Aragon Pharmaceuticals, Inc. *

Statistical Analysis Plan

A Phase 3 Randomized, Placebo-controlled, Double-blind Study of Apalutamide Plus Androgen Deprivation Therapy (ADT) Versus ADT in Subjects with Metastatic Hormone-sensitive Prostate Cancer (mHSPC)

Protocol 56021927PCR3002; Phase 3

JNJ-56021927 (apalutamide)

*Aragon Pharmaceuticals, Inc. is a wholly-owned subsidiary of Johnson & Johnson. Janssen Research & Development, LLC is part of the Janssen Pharmaceutical Companies of Johnson & Johnson and provides various services to its affiliated company, Aragon Pharmaceuticals, Inc.

Status: Approved
Date: 10 January 2019
Prepared by: Janssen Research & Development, LLC
Document No.: EDMS-ERI-96238676; 1.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

LIST OF TABLES

AMENDMENT HISTORY

None

ABBREVIATIONS

ADT	Androgen Deprivation Therapy
AE	Adverse Event
ALP	Alkaline Phosphatase
AR	Androgen Receptor
BPI-SF	Brief Pain Inventory-Short Form
BFI	Brief Fatigue Inventory
CI	Confidence Interval
CRF	Case Report Form
CS	Compound Symmetry
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DAO	Data as Observed
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic Case Report Form
eDISH	Evaluation of Drug-induced Serious Hepatotoxicity
EQ-5D-5L	Euro-QoL
EU	European Union
FACT-P	Functional Assessment of Cancer Therapy – Prostate
GnRHa	Gonadotropin Releasing Hormone Analog
HR	Hazard Ratio
IDMC	Independent Data Monitoring Committee
ITT	Intent-to-Treat
IWRS	Interactive Web Response System
LDH	Lactate Dehydrogenase
LS means	Least Squares Means
MedDRA	Medical Dictionary for Regulatory Activities
mHSPC	Metastatic Hormone Sensitive Prostate Cancer
MRI	Magnetic Resonance Imaging
MRU	Medical Resource Utilization
NA	North of America
NCI	National Cancer Institute
OS	Overall Survival
PCWG2	Prostate Cancer Working Group
PD	Pharmacodynamic(s)
PFS	Progression-free Survival
PhRMA	Pharmaceutical Research and Manufacturing Association
PK	pharmacokinetic(s)
PRO	Patient-reported Outcome
PSA	Prostate Specific Antigen
RECIST	Response Evaluation Criteria in Solid Tumors
rPFS	Radiographic Progression-free Survival
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
SRE	Skeletal-related Event
VAS	Visual Analogue Scale
WHO	World Health Organization

1. INTRODUCTION

This document describes the planned statistical analyses for Protocol 56021927PCR3002: A Phase 3 Randomized, Placebo-controlled, Double-blind Study of Apalutamide Plus Androgen Deprivation Therapy (ADT) Versus ADT in Subjects with Metastatic Hormone-sensitive Prostate Cancer (mHSPC). This statistical analysis plan (SAP) is intended to supplement the study protocol. Any deviations from this analysis plan will be described in the clinical study report.

1.1. Study Objectives

1.1.1. Primary Objective

The primary objective is to determine if the addition of apalutamide (JNJ-56021927) to ADT provides superior efficacy in improving radiographic progression-free survival (rPFS) or overall survival (OS) for subjects with mHSPC.

1.1.2. Secondary Objectives

- To evaluate clinically relevant improvements with addition of apalutamide to ADT including delays in pain progression and opioid use for prostate cancer, skeletal-related events, and the need for cytotoxic chemotherapy
- To characterize the safety of adding apalutamide to ADT in subjects with mHSPC
- To characterize the population pharmacokinetics (PK) and pharmacodynamics (PD) of apalutamide
- To evaluate the concentration of leuprolide and assess the PD effect of leuprolide on testosterone concentrations when used alone or in combination with apalutamide
- To evaluate the treatment effectiveness with the addition of apalutamide to ADT for the subpopulations of subjects with low-volume or high-volume mHSPC

1.1.3. Other Objectives

- To evaluate exploratory biomarkers predictive of response and resistance to treatment
- To evaluate patient relevant outcomes including symptoms (ie, pain, fatigue, urination) and function (ie, physical, emotional, social) and health-related quality of life
- To evaluate improvements in other clinically relevant endpoints with the addition of apalutamide to ADT
- To collect medical resource utilization (MRU) data that may be used in future economic modeling (the construction and reporting of the economic model will be conducted separately from this study)

1.2. Study Design

This is a randomized, double-blind, placebo-controlled, multinational, multicenter Phase 3 study to determine if subjects with mHSPC will benefit from the addition of apalutamide to ADT. Enrollment of approximately 1,000 subjects is planned for this study. Subjects who meet all of the inclusion criteria and none of the exclusion criteria will be stratified by Gleason score at diagnosis, prior docetaxel use, and geographic region, and then randomly assigned in a 1:1 ratio to the apalutamide plus ADT or matching placebo plus ADT group.

The study will include a Screening Phase of up to 28 days before randomization to establish study eligibility. Subjects will receive treatment in 28-day treatment cycles during the Treatment Phase until

disease progression or the occurrence of unacceptable treatment-related toxicity. Subjects must discontinue study drug with documented clinical progression. After discontinuation of study drug, subjects will have an End-of-Treatment Visit within 30 days after the last dose of study drug. In the event of a positive study result at either of the interim analyses or at the final analysis, all subjects in the Treatment Phase will have the opportunity to enroll in an Open-label Extension Phase, which will allow subjects to receive active drug (apalutamide) for up to 3 years.

Subjects will be monitored for safety starting from the signing of informed consent until 30 days after the last dose of study drug. Adverse events including laboratory AEs will be graded and summarized using the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) Version 4.03. Dose modifications will be made as required according to dose modification rules. An Independent Data Monitoring Committee (IDMC) will be commissioned for the study to provide recommendation during the planned interim efficacy analysis and regular safety review.

1.3. Statistical Hypotheses for Study Objectives

The hypothesis of the study is that apalutamide plus ADT compared with ADT alone will improve rPFS or OS or both, and have an acceptable safety profile in subjects with mHSPC.

The dual-primary endpoints are radiographic progression-free survival (rPFS) and overall survival (OS). The general hypothesis used to address the primary objective is as follows:

H_0: The survival distributions (rPFS or OS) of the experimental group (apalutamide +ADT) $S_E(t)$, and that of the control group (Placebo+ADT), $S_C(t)$, are equal:

$S_E(t) = S_C(t)$, for all $t > 0$

H_1: The survival distributions (rPFS or OS) are not equal:

$S_E(t) \neq S_C(t)$, for some $t > 0$,

where rPFS, as assessed by investigators, is defined as the duration from the date of randomization to the date of first documentation of radiographic progression, or death from any cause, whichever occurs first. Overall survival (OS) is defined as the time from randomization to the date of death from any cause.

1.4. Sample Size Justification

An overall type I error of 5% is planned for this study. This study utilizes the dual-primary endpoints of rPFS and OS with a 0.005 level of significance allocated for the rPFS endpoint and 0.045 is allocated for OS.

It is assumed that the failure distribution of the dual-primary endpoint of rPFS follows an exponential distribution with a constant hazard rate. It is estimated that approximately 368 rPFS events would be required to provide at least 85% power in detecting a hazard ratio (HR) of 0.67 (median rPFS of 20 months for the control group [ADT] versus 30 months for the treatment group of apalutamide plus ADT) at a 2-tailed significance level of 0.005. The assumption of 20 months for the control is an estimate based on published data.[1,2,3]

The study will also provide sufficient power (approximately 80%) to detect a HR of 0.75 in the dual-primary endpoint of overall survival (OS) based on an assumed OS median of 44 months [3] for the control group (ADT) (i.e., 44 months versus 59 months). Approximately 410 death events will be required to detect the assumed HR at a 2-tailed significance level of 0.045 with enrollment duration of

approximately 30 months (approximately 1,000 subjects). The total study duration will be approximately 54 months to obtain 410 deaths.

1.5. Randomization and Blinding

1.5.1. Randomization

Subjects who meet all the inclusion criteria and none of the exclusion criteria will be randomized in a 1:1 ratio to receive apalutamide plus ADT or matching placebo plus ADT using permuted block randomization. Subjects will be stratified by:

- Gleason score at diagnosis (≤7 vs. >7)
- Region (NA/EU vs. Other Countries)
- Prior docetaxel use (Yes vs. No)

1.5.2. Blinding

This is a double-blind study. All subjects and study team members associated with the study conduct are to remain blinded to treatment group assignment until the study is unblinded. Unblinding can occur in the case of a safety or a medical emergency, or for conducting data review by the IDMC as outlined in the IDMC Charter. Unblinding may also be performed after the subject discontinues from the Treatment Phase of study because of radiographic progression and the investigator feels this information is essential to determine the next course of therapy. Unblinding a subject for this situation may only take place after discussion with the sponsor's medical officer. Subjects who have had their treatment assignment unblinded should be discontinued from the Treatment Phase and entered in the Follow-up Phase.

2. GENERAL ANALYSIS DEFINITIONS

Study Day: For safety, study day will be calculated in reference to the date of first dose. Study Day 1 corresponds to the date the subject receive first dose of study drug. For efficacy analysis, Study Day will be calculated in reference to the date of randomization.

Cycle: For the purpose of the study, a treatment cycle is defined as 28 days. Subjects will begin taking study drug on Day 1 of Cycle 1.

Baseline Value: Unless otherwise specified, the baseline value will be defined as the closest measurement prior to or on the day of the first dose of study drug. Change from baseline will be defined as (post-baseline value – baseline value).

Treatment Duration: Treatment duration will be defined as the duration of time from the date of the first dose of study drug to the date of last dose of study drug.

Time to event: Time to event calculations will be defined as the time from randomization to the date of the event of interest. Time to event or duration of event endpoints will be based on the actual date of the event, not visit number or visit label.

2.1. Visit Windows

The Treatment Phase will begin at Cycle 1 Day 1 of treatment and will continue until study drug is discontinued. Study treatment should be initiated (Cycle 1 Day 1) within 3 calendar days after randomization. Visits for each cycle will have a ±2 day window except for computed tomography (CT)/magnetic resonance imaging (MRI) and bone scans (imaging visits may occur up to 8 days before scheduled cycles).

2.2. Pooling Algorithm for Analysis Centers

There is no plan in pooling the centers (study sites) for efficacy or safety analyses, unless analysis by site or pooled sites is warranted. However, geographical region (NA and EU versus Other Countries) is used as a stratification factor which pools sites by region and will be used for subgroup analysis.

2.3. Analysis Sets

The following analysis sets will be used for this study:

Intent-to-Treat (ITT) Population: The ITT population includes all randomized subjects and will be classified according to their assigned treatment group, regardless of the actual treatment received. Subject disposition and efficacy analyses will be performed on data from the ITT population.

Safety Population: The safety population includes all subjects who received at least 1 dose of study drug as treated.

Biomarker Population: The biomarker population includes randomized subjects who have at least 1 biomarker sample collected.

2.4. Definition of Subgroups

In order to assess the consistency of treatment benefit with respect to the dual-primary efficacy endpoints of rPFS and OS, univariate analyses will be performed for the following important subgroups:

- Baseline Eastern Cooperative Oncology Group (ECOG) performance status (PS) grade (0; 1)
- Geographic region (EU and NA; Other Countries)
- Bone metastasis only at baseline (yes; no)
- Visceral disease at baseline (yes; no)
- Gleason Score at diagnosis (≤7; > 7)
- Prior docetaxel use (yes; no)
- Age (< 65; ≥65; ≥75)
- Baseline PSA above median (yes; no)
- Baseline LDH above upper limit normal (yes; no)
- Baseline ALP above upper limit normal (yes; no)
- mHSPC (high volume; low volume)
- Number of bone lesions(<=10, >10)

High-volume mHSPC is based on an adaptation of the CHAARTED study and is defined as 1) visceral metastases and at least 1 bone lesion or 2) at least 4 bone lesions, with at least 1 bone lesion outside of the vertebral column or pelvis (see also Section 1.2).[3] Low-volume mHSPC is defined as the presence of bone lesion(s) not meeting the definition of high-volume mHSPC.

Subgroup analysis of a country may be conducted for filing purpose.

3. INTERIM ANALYSIS AND INDEPENDENT DATA MONITORING COMMITTEE REVIEW

3.1. Interim Analysis

There will be no interim analysis planned for the dual-primary endpoint of rPFS.

For the OS endpoint, 2 interim analyses are planned for this study after observing approximately 50% (~205 events) and 70% (~287) of the total number of required (410) events (Table 1) [redacted]

3.2. Independent Data Monitoring Committee (IDMC)

An IDMC will be commissioned for the study to perform regular safety review and the planned interim analyses. The IDMC will review the progress of the study and cumulative masked (ie, A vs B instead of actual treatment designation) safety data on a periodic basis as well as serve as the reviewer of the efficacy analysis.

Complete details regarding the composition and governance of the IDMC will be outlined in the IDMC Charter.

4. SUBJECT INFORMATION

4.1. Demographics and Baseline Characteristics

The following parameters will be summarized by treatment group and overall using the ITT population:

- Age, race, ethnicity, weight, height
- Geographic region
- Baseline ECOG PS grade
- Baseline PSA
- Baseline hemoglobin
- Baseline lactate dehydrogenase
- Baseline alkaline phosphatase
- Baseline BPI-SF pain score
- Tumor stage at diagnosis (T; N; M)
- Gleason score at diagnosis (≤7; > 7)
- Extent of disease (bone; node; viscera; other)
- Number of bone lesions at baseline (<=10; >10)
- Previous cancer treatment (surgery; radiotherapy; hormonal; other)
- Prior docetaxel use (yes; no) and number of cycles of prior docetaxel use
- Subjects with low volume or high volume mHSPC

4.2. Disposition Information

Subject enrollment and disposition will be summarized by treatment group. The summary of subject disposition will display the number of subjects randomized, and the numbers of subjects in ITT, safety. The summary will also include the number and percent of subjects who completed the study and who

discontinued from the study and study treatment, respectively, and reason for discontinuation as documented in the subject status case report form.

4.3. Treatment Compliance and Extent of Exposure

The safety population will be used to summarize drug exposure, treatment compliance, and dose modifications by treatment group.

Treatment duration will be defined as the duration from the date of the first dose of study drug to the date of last dose of study drug. The number of tablets taken will be calculated based on the number of tablets dispensed at the study visits minus the number of returned tablets.

The overall percent treatment compliance will be defined as the number of tablets taken during the study divided by the expected number of tablets, multiplied by 100. Each subject should be taking a maximum of 4 tablets of study drug per day while on the study. A subject's expected number of tablets will be calculated as the number of assigned tablets per day multiplied by treatment duration.

Subjects with at least one dose modification and the reason for the dose modification will be summarized by treatment group.

4.4. Protocol Deviations

Protocol deviations and eligibility deviations will also be summarized by treatment group. Protocol deviations will be reviewed on a case-by-case basis and assessed if they are considered major deviations for this study. The final list will be compiled before database lock. Examples of major protocol deviations may include, but are not limited to, the following:

- Deviation from inclusion/exclusion criteria that may affect efficacy endpoints
- Major study drug dosing errors or dose modifications that are not within the protocol specifications that may compromise subject safety or efficacy assessments.
- Administration of prohibited concomitant medication during the course of the study treatment period
- Any other deviation that impacts subject safety

4.5. Pre-study and Concomitant Therapy

Concomitant medications, other than study treatment, and medications taken prior to starting study treatment will be summarized for all subjects using the Safety population analysis set by treatment group. Medications are considered concomitant if taken during the treatment period (from the date of first dose through 30 days after the last dose of study drug, except subsequent therapies for prostate cancer). Medications will be summarized by World Health Organization (WHO) drug therapeutic class and generic medication name.

5. EFFICACY

This section outlines the planned analyses of the primary, secondary, and other efficacy endpoints of the study.

Efficacy analyses will be performed in the ITT population, incorporating the randomization stratification factors as documented in the interactive web response system (IWRS). All continuous variables will be summarized using number of subjects (n), mean, standard deviation (SD), median, minimum, and maximum. Categorical variables will be summarized with count and percentage.

Time-to-event endpoints will be summarized using the Kaplan-Meier method [5] and displayed graphically where appropriate. Median event times and two-sided 95% confidence interval (CI) for each treatment group will be provided. Stratified Cox proportional-hazard models will be used to estimate the hazard ratio (HR) and its 95% CI. The stratification factors to be used in the analysis are Gleason score at diagnosis (≤7 vs. >7), geographic regions (NA/EU vs. Other Countries) and prior docetaxel use (yes vs. no).

The testing for the time-to-event endpoints will be based on the stratified log rank test; non-stratified log rank test will be performed as a sensitivity analysis, as appropriate. Multivariate Cox regression analysis, adjusting for important selected prognostic factors (See Section 5.2.1.3 below), will be performed as supportive analyses for rPFS and OS.

The proportional hazard assumption will be assessed graphically by plotting log (-log [estimated survival distribution function]) against log (survival time). The resulting graphs should have approximately parallel lines when the assumption holds. If the proportional hazards assumption is reasonably met, then the HR will be used as an estimate of treatment effect. If the proportional hazards assumption is violated, then the inference remains statistically valid for testing equality in survival distributions, but treatment effect will only be estimated using the median time to event in each treatment group.

Endpoints with binary outcome will be summarized by descriptive statistics (count and percentage) by the treatment group. The relative risk (treatment: control) will be reported along with the corresponding two-sided 95% CI. The two treatment groups will be compared by using the chi-square test; Fisher's exact test may be used if the expected counts in some of the cells are less than 5.

5.1. Analysis Specifications

5.1.1. Level of Significance

In general, a two-sided significance level of $\alpha = 0.05$ will be used for all hypothesis testing and all CIs will be calculated on the two-sided 95% confidence level, unless otherwise specified. [redacted]

5.1.2. Data Handling Rules

In general, no imputation method is planned for handling missing or incomplete data. Sensitivity analyses with censoring rules may be conducted if warranted and will be documented in the clinical study report.

5.2. Primary Endpoints

5.2.1. Dual-Primary Endpoint of rPFS

5.2.1.1. Definition

The dual-primary efficacy endpoint is radiographic progression-free survival (rPFS), as assessed by the investigator, and is defined as the duration from the date of randomization to the date of first documentation of radiographic progressive disease or death due to any cause, whichever occurs first. Radiographic progression will be assessed by soft tissue lesion by computed tomography (CT) or magnetic resonance imaging (MRI) per modified RECIST 1.1[6] or by bone lesion progression on bone scan.

5.2.1.2. Censoring Rules

Subjects that withdraw from the study (i.e., withdrawal of consent, lost to follow-up) or receive selected new systemic anti-cancer therapy without documented disease progression will be censored on the date of the last tumor assessment. Subjects with no evidence of radiographic progressive disease or death will be censored on the date of the last tumor assessment. If there was no tumor assessment performed after the baseline visit, the subject will be censored on the date of randomization. Censoring rules are summarized as shown in Table 2.

Table 2: Censoring Rules

Scenario	Censoring rule
No tumor assessment at Baseline or No tumor assessment after Baseline	Censored on the date of randomization
Subjects who are lost to follow-up or withdraw from study	Censored on the date of the last tumor assessment
Subjects who receive selected new systemic anti-cancer therapy prior to documented disease progression or death	Censored on the date of the last tumor assessment prior to the start of the new systemic anti-cancer therapy or death
Subjects with no evidence of radiographic progressive disease or death	Censored on the date of the last tumor assessment
Subjects with 2 or more consecutive missing assessments followed by evidence of radiographic progressive disease or death	Censored on the date of the last tumor assessment before the missing assessments.

5.2.1.3. Analysis Methods

The primary analysis for the comparison of the rPFS distributions between the two treatment groups will be carried out using the stratified log rank test at a two-sided significance level of 0.005. At the time of primary analysis of rPFS it is projected that approximately 50% of the total number of required events for the OS analysis will be observed. Stratified Cox proportional-hazard model will be used to obtain the HR and its 95% confidence interval. Non-stratified log rank test will be performed as a sensitivity analysis.

To assess the consistency of treatment benefit with respect to the primary efficacy endpoint of rPFS across important subgroups, forest plots will be provided for subgroups as defined in Section 2.4. The comparison between the two treatment groups will be evaluated using the hazard ratio with its 95% CI from a univariate Cox regression model in each subgroup. There is no planned interim analysis on rPFS. Additional rPFS subgroup analyses on subjects with low- or high-volume mHSPC disease may be performed as an update at the time of subsequent OS analysis without alpha spending assigned.

Multivariate Cox regression analysis, adjusting for important selected prognostic factors, will also be performed as supportive analysis, if appropriate. The adjusted hazard ratio and its 95% confidence interval for treatment and each factor will be provided. The following baseline covariates will be considered for inclusion in the model:

- PSA
- Lactate dehydrogenase
- Alkaline phosphatase
- Hemoglobin
- Pain at baseline
- Age
- ECOG PS grade (0 vs. 1)
- Number of bone lesions at baseline (<=10 vs. >10)
- Presence of visceral disease (yes vs. no)
- Receipt of localized therapy (yes vs. no)

- Geographic region (NA/EU vs. other countries)
- Gleason score (<=7 vs. >7)
- Prior docetaxel use (yes vs. no)

5.2.1.4. Audit Plan

To evaluate the lack of bias in the investigator's assessment of rPFS, an audit plan will be implemented. The primary audit plan will utilize the method proposed by Dodd *et al*[7] (NCI method). The audit plan proposed by Amit *et al*[8] (PhRMA method) may be implemented as a supportive plan. All scans will be collected in a central location for a blinded independent review.

Stratified simple random sampling will be used to make sure the sample is truly representative of the entire population. 600(~60%) subjects will be randomly selected for IRC review, stratified by the same factors used for stratified randomization, Gleason score at diagnosis, prior docetaxel use, and geographic region. The SAS code for this is:

proc surveyselect data = all_rand method = SRS rep = **1** sampsize = **600** seed = 12345678 out = random_sample;

id usubjid;

STRATA Gleason docetaxel region ;

run;

Based on the audited sample, we can calculate $\hat{\theta}_{LA}$, the log(HR) based on INV assessment of rPFS, and $\hat{\theta}_{CA}$, the log(HR) based on IRC assessment, and its standard error $\sqrt{\hat{V}_L}$. The correlation ρ between $\hat{\theta}_{LA}$ and $\hat{\theta}_{CA}$, can be estimated through bootstrapping all subjects in the sample with replacement for both INV and IRC. Let $\hat{\theta}_{L\bar{A}}$ be log(HR) based on the INV assessment of the unaudited sample , δ the proportion of the audited sample, $\sqrt{\hat{V}_L}$ the standard error of $\hat{\theta}_L$ based on the overall INV evaluation, then we can estimate $\tilde{\theta}_c$, Log (HR) for full IRC assessment with this formula:

$$\tilde{\theta}_c = \hat{\theta}_{CA} + \hat{\rho}\sqrt{\delta}(1-\delta)\sqrt{\frac{\hat{V}_{CA}}{\hat{V}_L}}(\hat{\theta}_{L\bar{A}} - \hat{\theta}_{LA}) \qquad (2)$$

with variance estimated as $\hat{V}_{CA}$ {1- $\hat{\rho}^2(1-\delta)$ } . These will be used to verify the investigators' assessment.

100% review of the scans may be done upon the request of HAs or the company's decision . Other supportive methods may be applied to the same sample if appropriate or upon request.

For the selected subjects, scans from discontinued subjects will be centrally reviewed when all scans are available, while scans from ongoing subjects may be centrally reviewed on an ongoing basis when they are available . The latest scans close to or on the date of reaching the prespecified rPFS events will be reviewed after that date. Scans after that date won't be centrally reviewed.

5.2.1.5. Sensitivity analysis for the rPFS

Per FDA's request, a sensitivity analysis for rPFS will be performed based on the central review data, where the date of progression is defined as the date of the scan showing >=2 new bone lesions compared to the nadir of bone lesions.

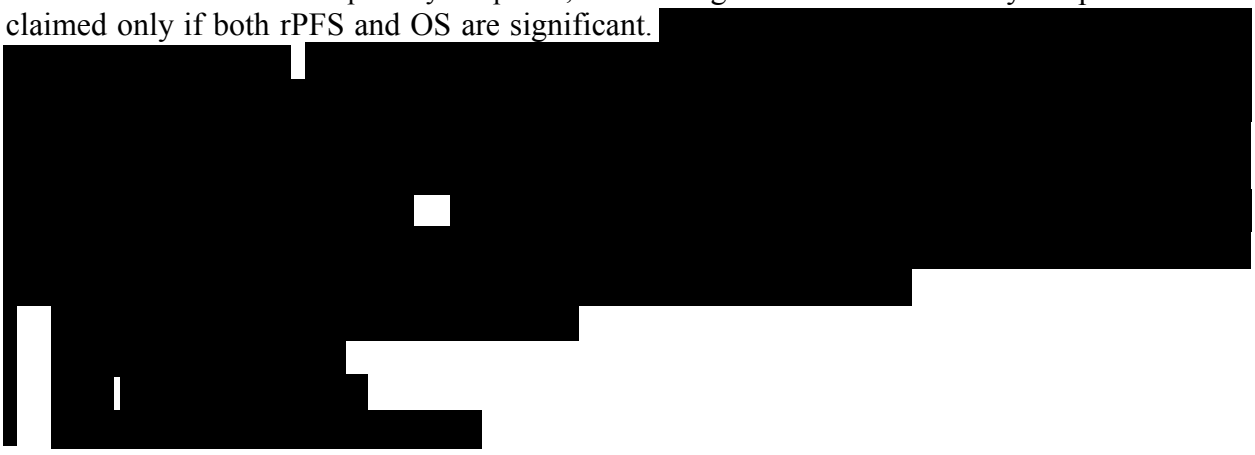
5.2.2. Dual-Primary Endpoint of OS

The efficacy endpoint of OS is defined as the time from date of randomization to date of death from any cause. Subjects alive at time of analysis will be censored on the last date the subject was known to be alive.

The primary analysis for the comparison of the OS distributions between the two treatment groups will be carried out using the stratified log rank test at a two-sided significance level of 0.045. The analysis methods for OS will be similar to those outlined for the analysis of rPFS (see Section 5.2.1.3). In addition, 2-year, 3-year survival rates will be estimated using the Kaplan-Meier method. Subgroup analyses on subjects with low- or high-volume mHSPC disease will be performed without alpha spending assigned.

5.3. Secondary Endpoints

The analyses of the secondary endpoints will be performed at the time of the primary analysis of the dual-primary endpoint of rPFS and the first interim analysis of the dual-primary endpoint of OS using similar methods as the dual-primary endpoints, and the significance of secondary endpoints can be claimed only if both rPFS and OS are significant.



This procedure controls the overall level of significance at the 2-tailed, with α set at 0.05. Subgroup analysis (for high volume and low volume mHSPC subgroups) on the secondary endpoints may be performed for exploratory purposes.

5.3.1. Definition

- Time to pain progression is defined as the time from the date of randomization to the date of the first observation of pain progression. Pain progression is defined as an average increase by 2 points from

baseline to >4 in the BPI-SF worst pain intensity (item 3) with no decrease in opioids confirmed ≥3 weeks apart or initiation of chronic opioids (see definition below), whichever occurs first. Subjects who did not experience pain progression at the time of the analysis will be censored on the last known date with no pain progression. Sensitivity analyses may be performed where increase of pain score to >4 is not required, or confirmation is not required.

- Time to SRE is defined as the time from the date of randomization to the date of the first observation of an SRE. An SRE is defined as the occurrence of symptomatic pathological fracture, or spinal cord compression, or radiation to bone, or surgery to bone. Subjects who did not experience an SRE at the time of the analysis will be censored on the last known date with no SRE.
- Time to chronic opioid use is defined as the time from date of randomization to the first date of confirmed chronic opioid use. For subjects entering the study without receiving opioids, chronic opioid use is defined as administration of opioid analgesics lasting for ≥3 weeks for oral or ≥7 days for non-oral formulations. For subjects entering the study already receiving opioids, chronic opioid use is defined as a ≥30% increase in total daily dose of the opioid analgesics lasting for ≥3 weeks for oral or ≥7 days for non-oral formulations. Subjects with no chronic opioid use at the time of the analysis will be censored on the last known date with no chronic opioid use.
- Time to initiation of cytotoxic chemotherapy is defined as the time from the date of randomization to the date of initiation of cytotoxic chemotherapy. Subjects who did not initiate cytotoxic chemotherapy at the time of the analysis will be censored on the last known date with no cytotoxic chemotherapy.

If there is heavy censoring due to death for some secondary endpoints, sensitivity analyses counting death as event may be performed if deemed clinically meaningful.

5.3.2. Censoring Rules

For all these time-to-events endpoints, subjects with no on-study assessment or no baseline assessment will be censored on the date of randomization (see Sections 5.2.1.2 and 5.3.1).

5.3.3. Analysis Methods

The testing of these secondary efficacy endpoints will be based on the stratified log rank test. Secondary efficacy endpoints will be summarized using the Kaplan-Meier method and displayed graphically where appropriate. Cox proportional hazard models will be used to estimate the HR and its 95% confidence interval.

5.4. Additional Analysis for OS Endpoint

The following sensitivity analyses for the OS may be carried out as appropriate if it is deemed useful to aid in the interpretation of the results.

5.4.1. Covariate Effects

Multivariate analysis of OS adjusting for baseline prognostic factors will be performed as outlined in Section 5.2.1.3. The results may be used to support the findings obtained from the primary analysis. In the event that some of the prognostic factors are not highly prognostic for this endpoint, additional exploratory analysis may be conducted by first examining each baseline factors in a univariate analysis. The factors with p-value ≤0.05 will be included in the multivariate model.

5.4.2. Subsequent Therapies

To assess the impact of the use of subsequent therapies on the treatment effect on the OS, a time-dependent analysis using Cox regression may be performed. A stratified Cox model will include treatment variable and a time-dependent covariate for the status change with respect to receiving

subsequent therapy. Stratified hazard ratio and its 95% confidence interval for treatment will be provided. This analysis may be conducted if a significant number of subjects receive subsequent therapies.

5.4.3. Crossover

In the event that a large number of subjects in the placebo group crossover to an AR signaling inhibitor, at least one of the following analyses may be used, if appropriate, in estimating the true treatment effect on OS:

- The rank preserving failure time model as described by Robins and Tsiatis[10]
- Inverse Probability Censoring Weighted (IPCW) log-rank test as described by Cole and Hernan[11]
- Iterative Parameter Estimate (IPE) method as described by Branson and Whitehead[12]

5.5. Other Endpoints

5.5.1. Definition

Other endpoints are defined in the following table.

Table 3: Other Endpoints

Endpoint	Description	Analysis Population
Patient Reported Outcomes and Analgesics Use	The patient-reported outcomes will include the following: • FACT-P: FACT-P total score, each subscale score (PWB, SFWB, EWB, FWB, FACT-G, PCS), and TOI. • BPI-SF: 4 pain severity items (worst, least, average and now) and 7 pain interference items • BFI: 3 severity items (now, usual and worst) and 6 fatigue interference items • EQ-5D-5L: 5 domains and EQ visual analogue scale (VAS) • Analgesics use See Section 5.5.3 for detailed description of endpoints and analysis.	PRO Population
Time to symptomatic local progression	The time from randomization to the date of the first occurrence of urethral obstruction or bladder outlet obstruction. Subjects with no symptomatic local progression at the time of analysis will be censored on the last known date the subject was known to have not progressed.	ITT Population
Time to PSA progression	The time from the date of randomization to the date of PSA progression based on Prostate Cancer Working Group 2 (PCWG2) criteria.[13] Subjects with no PSA progression at the time of analysis will be censored on the last known date with no PSA progression.	ITT Population
Prostate cancer-specific survival	The time from randomization to the date of death from prostate cancer-specific cause. Subjects alive at time of analysis will be censored on the last date the subject was known to be alive or lost to follow-up. Subjects who die from other than prostate cancer-specific causes at time of analysis will be censored on the death date.	ITT Population

Table 3: Other Endpoints

Endpoint	Description	Analysis Population
PFS2	The time from date of randomization to date of disease progression on the first subsequent therapy for prostate cancer or death, whichever occurs first. Subjects with no disease progression on the first subsequent therapy will be censored on the last known date the subject was known to have not progressed. Subjects not on subsequent therapy will be censored on the last known date of no progression.	ITT Population
Time to ECOG PS deterioration	The time from the date of randomization to the first date of deterioration in ECOG PS (defined as the worsening of ECOG PS grade by at least one point). Subjects who have no deterioration in ECOG PS grade at the time of analysis will be censored at the last known date of no deterioration. Subjects with no on-study assessment or no baseline assessment will be censored at date of randomization.	ITT Population

5.5.2. Analysis Methods

Analyses of the other endpoints will be performed on the ITT population, unless otherwise specified. Analysis of time-to-event endpoints and endpoints with binary outcome will be carried out as described above (see start of Section 5).

5.5.3. Patient Reported Outcome (PRO) and Analgesics Use

A separate statistical analysis plan will be provided for PRO data.

6. SAFETY

6.1. Adverse Events

Subjects will be assessed for AEs at each monthly clinic visit while on the study. Adverse events will be graded according to the NCI-CTCAE Version 4.03 and coded to preferred term and system organ class (SOC) using the most recent version of MedDRA.

All AEs reported during the AE reporting period (inclusive of the 30-day post last dose of study drug period) will be considered as treatment-emergent adverse events and will be summarized by treatment group as treated using all subjects in the safety population.

For each treatment group, AE incidence rates will be summarized with frequency and percentage by SOC and preferred term, with all subjects treated in that treatment group as the denominator, unless otherwise specified. In addition, AE incidence rates will also be summarized by severity and relationship to study drug. Treatment-related AEs are those judged by the Investigator to be at least possibly related to the blinded study drug. Subjects with multiple occurrences of events will only be counted once at the maximum severity to study drug for each preferred team, SOC, and overall. Deaths that occur within 30 days after the last dose of study drug are defined as on-study deaths.

AE onset date will not be imputed if it's completely missing. Partial AE onset date will be imputed as follows:

- If the AE onset date is missing day only, it will be set to:
    - the first day of the month that the AE occurred, if the month/year of the onset of AE is different from the month/year of the study treatment start;
    - the day of the study treatment start, if the month/year of the onset of AE is the same as month/year of the study treatment start but different from the month/year of the AE resolution date;

  - the day of the study treatment start or the day of AE resolution date, whichever is earlier if the month/year of the onset of AE and month/year of the study treatment start and month/year of the AE resolution date are same;

- If the AE onset date is missing both day and month, it will be set to the earliest of:
  - January 1 of the year of onset, as long as this date is after the study treatment start;
  - the day and month of the study treatment start, if this date is the same year as that the AE occurred;
  - the AE resolution date.

Summary tables of the following AEs will be provided:

- Overall summary of AEs: the number and percentage of subjects who experienced any AE, any serious adverse event (SAE), any treatment-related AE, any treatment-related SAE, any discontinuations due to an AE, and any deaths
- All AEs by SOC and preferred term All AEs by SOC, preferred term, and toxicity grade
- Grades 3 or 4 AEs by SOC and preferred term
- Treatment-related AEs by SOC and preferred term
- Treatment-related AEs by SOC, preferred term, and toxicity grade
- Treatment-related Grades 3 or 4 AEs by SOC and preferred term
- AEs that led to study drug discontinuation by SOC and preferred term. Study drug discontinuation will be determined from the treatment disposition CRF (where reason for discontinuation is "Adverse Event") and the specific AE will be determined from the AE CRF page (where action taken is "Drug withdrawn")
- AEs that led to study drug discontinuation by SOC, preferred term, and toxicity grade
- AEs that led to dose modification by SOC, preferred term, and toxicity grade
- All SAEs by SOC and preferred term
- All SAEs by SOC, preferred term, and toxicity grade
- Deaths will be summarized by time period (on-study vs. during follow-up) and cause of death.

Subject listings of all AEs by toxicity grade, Grades 3 or 4 AEs, all SAEs, AEs that led to study drug discontinuation, dose modification, and all deaths will be provided as well.

Narratives will be written for the following subjects in the final clinical study report:

- Subjects who die ≤ 30 days after the last dose of study drug
- Subjects who discontinue study drug due to treatment-emergent AE
- Subjects who experience a treatment-emergent SAE
- Subjects who experience seizure
- Subjects who experience Grade 3 or higher treatment-emergent major adverse cardiovascular events (MACE) including ischemic heart disease, cardiac failure, cerebrovascular disease, and arrhythmias
- Subjects who experience treatment-emergent primary cancers

Subjects who experience other Grade 3 or higher treatment-emergent adverse events of special interest.

6.2. Clinical Laboratory Tests

Only data collected by the central laboratory will be summarized. Local laboratory data collected for the purposes of planning treatment administration, dose modification, or monitoring adverse events, will not be summarized but will be displayed in the lab listings if needed.

Normal ranges will be used to identify values that are outside the normal ranges and abnormal laboratory results will be graded according to the NCI-CTCAE Version 4.03.

Descriptive statistics will be provided for each test result and for the change from baseline by visit.

A shift summary of baseline grade by maximum post baseline NCI-CTCAE toxicity grade will be presented, as appropriate for selected parameters. For each laboratory parameter, the baseline laboratory value will be defined as the last laboratory value collected on or prior to the date of the first dose of study drug. A listing of subjects who develop toxicities of Grade ≥ 3 will be provided for each lab type.

Liver function test data will be summarized based on eDISH and Hy's Law criteria [16]. Listings will be provided for subjects who meet the eDISH criteria and Hy's Law criteria, respectively.

6.3. Vital Signs

Each vital sign (temperature, blood pressure (systolic and diastolic), respiration rate, and heart rate) and change from baseline for blood pressure will be summarized and presented by treatment group and study visit. Subjects with clinically significant abnormalities in vital signs as compared to baseline will be listed.

Data will be summarized and presented according to the following categories:

Parameter	Criteria for Clinically Significant Abnormality
Systolic Blood Pressure	Absolute result > 160 mmHg and increase from baseline > 20 mmHg
	Absolute result < 90 mmHg and decrease from baseline > 20 mmHg
Diastolic Blood Pressure	Absolute result > 100 mmHg and increase from baseline > 10 mmHg
	Absolute result < 50 mmHg and decrease from baseline > 10 mmHg
Weight	5 - < 10% weight loss from baseline
	10 - < 20% weight loss from baseline
	>= 20% weight loss from baseline
	5 - < 10% weight gain from baseline
	10 - < 20% weight gain from baseline
	>= 20% weight gain from baseline

6.4. Electrocardiogram

Electrocardiograms (ECGs) (12-lead) will be recorded at Screening. Abnormalities noted at screening will be included in the medical history.

7. PHARMACOKINETICS

Pharmacokinetic assessment (trough PK samples) will be conducted. Pre-dose blood samples for analysis of apalutamide and the active metabolite JNJ-56142060 concentrations will be collected on Day 1 of Cycles 2, 3, 4, 5, and 6.

Descriptive statistics including mean and geometric mean values will be summarized.

Pharmacokinetic samples will be collected from at least 60 subjects who received or will receive leuprolide acetate as the GnRHa at the time of randomization. Samples will be collected on Day 1 of Cycles 1, 3, 4, 5, and 6 for analysis of leuprolide and testosterone concentrations.

Descriptive statistics of leuprolide PK data will be summarized by treatment group and dose of leuprolide acetate. Statistical analysis comparing leuprolide concentrations when administered (with placebo) alone versus in combination with apalutamide will be performed. The percentage of subjects with testosterone concentrations <50 ng/dL) will also be summarized descriptively by treatment group.

The results of these analyses will be provided in a separate report.

8. BIOMARKERS

Exploratory biomarker analysis of AR anomalies and other markers previously shown to be responsible for resistance to apalutamide will be conducted and will be reported in a separate document. To determine the appropriate biomarker threshold, statistical method for evaluating treatment effect on biomarker subgroup may be utilized as appropriate. [17]

9. MEDICAL RESOUURCE UTILIZATION

Data collected on medical resource utilization (MRU) will be used in the construction of economic model. The modeling analysis and reporting will be summarized in a separate report.

REFERENCES

1. Gravis G, Fizazi, F, Joly F, et al. Androgen-deprivation therapy alone or with docetaxel in non-castrate metastatic prostate cancer (GETUG-AFU 15): a randomised, open-label, phase 3 trial. *Lancet Oncol* 2013;14:149-158.
2. James ND, ND, Sydes MR, Clarke NW et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet* 2016;387:1163-1177.
3. Sweeney C, Chen Y-H, Carducci MA et al. Chemohormonal therapyin metastatic hormone-sensitive prostate cancer*New England Journal of Medicine* 2015; 373: 737-746.
4. Wang SK and Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics*, 1987; 43: 193-199
5. Kaplan EL and Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53:457-81
6. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45:228-47
7. Dodd LE, Korn EL, Freidlin B, Gray R, Bhattacharya S. An audit strategy for progression-free survival. *Biometrics* 2011;67: 1092–9.
8. Amit O, Mannino F, et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis. *European Journal of Cancer* 2011; 47: 1772-1778
9. Hochberg, A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; 75(4):800-802
10. Robins JM and Tsiatis AA. Correcting for non-compliance in randomized trials using Rank Preserving Structural Failure Time models. *Commun Statist Theory Meth* 1991; 20:2609-2631
11. Cole SR and Hernán MA Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine* 2004; 75:45-49
12. Branson M and Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. *Statistical Medicine* 2002; 21: 2449-2463.
13. Scher HI, Halabi S, Tannock I, et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: Recommendations of the prostate cancer clinical trials working group. *J Clin Oncol* 2008; 26:1148-1159.
14. http://www.mdanderson.org/education-and-research/departments-programs-and-labs/departments-and-divisions/symptom-research/symptom-assessment-tools/BPI_UserGuide.pdf
15. Cella D, Nichol MB, et al. 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* 2009;12(1): 124-129.
16. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Food and Drug Administration. Guidance for industry. Drug-induced liver injury: premarketing clinical evaluation. Drug Safety. 2009.
17. Jiang W, Freidlin B and Simon R. Biomarker Adaptive Threshold Design: A Procedure for Evaluating Treatment with Possible Biomarker-defined Subset Effect. *Journal on National Cancer Institute* 2007; 99:1036-1043.
18. Wiens BL and Dmitrienko A. The fallback procedure for evaluating a single family of hypotheses. Journal of Biopharmaceutical Statistics. 2005; 15(6): 929–942.

ATTACHMENT 1: ADDITIONAL EXPLORATORY ANALYISIS TO SUPPORT HEALTH ECONOMICS, MARKET ACCESS AND REIMBURSMENT (HEMAR)

1. DEFINITION OF SUBGROUPS

Analysis will be performed to determine whether the treatment effect is consistent among subgroups. Analysis will be conducted for the ITT population and for the following subgroups

- For subjects that are considered high risk defined as having two of the three following high-risk factors associated with poor prognosis:
  - A Gleason score of 8 or more on a scale of 2 to 10, with higher scores indicating more aggressive disease
  - At least three bone lesions
  - The presence of measurable visceral metastasis.

- For subjects not fulfilling high risk criteria and considered low risk

2. TIME TO EVENT ENDPOINTS FOR SUBGROUP ANALYZES

Kaplan- Meier estimates will be used to estimate the distribution of time to event by treatment arm based on all the ITT subjects. Data will be calculated and summarized with descriptive statistics. The following time to vent endpoints will be analysis by predefined subgroups as defined in section 1.

- rPFS
- Time to pain progression
- Time to chronic opioid use
- Time to chemotherapy
- OS
- EQ-5D and FACT P

3. EXPOSURE ADJUSTED INCIDENCE RATES (EAIR)

3.1. Restriction on the first event

The analysis restricts on the occurrence of the first event per patient and ignores the existence of later (multiple) events as these cannot be assumed to occur independent of previous events (e.g.: patients suffering from infections may have in general a higher risk of having other complications and may even have a higher risk of getting other infections). The occurrence of multiple events is subject to another analysis considering the absolute number of adverse events per patient.

For these reasons the EAIR should be interpreted as 'rate until the first event occurs'. Rates estimated from several patients can be averaged on the level of a preferred term (PT), of a system organ class (SOC), or on a global level (see below).

The interpretation of EAIRs is simple and consistent on the preferred-term level only, and can be expressed as "Average number of TEAEs per preferred-term emerging per person-month of exposure".

The aforementioned considerations apply in the same way to EAIRs estimated on the global level: when EAIRs are collapsed into the global estimate (first analyses), the estimate can be interpreted as the "Average number of TEAEs emerging per person-month and PT", because estimation has been performed on a 'per PT'-basis (per average or typical PT among all PTs). Comparing EAIRs on the level of the SOC or on the global level involves data destruction because a patient's information is reduced to the first TEAE only (and possibly to a TEAE of marginal relevance among many TEAEs with higher clinical relevance).

The EAIR analysis focuses on the 'speed' by which TEAEs emerge. The analysis restricts on the first event of a patient because independence of TEAEs cannot be assumed. The necessity to restrict on the first event entails considerable data destruction when deriving SOC-specific EAIRs or the EAIR on a global level. To overcome this, the 'per PT'-analysis, which is reported in both Tables identically, is preferable.

Comparing EAIRs between the analyses outlined below on a SOC-specific or a global level demonstrates that the 'per PT'- method makes the interpretation of results more difficult. However, it can be suggested that this method provides a more robust approach when the two treatment arms are to be compared on a SOC-specific or global level. A t-Test like comparison of PT-specific estimates between the two treatment arms may provide a more robust, comprehensive and easy-to-communicate way of visualizing and comparing results.

3.2. Duration of exposure: censored & non-censored

The incidence rate for a patient is derived from the duration of exposure to treatment of that patient. When averaging incidence rates a patient's duration of exposure is given either A) by the time when the event has occurred (non-censored data), or B) by the total duration of treatment in case the patient does not show the adverse event in question (censored data). Depending on whether a patient has an adverse event or not, the duration of exposure enters the denominator in its non-censored or censored form, respectively.

3.3. Incidence rate per patient

The incidence rate for a specific event of a patient i is the reciprocal of time t when the first event occurs:

$EAIRi=1ti$.

3.4. Average EAIR

The EAIR averaged over all patients is

$EAIR=\Sigma TEAEini=1\Sigma tini=1$,

whereby

a) a TEAE enters the sum in the nominator unweighted (TEAEi =1, otherwise TEAEi =0), and

b) the duration of exposure enters the denominator as described before: $ti=\{$ $time\ of\ TEAE$ if occurring (non-censored data) $total\ duration\ of\ treatment$ if no event occurs (censored data).

3.5. EAIRs on the level of a SOC and on the global level on a 'per-PT' basis

3.5.1. Average EAIR per PT

The EAIR for a specific PT is an average over all patients, i. e.

$EAIRPT = \Sigma TEAEPT,ini = 1 \Sigma tPT,ini = 1$,

whereby the number of TEAEs and durations of exposure enter the nominator and the denominator.

3.5.2. Average EAIR per SOC

The average EAIR per SOC considers the first event of each patient within the SOC. The denominator includes the exposure time of each adverse event of all PTs within the SOC, per patient, i. e.

$EAIRSOC = \Sigma TEAESOC,i \Sigma 1 tPT,in\ PTs\ per\ SOCPT = 1 ni = 1$,

where TEAESOC,i is the first event per patient per SOC and tPT,i is the exposure time for a specific preferred term of a given patient.

Note: This EAIR is an incidence rate per average (or typical) preferred term in that SOC (cf. 3.6.1).

3.5.3. Average EAIR on a global level

The average EAIR on a global level only considers the first event per patient across all events. The denominator includes the exposure times of all PTs, i. e.

$EAIRglobal = \Sigma TEAEi \Sigma 1 tPT,in\ PTsPT = 1 ni = 1$,

where TEAEi is the first event of a patient overall and the tPT,i's are PT-specific exposure times of that patient.

Note: This EAIR is an incidence rate per average (or typical) preferred term.

3.6. Second analyses

3.6.1. Average EAIR per PT

The EAIR for a specific PT is an average over all patients as described before, i. e.

$EAIRPT = \Sigma TEAEPT,ini = 1 \Sigma tPT,ini = 1$,

whereby the number of TEAEs and durations of exposure enter the nominator and the denominator.

3.6.2. Average EAIR per SOC

The average EAIR per SOC considers the first event per patient per SOC only, and only one (the corresponding) exposure time in the denominator (confer before, where the denominator in the EAIRSOC depends on the number of PTs per SOC):

$EAIRSOC = \Sigma TEAESOC,ini = 1 \Sigma tSOC,ini = 1$,

Note: This EAIR is an incidence rate per SOC.

3.6.3. Average EAIR on a global level

The average EAIR on a global level considers the overall first event per patient only, and only one (the corresponding) exposure time in the denominator (confer before, where the denominator in the EAIRSOC depends on the overall number of PTs):

$EAIRglobal = \Sigma TEAEini = 1 \Sigma tini = 1$,

whereby TEAEi represents the first TEAE among all TEAEs of patient i and ti as before (time when TEAE occurs (non-censored data) or total duration of treatment if no event occurs (censored data))