighted combination of the levels of the five dimension scales. The weights are based on value sets which are country-specific. The country specific code for the U.K. will be used for all sites in this study since the health economics modeling will target the U.K. population for developing the core economic model. Each patient's 5 digit health states code (response to question 1,2,3,4, and 5 concatenated (ex., 41325 results in a utility score of 0.193)) is converted to a utility score using the EQ-5D-5L value set, available in the cross-walk index value calculator which can be downloaded from the web site.

Since utility score depends on the combination of all items' responses, any missing response results in a missing utility score. In the U.K. value set, utility scores ranges from the lowest possible score for a patient of -0.594 (when all responses are '5') to 1 (when all responses are '1').

Analyses

Time to worsening for utility score is defined as the time (in months) from randomization to the first occurrence of worsening in score relative to baseline (any decrease of at least 0.08), clinical disease progression (excluding radiographic and PSA progression), or death, whichever is earlier. If no event is experienced, the censoring date will be time of the last QoL assessment. This analysis will include only assessments before the crossover, if any.

Additionally, time to worsening defined as the time (in months) from randomization to the first occurrence of worsening in score relative to baseline (any decrease of at least 0.08) or death, whichever is earlier will also be analyzed. If no event is experienced, the censoring date will be date of the last QoL assessment. This analysis will include only assessments before the crossover, if any.

Time to worsening will be estimated using the Kaplan-Meier method. The median time along with 95% confidence intervals will be presented by treatment group. A stratified Cox regression model will be used to estimate the hazard ratio of time to definitive deterioration, along with 95% confidence interval (using randomization stratification).

Descriptive statistics will be used to summarize the original scores, as well as change from baseline, of the EQ-5D-5L at each scheduled assessment timepoint for each treatment group. Participants with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

In addition, a repeated measurement analysis model for longitudinal data will be used to estimate differences in utility scores of the EQ-5D-5L between treatment arms. Baseline value (as a continuous variable), treatment arm, time (as visit number), treatment by time interaction, baseline by time interactions, and randomization stratification factors as main effects will be the fixed factors. A general (unstructured) variance-covariance matrix for the repeated measures for a single patient will be used. In case of problems with fitting the model, as an alternative, a heterogeneous Toeplitz and AR(1) structures will be considered to reduce the number of the parameters of the model. The differences in least square means between the treatment arms and corresponding 95% confidence interval at selected timepoints will be presented.

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

The higher the FACT-P score (for all subscales and total scales), the better the QoL.

Table 2-13 FACT-P scales and subscales

Scale/Sub-scale Name	Number of Items	Scale Range	FACT-P Item numbers	Threshold for worsening*
Subscale				
Physical Well-Being (PWB)	7 items	0-28	GP1-GP7	3

Scale/Sub-scale Name	Number of Items	Scale Range	FACT-P Item numbers	Threshold for worsening*
Social/Family Well-Being (SFWB)	7 items	0-28	GS1-GS7	3
Emotional Well-Being (EWB)	6 items	0-24	GE1-GE6	3
Functional Well-Being (FWB)	7 items	0-28	GF1-GF7	3
Prostate Cancer Subscale (PCS)	12 items	0-48	All items in "Additional Concerns" section	3
PCS pain-related subscale (PRS)**	4 items	0-16	P1, P2, P3, GP4	2
FACT Advanced Prostate Symptom Index-8 (FAPSI-8)***	8 items	0-32	GP1, GP4, GE6, C2, P2, P3, P7, P8	3
Total Scale				
Trial Outcomes Index (TOI) score	3 subscales	0-104	PWB, FWB, PCS	9
FACT-G (General)	4 subscales	0-108	PWB, SFWB, EWB, FWB	9
FACT-P Total	39 items	0-156	All	10

*Minimally important difference for both 1) decrease from baseline for within subject change and 2) between group differences for treatment comparisons.

** Calculated using the 4 questions on pain in the FACT-P, but the scores are reversed

such that higher score indicates better health and less pain. A decrease in score signifies pain progression Cella et al. 2009)

*** Symptom index of important clinician-rated symptoms/concerns to monitor when assessing value of treatment for advanced prostate cancer (FAPSI-8; Yount et al. 2003)

Scoring of FACT-P subscales and total scores (Trial Outcome Index (TOI), FACT-G Total Score (G for general), and FACT-P Total Score (P for prostate)) are shown below. These are from the FACT-P Scoring Guidelines (Version 4). Item codes in scoring guidelines are shown on the FACT-P form in [Appendix 16.5 of the protocol].

1. Record answers in "item response" column. If missing, mark with an X.
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.

4. Add subscale scores to derive total scores (TOI, FACT-G, FACT-P).
5. Handling missing items. The FACT scale is considered to be an acceptable indicator of patient quality of life as long as **overall item response rate** is greater than 80% (e.g., at least 32 of 39 FACT-P items completed). In addition, a total score should only be calculated if ALL of the component subscales have valid scores. For subscales, as long as **more than** 50% of the items were answered (e.g., a minimum of 4 of 7 items, 4 of 6 items, etc), prorate the subscale score by following the scoring instructions below, producing an observed sum weighted by the inverse of the proportion of observed items.

Subscale	Item Code	Reverse item?		Item response		Item Score	
Physical Well-Being (PWB)	GP1	4	-		=		
	GP2	4	-		=		
	GP3	4	-		=		
Score range:	GP4	4	-		=		
0-28	GP5	4	-		=		
	GP6	4	-		=		
	GP7	4	-		=		
				Sum individual item scores:			
				Multiply by 7:			
				Divide by number of items answered:			= PWB subscale score
Social/Family Well-Being (SFWB)	GS1	0	+		=		
	GS2	0	+		=		
	GS3	0	+		=		
Score range:	GS4	0	+		=		
0-28	GS5	0	+		=		
	GS6	0	+		=		
	GS7	0	+		=		
				Sum individual item scores:			
				Multiply by 7:			
				Divide by number of items answered:			= SFWB subscale score
Emotional Well-Being (EWB)	GE1	4	-		=		
	GE2	0	+		=		
	GE3	4	-		=		
Score range:	GE4	4	-		=		
0-24	GE5	4	-		=		
	GE6	4	-		=		
				Sum individual item scores:			
				Multiply by 6:			
				Divide by number of items answered:			= EWB subscale score
Functional Well-Being (FWB)	GF1	0	+		=		
	GF2	0	+		=		
	GF3	0	+		=		
Score range:	GF4	0	+		=		
0-28	GF5	0	+		=		
	GF6	0	+		=		
	GF7	0	+		=		
				Sum individual item scores:			
				Multiply by 7:			
				Divide by number of items answered:			= FWB subscale score

Subscale	Item Code	Reverse item?		Item response		Item Score
Prostate Cancer Subscale (PCS)	C2	4	-		=	
	C6	0	+		=	
	P1	4	-		=	
Score range: 0-48	P2	4	-		=	
	P3	4	-		=	
	P4	0	+		=	
	P5	0	+		=	
	P6	4	-		=	
	P7	4	-		=	
	BL2	4	-		=	
	P8	4	-		=	
	BL5	0	+			
				Sum individual item scores:		
				Multiply by 12:		
				Divide by number of items answered:		= PC Subscale score

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

_____ + _____ + _____ = _____ = FACT-P TOI
(PWB score) *(FWB score)* *(PCS score)*

To derive a FACT-G Total score:
Score range: 0-108

_____ + _____ + _____ + _____ = _____ = FACT-G Total score
(PWB score) *(SFWB score)* *(EWB score)* *(FWB score)*

To derive a FACT-P Total score:
Score range: 0-156

_____ + _____ + _____ + _____ + _____ = __ = FACT-P Total score
(PWB score) *(SFWB score)* *(EWB score)* (FWB *score*) (PCS *score*)

To derive Pain-related subscale (PRS):

Subscale	Item Code	Reverse item?		Item response		Item Score
Pain-related subscale	P1	4	-		=	
	P2	4	-		=	
	P3	4	-		=	
Score range: 0-16	GP4	4	-		=	
				Sum individual item scores:		
				Multiply by 4:		
				Divide by number of items answered:		= PR Subscale score

To derive FACT Advanced Prostate Symptom Index-8 (FAPSI-8):

Using the 8 items GP1, GP4, GE6, C2, P2, P3, P7, P8, do the following:

Reverse code individual items as needed following guidelines in scores above.

Sum individual item scores.

Multiply by 8.

Divide by the number of items answered.

Analyses

Time to worsening for FACT-P, FAPSI-8, TOI, and FACT-G is defined as the time (in months) from randomization to the first occurrence of worsening in score relative to baseline (as shown in Table 2-13), clinical disease progression , or death, whichever is earlier. If no event is experienced, the censoring date will be time of the last QoL assessment. This analysis will include only assessments before the crossover, if any.

Additionally, time to worsening defined as the time (in months) from randomization to the first occurrence of worsening in score relative to baseline or death, whichever is earlier will also be analyzed (i.e. not counting clinical progressions as events). If no event is experienced, the censoring date will be date of the last QoL assessment. This analysis will include only assessments before the crossover, if any.

Time to worsening will be estimated using the Kaplan-Meier method. The median time along with 95% confidence intervals will be presented by treatment group. A stratified Cox regression model will be used to estimate the hazard ratio of time to definitive deterioration, along with 95% confidence interval (using randomization stratification).

Descriptive statistics will be used to summarize the original scores, as well as change from baseline, of the scales and subscales at each scheduled assessment timepoint for each treatment group. Participants with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

In addition, a repeated measurement analysis model for longitudinal data will be used to estimate differences in scores of the FACT-P, FAPSI-8, TOI, and FACT-G between treatment arms. Baseline value (as a continuous variable), treatment arm, time (as visit number), treatment by time interaction, baseline by time interactions, and randomization stratification factors as

main effects will be the fixed factors. A general (unstructured) variance-covariance matrix for the repeated measures for a single patient will be used. In case of problems with fitting the model, as an alternative, a heterogeneous Toeplitz and AR(1) structures will be considered to reduce the number of the parameters of the model. The differences in least square means between the treatment arms and corresponding 95% confidence interval at selected timepoints will be presented.

2.11.3 Brief Pain Inventory – Short Form (BPI-SF)

The Brief Pain Inventory - Short Form (BPI-SF) is shown in protocol Appendix 8. The higher the BPI-SF score, the worse the pain.

Description and Scoring

The BPI-SF consists of 4 questions regarding pain intensity, 2 questions on the use of analgesics, and 7 questions on how the level pain has interfered with the subject's life. Intensity items consist of an 11-response rating scale scored from 0 ("No Pain") to 10 ("Pain As Bad As You Can Imagine"). Interference items consist of scores from 0 ("Does Not Interfere") to 10 ("Completely Interferes").

Table 2-14 BPI-SF scales

Scale Name	Number of Items	Scale Range	BPI-SF Item numbers	Threshold for worsening*
Individual Item Scales				
Worst pain intensity	1	0-10	3	Either of ≥30% of baseline or ≥2-point increase
Least pain intensity	1	0-10	4	
Average pain intensity	1	0-10	5	
Pain intensity right now	1	0-10	6	
Summary Scales				
Pain Intensity Scale	4	0-10	3-6	≥30% of baseline or ≥2-point increase
Pain Interference Scale	7	0-10	9a-9g	≥30% of baseline or ≥2-point increase

*Minimally important difference for both 1) increase from baseline for within subject change and 2) between group differences for treatment comparisons.

BPI-SF Intensity is the mean of non-missing items of the 4 items in the table above, if there are 3 or more items not missing; otherwise this scale is set to missing.

BPI-SF Interference scale is the mean of non-missing items of the 7 items in the table above, if there are 4 or more items not missing; otherwise this scale is set to missing.

Analyses

Time to worsening of Worst Pain Intensity (item 3), also called Time to Disease Related Pain (TDRP), Time to worsening of Pain Intensity Scale, and Time to worsening of Pain Interference Scale are defined as the time (in months) from randomization to the first occurring of 1) an increase of worsening threshold (as shown in Table 2-14) compared to baseline, 2) clinical disease progression (excluding radiographic and PSA progression), or 3) death. If no event is experienced, the censoring date will be time of the last BPI-SF assessment. This analysis will include only assessments before the crossover, if any.

In addition, Time to Disease Related Pain (TDRP), Time to worsening of Pain Intensity Scale, and Time to worsening of Pain Interference Scale defined as the time (in months) from randomization to the first occurrence of an increase of worsening threshold relative to baseline or death, whichever is earlier will also be analyzed (i.e. not counting clinical progressions as events). If no event is experienced, the censoring date will be date of the last BPI-SF assessment. This analysis will include only assessments before the crossover, if any.

Time to Disease Related Pain (TDRP), Time to worsening of Pain Intensity Scale, and Time to worsening of Pain Interference will be estimated using the Kaplan-Meier method. The median time along with 95% confidence intervals will be presented by treatment group. A stratified Cox regression model will be used to estimate the hazard ratio of time to definitive deterioration, along with 95% confidence interval (using randomization stratification).

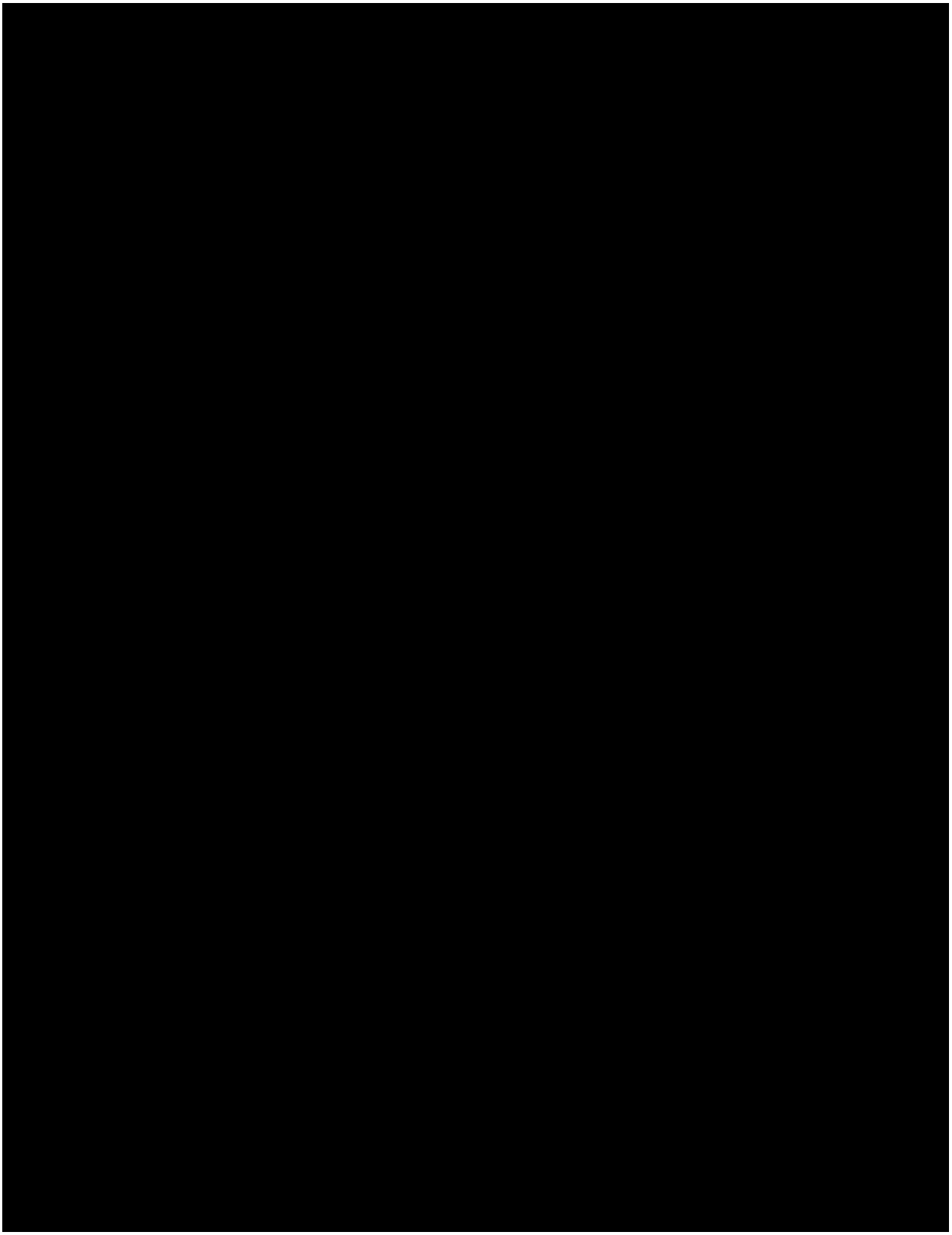
Additionally, the time to improvement following initial pain worsening in Pain Intensity Scale and Pain Interference Scale will be analyzed using mixture distribution methodology described by Ellis et al 2008. Time to pain improvement is defined as time from worsening of intensity or interference to a score $\leq$ baseline.

Descriptive statistics will be used to summarize the original scores, as well as change from baseline, of the BPI-SF at each scheduled assessment timepoint for each treatment group. Participants with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

In addition, a repeated measurement analysis model for longitudinal data will be used to estimate differences in scores of the BPI-SF between treatment arms. Baseline value (as a continuous variable), treatment arm, time (as visit number), treatment by time interaction, baseline by time interactions, and randomization stratification factors as main effects will be the fixed factors. A general (unstructured) variance-covariance matrix for the repeated measures for a single patient will be used. In case of problems with fitting the model, as an alternative, a heterogeneous Toeplitz and AR(1) structures will be considered to reduce the number of the parameters of the model. The differences in least square means between the treatment arms and corresponding 95% confidence interval at selected timepoints will be presented.

[redacted]

[redacted]



2.14 Interim analysis

No interim is planned for rPFS. The final rPFS analysis will only be carried out after all participants have been randomized.

Key secondary endpoint: Overall survival (OS)

A hierarchical testing procedure will be adopted and the statistical tests for OS will be performed only if the primary efficacy endpoint rPFS is statistically significant.

A maximum of four analyses (three interims and one final) is planned for OS: The first interim OS was performed at the time of the primary rPFS analysis, which was expected to be after

approximately 74 deaths (25% information fraction) were observed. The second interim analysis will be performed when ~ 125 of the ~ 297 targeted OS events (~42% information fraction) are observed. The second interim analysis is expected to occur around 24 months from the date of first participant randomized in the study. At the time of the third interim analysis (at approximately 75% information fraction) approximately 223 deaths are expected. The final analysis for OS is planned when approximately 297 deaths are observed (approximately 57 months from date of first participant to be randomized). The third interim analysis is expected to occur around 38 months from the date of first participant randomized in the study.

An α-spending function according to Lan et al 1983) (O'Brien-Fleming) as implemented in East 6.4, along with the testing strategy outlined below will be used to maintain the overall type I error probability. This guarantees the protection of the 2.5% overall level of significance across the two hypotheses and the repeated testing of the OS hypotheses in the interim and the final analyses (Glimm et al 2010).

The trial allows for the stopping of the study for a superior OS result, provided the primary endpoint rPFS has already been shown to be statistically significant favoring the test treatment arm. Further, the exact nominal p-values that will need to be observed to declare statistical significance at the time of these analyses for OS will depend on the number of OS events that have been observed at the time of these analyses and the α already spent for OS at the time of earlier analyses.

An α-spending function according to a four-look (Lan et al 1983) group sequential design with (O'Brien-Fleming) type stopping boundary (as implemented in East 6.4) will be used to construct the efficacy stopping boundaries. Based on the choice of α-spending function described above and if the interim analyses are performed exactly after the planned OS events, the efficacy boundaries in terms of p-value scale (or equivalently Z-statistic scale) at the interim are calculated as p=0.000000001, p=0.00055 and p = 0.01 (or Z = -5.996, Z = -3.263 and Z = -2.345). The observed (i.e. nominal) p-value has to be smaller than 0.000000001, 0.00055 and 0.01 (or equivalently the observed Z-statistic has to be < Z-statistic scale boundary = -4.34 and -2.338) to conclude superior efficacy at the interim analyses. Since the observed number of events at the interim analyses may not be exactly equal to the planned OS events, the efficacy boundary will need to be recalculated using the pre-specified α-spending function and based on the actual number of observed events at interim and the total number of targeted events to calculate the exact information fraction. The observed p-value (or Z-test statistic) at the interim analysis will then be compared against the re-calculated efficacy boundary.

If the study continues to the final OS analysis, the final OS analysis will be performed when approximately 297 OS events have been observed. In practice, the final analysis will be based on the actual number of OS events observed at the cut-off date for the final OS analysis and α already spent at the interim analyses. The boundary for the final analysis will be derived accordingly from the pre-specified α-spending function such that the overall significance level across all analyses is maintained at 0.025.

2.15 Other analyses related to delayed dosing due to drug supply challenges of [^{177}Lu]Lu-PSMA-617

Interruptions of scheduled dosing due to study drug supply challenges may have different effects on individual participants randomized to [^{177}Lu]Lu-PSMA 617 arm depending on when they entered the trial. The trial experienced numerous interruptions of study drug supply, primarily from a global temporary production halt. However, this analysis may also include any sporadic dosing issues of patients impacted by individual batch failures or other production issues.

2.15.1 Impact on delay of [^{177}Lu]Lu-PSMA-617 administration and protocol deviations

To evaluate the impact of an interruption of study drug supply, and assess how many study participants are affected, the study team will review the number of patients with [^{177}Lu]Lu-PSMA-617 infusions outside of the study protocol pre-specified windows, i.e:

- Administration delayed by 17 days after randomisation at treatment initiation (C1D1)
- Administration delayed by 1 week to 4 weeks (C2-C6)
- Study treatment delayed by more than 4 weeks: (C2 –C6)

The risk of delayed study treatment due to large scale interruption of study drug supply was investigated to assess the potential impact on the expected treatment effect. The study team took proactive operational measures to minimize infusion delays and reduce impact on the overall integrity of the trial.

Resuming [^{177}Lu]Lu-PSMA-617 infusion after >4 weeks delay (>10 weeks from last infusion) and initiating [^{177}Lu]Lu-PSMA-617 infusion >17 days after randomization was not allowed as per Protocol Amendment v01. If either scenario happens for a number of participants due to an interruption of study drug supply, it will be recorded as an important protocol deviation up to release of Protocol Amendment v02 (where retreatment after >4 week delay was permitted if interruption due to drug supply issues). The number of participants with delayed treatment administration and the number of protocol deviations due to an interruption of study drug supply will be summarized by appropriate descriptive statistics (i.e. mean, proportion, standard deviation, median, minimum, and maximum).

2.15.2 Impact on primary estimand

We will examine the impact of an interruption of study drug supply on estimand attributes and highlight important considerations related to the interruption of study drug supply.

Strategies for handling intercurrent events not related to an interruption of study drug supply will remain unchanged.

The major interccurents event related to the interruption of study drug supply are

1. At least 4 week delay of [^{177}Lu]Lu-PSMA-617 administration in cycles 2-6

2. At least 18 days (from randomization) delay of [^{177}Lu]Lu-PSMA-617 administration in cycle 1.

For the purpose of the primary estimand, the delays due to an interruption of study drug supply are considered irrelevant in defining the treatment effect (treatment policy strategy). Under the treatment policy strategy, the estimated treatment effect reflects the effects of treatment regardless of delay of [^{177}Lu]Lu-PSMA-617 administration due to an interruption of study drug supply.

To target the treatment effect that is not confounded by the delays related to an interruption of study drug supply, a supplementary analysis will be defined.

As a supplementary estimand, a hypothetical strategy is also considered to handle this intercurrent events (delayed [^{177}Lu]Lu-PSMA-617 administration due to an interruption of study drug supply).

Treatment effect as if a production-related intercurrent event did not occur may be of interest. For participants with delayed treatment due to an interruption of study drug supply, the hypothetical scenario is that they would continue in the study in the same way as similar participants with an undelayed access to treatment. These participants would be censored at the last adequate tumor assessment before the 4-week delay mark or at randomization for delayed start of treatment.

The same analysis convention and summary measure as the primary estimand will be used for the supplementary estimand.

Furthermore, the assumption of proportional hazards (PH) will be investigated to ensure external validity and interpretability of the summary measure in terms of hazard ratio (HR). If the PH assumption is violated, supportive estimands with alternative summary measures will be considered (e.g, using a time-varying covariate-adjusted Cox model accounting for the interruption of study drug supply).

2.15.3 Impact on secondary estimand

The impact of an interruption of study drug supply on the secondary estimand is expected to be limited and there will be no change to the planned analysis.

2.15.4 Impact on safety

The interruption of study drug supply-related delay group is defined as "> 4 week delay in cycles 2-6 or >17 days delay in cycle 1" or "no major delay due to interruption of study drug supply" and participants in the [^{177}Lu]Lu-PSMA-617 arm would be classified into one of the two groups.

Overview of AEs for the [^{177}Lu]Lu-PSMA-617 arm by interruption of study drug supply - related delay group will be provided.

3 Sample size calculation

The sample size calculation is based on the primary variable rPFS and key secondary variable OS.

Overall, a total study sample size of approximately 450 participants will be randomly assigned to each treatment arm in a 1:1 ratio (225 participants in [^{177}Lu]Lu-PSMA-617 with BSC arm and 225 participants in ARDT with BSC arm). These calculations were made using the software package East 6.4.

3.1 Primary endpoint(s)

The sample size calculation is based on the primary variable rPFS. The hypotheses to be tested and details of the testing strategy are described in Section 2.5.2.

Based on available data, the median rPFS in the control arm is expected to be around 6 months in this population months (de Bono et al 2020, de Wit et al 2019, Komura et al 2019). It is expected that treatment with ^{177}Lu-PSMA-617 and BSC will result in a 44% reduction in the hazard rate for rPFS compared to the control arm i.e., an expected hazard ratio of 0.56, which corresponds to an increase in median rPFS by 4.7 months under the exponential model assumption (from 6 to 10.7 months).

In order to ensure at least 95% power to test any null hypothesis: rPFS hazard ratio = 1, versus the alternative hypothesis: rPFS hazard ratio = 0.56, it is calculated that a total of 156 rPFS events need to be observed. This calculation assumes analysis by a one-sided log-rank test at the overall 2.5% level of significance, participants randomized to the two treatment arms in a 1:1 ratio.

Assuming that enrolment will continue for 10.5 months at a non-uniform rate for a total of 450 patients and losses to follow-up for rPFS of approximately 10% and 20% in the [^{177}Lu]Lu-PSMA-617 and BSC and control arm, respectively, the targeted 156 rPFS events are expected to occur at about 3 months after the randomization date of the last participant.

3.2 Secondary endpoint(s)

OS, as the key secondary variable, will be formally statistically tested, provided that the primary variable rPFS is statistically significant.

The median OS in the control arm is expected to be around 18 months (de Bono et al 2020, de Wit et al 2019, Komura et al 2019). It is hypothesized that treatment with [^{177}Lu]Lu-PSMA-617 will result in a 28% reduction in the hazard rate for OS compared to the control arm i.e., an expected hazard ratio of 0.72, which corresponds to an increase in median OS by 7 months under the exponential model assumption (from 18 months to 25 months).

Then in order to ensure 80% power to test the null hypothesis: OS hazard ratio = 1, versus the alternative hypothesis: OS hazard ratio = 0.72, it is calculated that a total of 297 deaths need to be observed. This calculation assumes analysis by a one-sided log-rank test at the overall 2.5% level of significance, participants randomized to the two treatment arms in a 1:1 ratio, and a 4-

look group sequential design with a Lan-DeMets (O'Brien-Fleming) alpha spending function using information fractions of (approximately 0.135, 0.42, 0.75, 1).

Assuming that enrolment will continue for 10.5 months at a non-uniform rate and losses to follow-up for OS of approximately 5%, a total of 450 participants will need to be randomized to observe the targeted 297 deaths events at about 43 months after the randomization date of the first participant. Based on the assumptions above and a sample size of 450 participants, the rPFS analysis is estimated to occur at approximately 13.5 months.

4 Change to protocol specified analyses

Section	Protocol	SAP update
SAP Section 2.2 updated as compared to the initial protocol released on 03-Nov-2020	**The Ga-PSMA-11 Full Analysis set (Ga-FAS)** includes all participants who received the administration of ^{68}Ga-PSMA-11. **This analysis set ~~will include all participants with ^{68}Ga-PSMA-11 PET/CT scan during screening assessment and~~** will be the basis of the safety and imaging summaries for the ^{68}Ga-PSMA-11 specific analyses. This includes screened participants that are not enrolled (i.e., not randomized).	**The Ga-PSMA-11 Full Analysis set (Ga-FAS)** includes all participants who received the administration of [^{68}Ga]-PSMA-11. This analysis set will be the basis of the safety and imaging summaries for the [^{68}Ga]-PSMA-11 specific analyses. This includes all screened participants who received [^{68}Ga]Ga-PSMA-11 and were randomized, and those who received [68Ga]Ga-PSMA-11 but not randomized.
SAP Section 2.5.2 updated as compared to the protocol released on 13-Jan-2022	The primary efficacy variable, rPFS, will be analyzed when 156 rPFS events are observed. Analyses will be based on the FAS population according to the randomized treatment group and strata assigned at randomization (strata formed using the randomization factor as obtained via IRT). The rPFS distribution will be estimated using the Kaplan-Meier method, and Kaplan-Meier curves, medians and 95% confidence intervals of the medians will be presented for each treatment group. The hazard ratio for rPFS will be calculated, along with its 95% confidence interval, from a	The primary efficacy variable, rPFS, was to be analyzed after approximately 156 rPFS events were observed. Analyses will be based on the FAS population according to the randomized treatment group and strata assigned at randomization (strata formed using the randomization factor as obtained via IRT). The rPFS distribution will be estimated using the Kaplan-Meier method, and Kaplan-Meier curves, medians and 95% confidence intervals of the medians will be presented for each treatment group. The hazard ratio for rPFS will be

	stratified Cox model using the same stratification factors as for the log-rank test.	calculated, along with its 95% confidence interval, from a stratified Cox model using the same stratification factors as for the log-rank test. [redacted]
SAP Section 2.6.2 updated as compared to the protocol released on 13-Jan-2022	• Therefore, a three-look group sequential design is considered for OS. • OS will be hierarchically tested in the following way: • The time point for the first OS interim analysis will be at the same time as the rPFS analysis, after approximately 25% of deaths (74 deaths) are expected to have been recorded in the clinical database. • If OS is not statistically significant at the first interim analysis, the time point for the second OS interim analysis will be after approximately 75% of deaths (223 deaths) are expected to have been recorded in the clinical database. • If OS is not statistically significant at the first or second interim analysis, a final analysis is planned at the time approximately 297 deaths have been recorded.	Therefore, a four-look group sequential design is considered for OS. OS will be hierarchically tested in the following way: • The time point for the first OS interim analysis was at the same time as the rPFS analysis at which a very limited information fraction was observed • OS was not statistically significant at the first interim analysis, the time point for the second OS interim analysis will be after approximately 42% of deaths (125 deaths) have been recorded in the clinical database. • If OS is not statistically significant at the second interim analysis, the time point for the third OS interim analysis will be after approximately 75% of deaths (223 deaths) have been recorded in the clinical database. • If OS is not statistically significant at the third interim analysis, a final analysis is

		planned at the time approximately 297 deaths have been recorded.
SAP Section 2.13 updated as compared to the protocol released on 13-Jan-2022	BOR for each participant is determined from the sequence of overall (lesion) responses according to the following rules: • CR = at least two determinations of CR at least 4 weeks apart before progression. • PR = at least two determinations of PR or better at least 4 weeks apart before progression (and not qualifying for a CR). • SD = at least one SD assessment (or better) > 7 weeks after **randomization** (and not qualifying for CR or PR). • PD = progression ≤ 17 weeks after **randomization** (and not qualifying for CR, PR or SD). • UNK = all other cases (i.e. not qualifying for confirmed CR or PR and without SD after more than 7 weeks or early progression within the first 17 weeks)	BOR for each participant is determined from the sequence of overall responses according to the following rules: • CR = at least two determinations of CR at least 4 weeks apart before progression. • PR = at least two determinations of PR or better at least 4 weeks apart before progression (and not qualifying for a CR). • SD = at least one SD assessment (or better) > 7 weeks after **treatment start** (and not qualifying for CR or PR). • Non-CR/Non-PD = at least one non-CR/non-PD assessment (or better) > 7 weeks after treatment start (and not qualifying for CR or PD). • PD = progression ≤ 17 weeks after **treatment start** (and not qualifying for CR, PR or SD). • UNK = all other cases (i.e. not qualifying for confirmed CR or PR and without SD after more than 7 weeks or early progression within the first 17 weeks)
	[redacted]	

	[redacted]	
SAP Section 2.8.1 updated as compared to the protocol released on 03-Nov-2020	Adverse events reported related to 68Ga-PSMA-11 that occur beyond the 14-day reporting window but occur before the initiation of randomized treatment are also ^{68}Ga-PSMA-11 TEAEs. Unrelated 68Ga-PSMA-11 adverse events that occur beyond 14 days will not be TEAEs.	Note: Adverse events reported related to 68Ga-PSMA-11 are also ^{68}Ga-PSMA-11 TEAEs, irrespective of time of onset or start of randomized treatment.
One subject was randomized after the data cut-off date for the primary rPFS analysis due to logistical reasons (i.e. drug supply and visit scheduling). Data for this subject was therefore not included in the primary analysis but will be included in subsequent analyses. Details for this subject will be clarified in the clinical study report.		

5 Appendix

5.1 Imputation rules

This section contains general imputation rules, a more detailed complex rules will be defined in the PDS.

5.1.1 Study drug

The following rule should be used for the imputation of the dose end date for a given study treatment component:

Scenario 1: If the dose end date is completely missing and there is no EOT page and no death date, the patient is considered as on-going:

The patient should be treated as on-going and the cut-off date should be used as the dose end date.

Scenario 2: If the dose end date is completely or partially missing and the EOT page is available:

Case 1: The dose end date is completely missing, and the EOT completion date is complete, then this latter date should be used.

Case 2: Only Year(yyyy) of the dose end date is available and yyyy < the year of EOT date:
Use Dec31yyyy

Case 3: Only Year(yyyy) of the dose end date is available and yyyy = the year of EOT date:
Use EOT date

Case 4: Both Year(yyyy) and Month (mm) are available for dose end date, and yyyy = the year of EOT date and mm < the month of EOT date:
Use last day of the Month (mm)

All other cases should be considered as a data issue and the statistician should contact the data manager of the study.

After imputation, compare the imputed date with start date of treatment, if the imputed date is < start date of treatment:
Use the treatment start date

Patients with missing start dates are to be considered missing for all study treatment component related calculations and no imputation will be made. If start date is missing then end-date should not be imputed.

5.1.2 AE and concomitant medication and safety assessment date imputation

Table 5-1 Imputation of start dates (AE, CM) and assessments (LB, EG, VS)

Missing Element	Rule
day, month, and year	• No imputation will be done for completely missing dates

day, month	• If available year = year of study treatment start date then
	  ○ If stop date contains a full date and stop date is earlier than study treatment start date then set start date = 01JanYYYY
	  ○ Else set start date = study treatment start date.
	• If available year > year of study treatment start date then 01JanYYYY
	• If available year < year of study treatment start date then 01JulYYYY

day	• If available month and year = month and year of study treatment start date then
	  ○ If stop date contains a full date and stop date is earlier than study treatment start date then set start date= 01MONYYYY.
	  ○ Else set start date = study treatment start date.
	• If available month and year > month and year of study treatment start date then 01MONYYYY
	• If available month and year < month year of study treatment start date then 15MONYYYY

Table 5-2 Imputation of end dates (AE, CM)

Missing Element	Rule (*=last treatment date plus <30> days not > (death date, cut-off date, withdrawal of consent date))
day, month, and year	• Completely missing end dates (incl. ongoing events) will be imputed by the end date of the on-treatment period*
day, month	• If partial end date contains year only, set end date = earliest of 31DecYYYY or end date of the on-treatment period *
day	• If partial end date contains month and year, set end date = earliest of last day of the month or end date of the on-treatment period*

Any AEs and ConMeds with partial/missing dates will be displayed as such in the data listings.

Any AEs and ConMeds which are continuing as per data cut-off will be shown as ‘ongoing’ rather than the end date provided.

The above imputations are only used for analyses of time to and duration of AEs and concomitant medications.

5.1.3 Incomplete date for anti-neoplastic therapies

Prior therapies

Start date:

The same rule which is applied to the imputation of AE/concomitant medication start date will be used with the exception that randomization date should be used in place of treatment start date.

End date: (if applicable)

Imputed date = min (randomization date, last day of the month), if day is missing;

Imputed date = min (randomization date, 31DEC), if month and day are missing.

If the end date is not missing and the imputed start date is after the end date, use the end date as the imputed start date.

If both the start date and the end date are imputed and if the imputed start date is after the imputed end date, use the imputed end date as the imputation for the start date.

Post therapies

Start date:

Imputed date = max (last date of study treatment + 1, first day of the month), if day is missing;

Imputed date = max (last date of study treatment + 1, 01JAN), if day and month are missing.

End date: No imputation.

5.1.4 Incomplete dates for disease progression prior to start of study drug

If day of PD associated with prior antineoplastic medication is missing then imputed PD date:

= min (midpoint between the end date of the prior antineoplastic medication and the end of the month, start date of study treatment, start date of prior medication from the next regimen), if end date of prior antineoplastic medication is in the same month as the PD date,

= min (15th of the month of the PD date, start date of study treatment, start date of prior medication from the next regimen), if end date of prior antineoplastic medication is in a month prior to the PD date

= 15th of the month of the PD date, if end date of prior antineoplastic medication is in a month after the PD date.

If both day and month of PD associated with prior antineoplastic medication are missing then imputed PD date:

= min (midpoint between the end date of the prior antineoplastic medication and the end of the year, start date of study treatment, start date of prior medication from the next regimen), if end date of prior antineoplastic medication is in the same year as the PD date

= min (July 1 of the year of the PD date, start date of study treatment, start date of prior medication from the next regimen), if end date of prior antineoplastic medication is in a year prior to the PD date

= July 1 of the year of the PD date, if end date of prior antineoplastic medication is in a year after the PD date

Completely missing PD dates will not be imputed. The start date of medication from the next regimen is based on the earliest start date of any medication(s) from the next regimen. For the mid-point calculation, if odd days in between, (e.g. last dose of medication is 27 June 2012, and end of the month is 30 June 2012), then use the next day from the midpoint calculation (e.g. mid-point is 29 June 2013).

5.1.5 Incomplete dates for disease progression on further antineoplastic therapies

If day of PD associated with further antineoplastic medication is missing then imputed PD date:

= min (midpoint between the end date of the further antineoplastic medication and the end of the month, start date of further medication from the next regimen), if end date of further antineoplastic medication is in the same month as the PD date,

= min (15th of the month of the PD date, start date of further medication from the next regimen), if end date of prior antineoplastic medication is in a month prior to the PD date

= 15th of the month of the PD date, if end date of further antineoplastic medication is in a month after the PD date.

If both day and month of PD associated with further antineoplastic medication are missing then imputed PD date:

= min (midpoint between the end date of the further antineoplastic medication and the end of the year, start date of further medication from the next regimen), if end date of further antineoplastic medication is in the same year as the PD date

= min (July 1 of the year of the PD date, start date of further medication from the next regimen), if end date of further antineoplastic medication is in a year prior to the PD date

= July 1 of the year of the PD date, if end date of further antineoplastic medication is in a year after the PD date

Completely missing PD dates will not be imputed. The start date of medication from the next regimen is based on the earliest start date of any medication(s) from the next regimen. For the mid-point calculation, if odd days in between, (e.g. last dose of medication is 27 June 2012, and end of the month is 30 June 2012), then use the next day from the midpoint calculation (e.g. mid-point is 29 June 2013).

5.1.5.1 Other imputations

Incomplete date of initial diagnosis of cancer and date of most recent recurrence

Missing day is defaulted to the 15th of the month and missing month and day is defaulted to 01-Jan.

Incomplete assessment dates for tumor assessment

All investigation dates (e.g. MRI scan, CT scan) must be completed with day, month and year. If one or more assessment dates are incomplete but other investigation dates are available, this/these incomplete date(s) are not considered for calculation of the assessment date and assessment date is calculated as the latest of all investigation dates (e.g. MRI scan, CT scan) if the overall response at that assessment is CR/PR/SD/UNK. Otherwise – if overall response is progression – the assessment date is calculated as the earliest date of all investigation dates at that evaluation number. If all measurement dates have no day recorded, the 1st of the month is used. If the month is not completed, for any of the investigations, the respective assessment will be considered to be at the date which is exactly between previous and following assessment. If a previous and following assessment is not available, this assessment will not be used for any calculation.

Applying the cut-off to tumor assessment

For tumor related assessments, if an evaluation has some assessments done prior to cut-off date and others after the cut-off date , then the evaluation is considered post-cut-off date and will be excluded from analysis.

5.2 AEs coding/grading

Adverse events are coded using the Medical dictionary for regulatory activities (MedDRA) terminology.

AEs will be assessed according to the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0.

The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening, and death. This grading system inherently places a value on the importance of an event, although there is not necessarily proportionality among grades (a grade 2 is not necessarily twice as bad as a grade 1).

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 . The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. The criteria to assign CTCAE grades are given in Novartis internal criteria

for CTCAE grading of laboratory parameters. The latest available version of the document based on the underlying CTCAE version 5.0 at the time of analysis will be used. For laboratory tests where grades are not defined by CTCAE v5.0, results will be graded by the low/normal/high (or other project-specific ranges, if more suitable) classifications based on laboratory normal ranges.

A severity grade of 0 will be assigned for all non-missing lab values not graded as 1 or higher. Grade 5 will not be used. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia.

Imputation Rules

CTCAE grading for blood differentials is based on absolute values. However, this data may not be reported as absolute counts but rather as percentage of WBC.

If laboratory values are provided as '<X' (i.e. below limit of detection) or '>X', prior to conversion of laboratory values to SI unit, these numeric values are set to X.

The following rules will be applied to derive the WBC differential counts when only percentages are available for a xxx differential

xxx count = (WBC count) * (xxx %value / 100)

Further derivation of laboratory parameters might be required for CTCAE grading. For instance, corrected calcium can be derived using the reported total calcium value and albumin at the same assessment using the following formula:

Corrected Calcium (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4]

In order to apply the above formula, albumin values in g/L will be converted to g/dL by multiplying by 0.1), calcium values in mmol/L will be converted to mg/dL by dividing by 0.2495. For calculation of laboratory CTC grades 0 and 1, the normal range for derived corrected calcium is set to the same limits (in mg/dL) as for calcium.

CTC grades for the derived absolute WBC differential counts (neutrophils, lymphocytes) and corrected calcium will be assigned as described above for grading

5.4 Statistical models

5.4.1 Analysis of time to events data

Hypothesis and test statistic

The null hypothesis stating that rPFS/OS survival distributions of the two treatment groups are equivalent will be tested against one-sided alternative.

Assuming proportional hazards for rPFS/OS, the following statistical hypotheses will be tested:

$$H_{01}: \theta_1 \geq 1 \quad vs. \quad H_{A1}: \theta_1 < 1$$

where θ_1 is the rPFS hazard ratio ([^{177}Lu]-PSMA-617 arm versus ARDT arm).

Stratified log-rank test adjusting for the strata used in the randomization will be implemented as follows: In each of the K strata separately, the LIFETEST procedure with STRATA statement including only the treatment group variable and with the TIME statement will be used to obtain the rank statistic S_k and variance var(S_k) where k=1, 2, …, K. The final test statistics will then be reconstructed as follows:

$Z=[S_1+\ldots+S_K]/\sqrt{[var(S_1)+\ldots+var(S_k)]}$.

One-sided p-value will be obtained using Z statistic.

Kaplan-Meier estimates

An estimate of the survival function in each treatment group will be constructed using Kaplan-Meier (product-limit) method as implemented in PROC LIFETEST with METHOD=KM option. The PROC LIFETEST statement will use the option CONFTYPE=LOGLOG.

Median survival for each treatment group will be obtained along with 95% confidence intervals calculated from PROC LIFETEST output using the method of Brookmeyer and Crowley 1982. Kaplan-Meier estimates of the survival function with 95% confidence intervals at specific time points will be summarized. The standard error of the Kaplan-Meier estimate will be calculated using Greenwood's formula Collett 1994.

Hazard ratio

Hazard ratio will be estimated by fitting the Cox proportional hazards model using SAS procedure PHREG (with TIES=EXACT option in the MODEL statement).

A stratified unadjusted Cox model will be, i.e. the MODEL statement will include the treatment group variable as the only covariate and the STRATA statement will include stratification variable(s).

Hazard ratio with two-sided 95% confidence interval will be based on Wald test.

Treatment of ties

The STRATA statement in LIFETEST procedure will be used to analyze time to event data with ties. The PHREG procedure in SAS with option TIES=EXACT will be used to fit the Cox proportional hazards model.

Checking proportionality of hazard assumption

Plots (SURVIVAL LOGSURV LOGLOGS) generated by LIFETEST procedure in SAS will be used to provide visual checks of the proportional hazard assumption.

- SURVIVAL plots estimated survivor functions. The shape of the curves should be basically the same if hazards are proportional.
- LOGSURV plots the cumulative hazard functions The larger cumulative hazard should be a multiple of smaller if hazards are proportional
- LOGLOGS plots log (cumulative hazard). The LOGLOG plot will show parallel curves if hazards are proportional.

5.4.2 Rank preserving structural failure time (RPSFT) model

5.4.2.1 Background

RPSFT models are a class of "structural" accelerated failure time models with time-dependent covariates (Robins and Tsiatis 1991). The term "structural" stems from social science and econometrics and means that event time is modelled as a counterfactual time: the time that would have been observed if the patient had received the control treatment. The RPSFT model is rank preserving: it assumes that if patient *i* failed before patient *j* when both followed a particular treatment, then patient *i* would fail before patient *j* when both followed any other particular treatment.

The RPSFT method estimates the treatment effect taking into account the treatment switching from the control arm to the experimental arm. It is randomization-based in the sense that all patients are analyzed in the treatment arm they were randomized to, as opposed to non-randomized approaches (not described in this guideline), which suggest to exclude crossover patients or to analyze them in the experimental treatment arm (Morden et al. 2011).

RPSFT model: counterfactual survival times

For patient i, let T_i be the observed survival time and U_i be the corresponding counterfactual time. U_i can only be observed (and equals T_i) for patients who were randomized to the control arm and did not switch treatment (Figure 5-1).

U_i is linked to T_i through the causal accelerated failure time model and a structural parameter ψ

$$U_i = \int_0^{T_i} e^{\psi L_i(t)} dt = A_i e^{\psi} + (T_i - A_i),$$

where $L_i(t) = 1$ if patient i received experimental treatment at time t and 0 otherwise, and A_i is the total time on experimental treatment. e^{ψ} is the multiplication (acceleration) factor by which the observed survival times would be shrunken (ψ<0) or extended (ψ>0) to obtain the respective counterfactual survival times (Robins and Tsiatis 1991).

RPSFT model: counterfactual censoring times

The counterfactual survival times $U_i(\psi)$ are computed from the observed survival times and therefore cannot be calculated for censored observations. We assume that, for a patient *i*, follow-up will end at time C_i, so that C_i is known as is the last contact date prior to the study cut off date. Under the accelerated failure time model, the censoring times may be defined as for the event times,

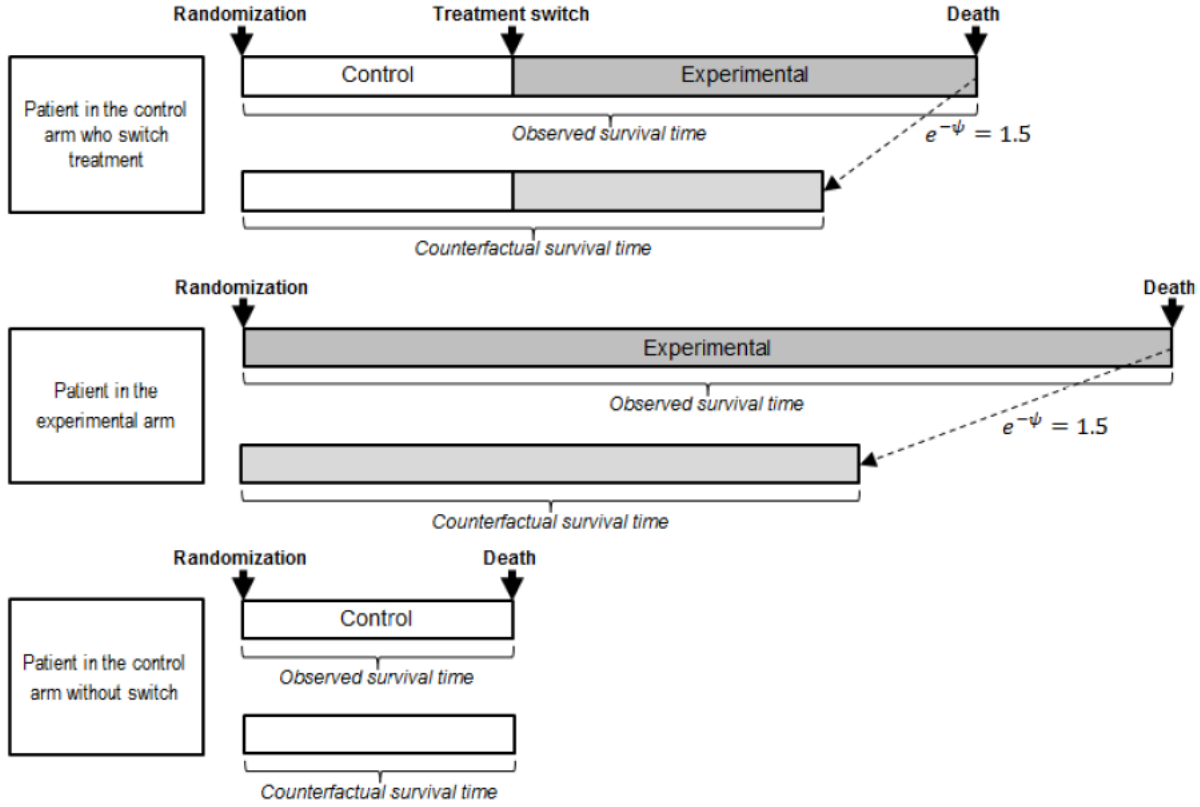
$$D_i = \int_0^{C_i} e^{\psi L_i(t)} dt = A_i e^{\psi} + (C_i - A_i)$$

However, this censoring is not appropriate because the censoring time depends on the received treatment $A_i(t)$ and hence may depend on prognosis. Thus, it could introduce informative censoring. Therefore, it is required to define the counterfactual censoring time as $C_i(\psi) = \min\left(C_i, C_i e^{\psi}\right)$, which denotes the earliest possible censoring time for all possible realizations

of $A_i(t)$. For patients who had an event, the study cut off date will be used in the calculation of the possible censoring times. This permits to preserve non-informative censoring (Korhonen et al. 2012).

Figure 5-1 shows the observed and counterfactual survival times for patients (1) who were on control treatment and then switched to experimental treatment; (2) who were treated with experimental treatment; (3) who were treated with control treatment only. Here, the acceleration factor is 1.5 ($\psi = -0.405$).

Figure 5-1 Observed and counterfactual survival times for three types of patients



If a patient discontinues experimental treatment, there are two approaches to calculate counterfactual survival and censoring times (Latimer and Abrams 2014):

- *on treatment* approach: counterfactual times are obtained by multiplying the acceleration factor with the total time spent on experimental treatment.

  This approach should be considered if the treatment effect is not expected to extend beyond experimental treatment duration, or if patients are likely to continue with another line of therapy after discontinuation.

- *treatment group* approach: counterfactual times are obtained by multiplying the acceleration factor with the total time from initiation of experimental treatment until end of follow up.

  This approach should be considered if the treatment effect is expected to extend beyond the experimental treatment duration, or if patients are likely to continue with the similar or the same treatment after discontinuation.

Estimation of the structural parameter ψ

When estimating the structural parameter ψ, the following times are used: for patients who had an event and for whom the counterfactual survival time U_i is calculated, the time variable is defined as $X_i = \min(U_i, C_i e^{\psi})$ and the event will be censored unless $X_i = U_i$; for patients who were originally censored and for whom the counterfactual censoring time $C_i(\psi)$ is calculated, the time variable is defined as $X_i = C_i(\psi)$.

The structural parameter ψ is estimated via grid search (G-estimation) using the following steps:

- Divide the range of possible values of ψ into a fine grid with a small step size (e.g. 0.001)
- Calculate $U_i(\psi)$ and $X_i(\psi)$ for patients with an event and $C_i(\psi)$ and $X_i(\psi)$ for patients without an event for each ψ in the grid.
- Compare the survival distributions of $X_i(\psi)$ between the two treatment arms as randomized with the test statistic used for the ITT analysis (e.g. log-rank, Wald). Note that for stratified randomization, a stratified version of the test should be used.
- Find the value $\hat{\psi}$ that minimizes the test statistic (or maximizes the p-value). In reality the test statistic is a step function in ψ and therefore there might be several values where the minimum is obtained, and the recommended choice then would be the maximum value of ψ as the most conservative approach. Test-based lower and upper 95% confidence limits are obtained as the values closest to the point estimate where the value of the test statistic changes from less than 3.84 to greater than 3.84 if the test statistic has a chi-squared distribution with 1 degree of freedom (or equivalently the p-value changes from greater than 0.05 to less than 0.05).

  Since counterfactual survival times are the times that would have been observed if no experimental treatment had been received, a good estimate of ψ should lead to very similar survival distributions, that is, a small test statistic.

Estimation of the treatment effect

For the estimated $\hat{\psi}$ the crossover-corrected hazard ratio (HR_c) can be obtained from a Cox regression model by transforming original survival times with the counterfactual times (White et al., 1999, Korhonen et al., 2012). The data for the model include the original times for patients in the experimental treatment arm, the original times for patients in the control arm who did not switch treatment, and the counterfactual times for patients in the control arm who switched treatment.

The symmetrical test-based $(1 - \alpha) \times 100\%$ confidence interval (CI) for the estimated hazard ratio can be calculated using the two-sided p-value from the test statistic of the ITT analysis.

The obtained CI will be consistent with the CI for the ITT analysis because each interval will simultaneously (not) contain the null.

The respective standard error of the estimated HR_c (Korhonen et al. 2012) is

$$SE = \frac{\left|\widehat{Log(HR_c)}\right|}{z_{1-p/2}}$$

From this it follows that the $(1-\alpha) \times 100\%$ CI of HR_c is

$$\exp\left(\widehat{Log(HR_c)} \pm z_{1-\alpha/2} \times SE\right)$$

Another calculation of the confidence interval for the hazard ratio is bootstrapping, which takes account of the uncertainty in the estimation of the structural parameter (White et al. 1999). The original survival data is bootstrapped, and for every sample the entire procedure is repeated and the 2.5 percentile and 97.5 percentile of all the hazard ratios obtained is then reported as the 95% confidence interval. This approach is not recommended, because the value of ψ may not always be estimable.

5.4.3 Inverse probability weighting (IPW) methods

5.4.3.1 Background

The objective of IPW methods is to assess the causal effect of *exposure to treatment* on the *outcome*, i.e., survival. As shown in the previous sections, in the presence of time-dependent confounders, the standard analyses may lead to biased estimation of treatment effect. Marginal structural models using IPW are one approach to construct a potentially unbiased estimate.

The main assumptions for IPW methods are:

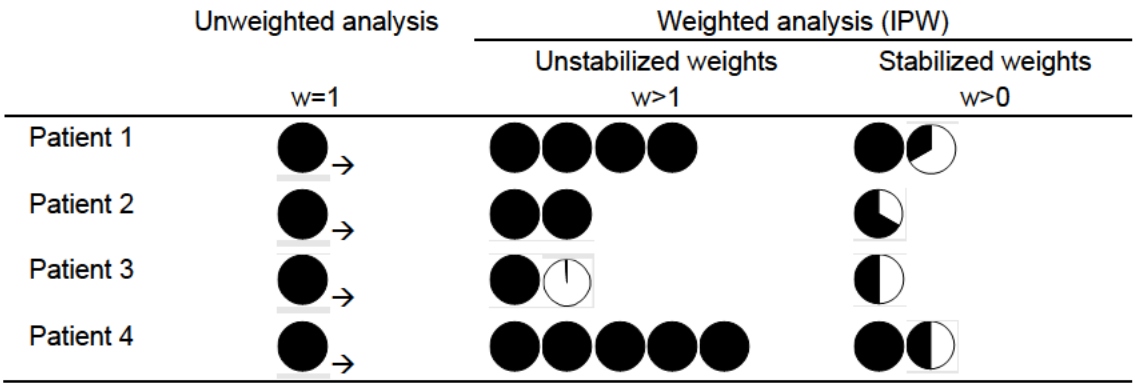
- Exchangeability: no unmeasurable confounding is present. This implies that the relevant confounders are known and have been assessed in the study.
- Positivity: for each level of the confounders, there must be exposed and unexposed patients. For example, this assumption is violated if there is a variable or a level for a variable that determines treatment switching.
- No misspecification of the models to derive the weights and the weighted Cox regression model. An example would be use of only a linear term for a variable for which a quadratic term would be more appropriate.

The basic idea of marginal structural models is to construct a pseudo-risk set in which the effect of the confounders with respect to treatment is removed. This is achieved by balancing the original study population: each patient is assigned a weight according to his or her baseline and on-treatment characteristics. This pseudo-risk set will afterwards be used in a weighted Cox regression model or any other survival analysis to estimate the unconfounded treatment effect.

The pseudo-risk set is usually constructed using Inverse-Probability Weighting (IPW), which determines a patient weight as the inverse of the probability of switching as a function of baseline and time-dependent covariates.

Since probability is defined between 0 and 1, the inverse will always be greater than 1 and is called unstabilized weight. A disadvantage of unstabilized weights in the weighted Cox regression is high variability of the weights and large size of the modified risk set compared to the original risk set. To overcome these limitations, stabilized weights are recommended. They are defined as the ratio of the probability of treatment conditional on past exposure and baseline covariates and the probability of treatment conditional on past exposure, baseline covariates, and time dependent cofounders. Therefore, stabilized weights can take values larger and smaller than 1 (see Figure 5-2). Stabilized weights can increase the efficiency of the estimate (Hernan et al. 2000) and are therefore recommended.

Figure 5-2 Illustration of the modified risk set



The weight here is a stabilized weight retaining the original number of patients and hence allowing the weight to be less than 1.

A patient's weight can vary over the entire observation period. Follow-up is divided into time intervals $(t_i, t_{i+1}]$ with constant values for time-dependent confounder x_i and weight w_i. The time intervals can be of equal or variable length for patient 1 as shown in Table 5-3, which uses the counting process notation. In this example, a patient is being observed from time 0 until the event at time t_3. The patient had three assessments for the time dependent covariates x_1, x_2, x_3. The pseudo-risk set thus becomes a "per time interval" pseudo-risk set and hence the weights are calculated for each time interval.

Table 5-3 Presentation of survival data using counting process notation

Patient	Time interval	Event (0=No/1=Yes)	Time-dependent covariate impacting outcome	Weight
1	$(0, t_1]$	0	x_1	w_1
1	$(t_1, t_2]$	0	x_2	w_2
1	$(t_2, t_3]$	1	x_3	w_3

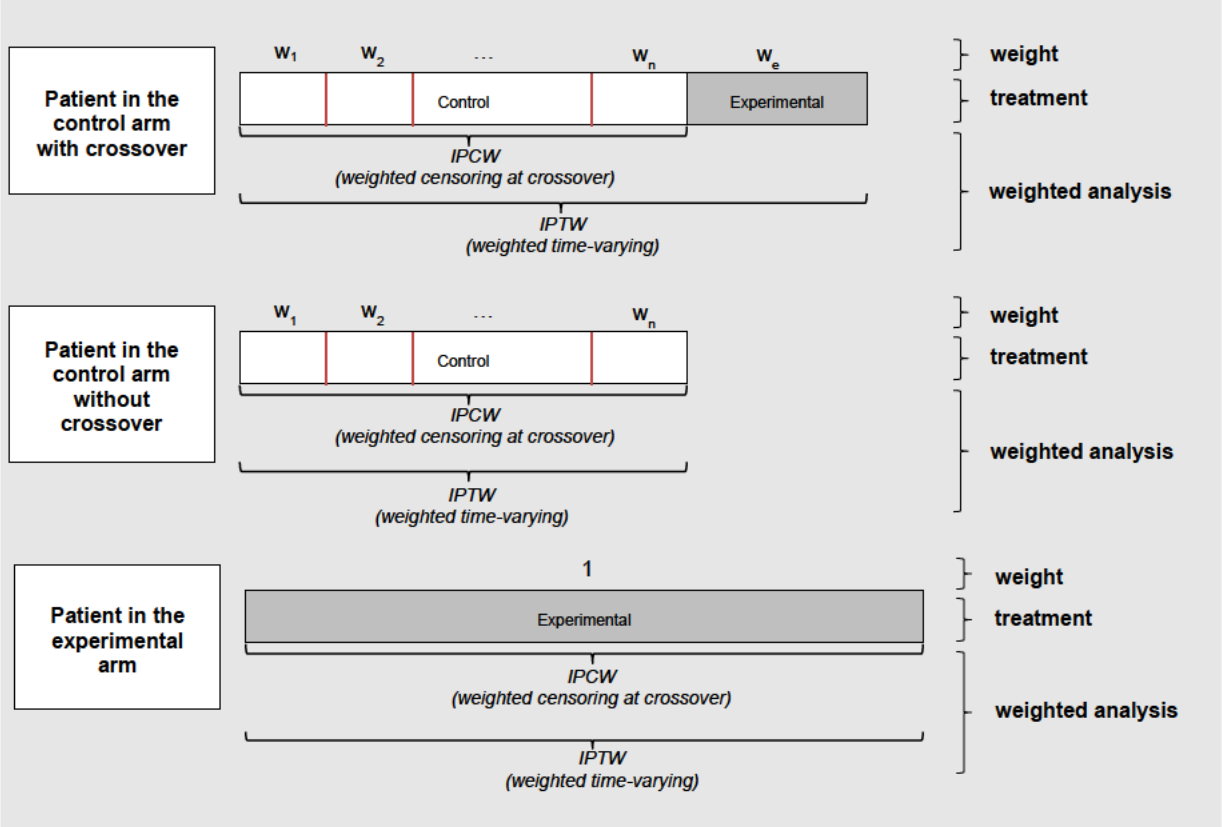
5.4.3.2 IPCW vs IPTW

Publications in the field often use two terms: IPCW and IPTW, inverse probability of censoring weighting and inverse probability of treatment weighting. Both approaches rely on the principles described above but model and use the data in a different way as explained below.

Considering the analysis of OS and accounting for treatment switching from control to experimental arm, IPCW uses data until the time of treatment switching and ignores data post-crossover, while IPTW uses data even beyond crossover.

Figure 5-3 shows three types of patients: (1) patients who were treated with control treatment and then crossed over to experimental treatment; (2) patients who were treated with experimental treatment; (3) patients who were treated with control treatment only.

Figure 5-3 Inverse probability weighting



For a patient who was on experimental treatment since start of study, a weight of 1 is assigned for the entire follow-up time. For patients on control treatment, the time is segmented and a weight is calculated for each time interval. Once a patient crosses over, the same weight is retained throughout the remaining follow-up time post crossover. It should be noted that the entire follow-up time is used in IPTW analyses whereas only the time until crossover is used in IPCW analyses. Therefore, the precision of an IPTW analysis may be higher.

IPCW can also be used to model different censoring reasons, such as censoring due to treatment switching, loss to follow-up, and discontinuation other than disease progression. More theory can be found in Hernan et al (2000), and a practical application from the MEGA study in coronary heart disease is discussed in Yoshida et al (2007).

5.4.3.3 Main implementation steps

There are four steps involved in estimating the crossover corrected treatment effect using IPW:

1. Selection of baseline and time-dependent covariates.
2. Preparation of dataset.
3. Calculation of weights: conditional probabilities of remaining on control treatment (IPTW) or remaining uncensored (IPCW).
4. Estimation of the crossover corrected treatment effect using weighted Cox regression.

5.4.4 Two-stage method

A simplified method presented in Latimer N 2014 may also be used to adjust for the crossover. This methods relies on the assumption that switching will likely occur soon after disease progression. Assuming patients are at a similar stage of disease at the time of progression, the method proposes to estimate the effect of the new treatment on extending survival from disease progression to death in the control group of patients who switch treatment.

To achieve this, a Weibull model (or any other Accelerated Failure Time model) will be fitted to data from patients in the control group only, including covariates measured at the time of crossover and a time-varying covariate indicating treatment switch. This will allow to estimate the treatment effect in patients who crossed over, compared to those who did not.

Counterfactual survival times are subsequently calculated as:

$$U_i = T_{A_i} + \frac{T_{B_i}}{\mu_B}$$

where T_{A_i} represents the time spent on control treatment, T_{B_i} represents the time spent on the new intervention and μ_B is the treatment effect (AF) in switching patients.

This approach makes the following assumptions:

1. No time dependent confounding between the time of disease progression and the time of crossover
2. Parametric assumptions to use the AFT model to estimate the treatment effect in patients who crossover.

6 References

Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38, 29 - 41.

Cella D, Nichol MB, Eton D, Nelson JB, Mulani P. Estimating Clinically Meaningful Changes for the Functional Assessment of Cancer Therapy—Prostate: Results from a Clinical Trial of Patients with Metastatic Hormone-Refractory Prostate Cancer. Value in Health 12(1): 124-129, 2009.

Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrical, 26, 404-413.

Collet D (1994). Modelling survival data in medical research. London, Chapman & Hall.

de Bono J, Mateo J, Fizazi K, et al (2020) Olaparib for Metastatic Castration-Resistant Prostate Cancer. N Engl J Med; 382(22):2091-102.

de Wit R, de Bono J, Sternberg CN, et al (2019) Cabazitaxel versus Abiraterone or Enzalutamide in Metastatic Prostate Cancer. N Engl J Med; 381(26):2506-18.

Ellis S, Carroll KJ, Pemberton K. Analysis of duration of response in oncology trials. Contemp Clin Trials. Jul 2008;29(4):456-65.

Glimm E, Maurer W and Bretz F (2010). Hierarchical testing of multiple endpoints in group-sequential trials. Statistics in Medicine, 29, 219-228.

Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000 Sep;11(5):561-70.

James M. Robins & Anastasios A. Tsiatis (1991) Correcting for non-compliance in randomized trials using rank preserving structural failure time models, Communications in Statistics - Theory and Methods, 20:8, 2609-2631

Komura et al (2019) Comparison of Radiographic Progression-Free Survival and PSA Response on Sequential Treatment Using Abiraterone and Enzalutamide for Newly Diagnosed Castration-Resistant Prostate Cancer: A Propensity Score Matched Analysis from Multicenter Cohort p. 8, 1251; doi:10.3390/jcm8081251.

Korhonen P, Zuber E, Branson M, Hollaender N, Yateman N, Katiskalahti T, Lebwohl D, Haas T. Correcting overall survival for the impact of crossover via a rank-preserving structural failure time (RPSFT) model in the RECORD-1 trial of everolimus in metastatic renal-cell carcinoma. J Biopharm Stat. 2012;22(6):1258-71

Lan KKG, DeMets DL (1983). Discrete sequential boundaries for clinical trials. Biometrika, 70, 649-53.

Latimer N, Abrams K. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.

Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL, Campbell MJ. Adjusting for treatment switching in randomised controlled trials - A simulation study and a simplified two-stage method. Stat Methods Med Res. 2017 Apr;26(2):724-751

Latimer NR, White IR, Abrams KR, Siebert U. Causal inference for long-term survival in randomised trials with treatment switching: Should re-censoring be applied when estimating

counterfactual survival times? Stat Methods Med Res. 2019 Aug;28(8):2475-2493. doi: 10.1177/0962280218780856. Epub 2018 Jun 25. PMID: 29940824; PMCID: PMC6676341.

Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Med Res Methodol. 2011 Jan 11;11:4.

Pickard A, Neary M, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes. 2007 Dec;5(70):139-44.

Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000 Sep;56(3):779-88.

Robins JM, Tsiatis A. (1991). Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. Communications in Statistics, 20:2609-2631.

Scher HI, Morris MJ, Stadler WM, et al (2016) Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. J. Clin. Oncol. 1402-18.

White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. Stat Med. 1999 Oct 15;18(19):2617-34.

Yount S, Cella D, Banik D, Ashraf T, Shevrin D. Brief assessment of priority symptoms in hormone refractory prostate cancer: the FACT Advanced Prostate Symptom Index (FAPSI). Health Qual Life Outcomes. 2003 Nov 21;1:69

Yoshida M, Matsuyama Y, Ohashi Y; MEGA Study Group. Estimation of treatment effect adjusting for dependent censoring using the IPCW method: an application to a large primary prevention study for coronary events (MEGA study). Clin Trials. 2007;4(4):318-28.