TRANSFORMER Trial Statistical Analysis Plan

Full Title of Study	The TRANSFORMER Trial (Testosterone Revival Abolishes Negative Symptoms, Fosters Objective Response and Modulates Enzalutamide Resistance)
Protocol Number	J14146
Protocol Version	1.10 (03 Nov., 2014)
IND Reference Number	N/A
Version of Plan	1.0
Date of Plan	27 March 2017 Replaces 28 October 2016
Sponsor Name	**Department of Defense (DOD)** **Prostate Cancer Research Program**

Revision History		
Revision	**Date**	**Description of Change**
1	March 27, 2017	
		Primary endpoint changed to PFS, including both radiographic and clinical progression
		Definition of radiographic progression clarified
		Secondary objectives revised
		Sample size changed to be consistent with protocol version 5.0

List of Abbreviations

Abbreviation	Full Term
BAT	Bipolar Androgen Therapy
BPI	Brief Pain Inventory
BSA	Body Surface Area
CR	Complete Response
CRPC	Castration-Resistant Prostate Cancer
CTC	Common Terminology Criteria
DMC	Data Monitoring Committee
ECOG	Eastern Cooperative Group
eCRF	Electronic Case Report Form
FACIT-F	Functional Assessment of Chronic Illness Therapy-Fatigue
HR	Hazard Ratio
IIEF	International Index of Erectile Function
I-PANAS-SF	International Positive and Negative Affect Schedule Short-Form
ITT	Intent to Treat
NCI	National Cancer Institute
ORR	Objective Response Rate

PCWG2	Prostate Cancer Working Group 2
PPP	Per Protocol Population
PR	Partial Response
PSA	Prostate-Specific Antigen
RECIST	Response Evaluation Criteria in Solid Tumors
rPFS	radiographic Progression-Free Survival
SAP	Statistical Analysis Plan
T	Testosterone
TRANSFORMER	Testosterone Revival Abolishes Negative Symptoms, Fosters Objective Response and Modulates Enzalutamide Resistance
WBC	White Blood Cell

1.0 Introduction

This draft Statistical Analysis Plan (SAP) describes the proposed statistical analysis of the study entitled: The TRANSFORMER Trial (Testosterone Revival Abolishes Negative Symptoms, Fosters Objective Response and Modulates Enzalutamide Resistance).

The purpose of this document is to ensure the credibility of the study outcomes by pre-specifying the statistical approaches and data handling conventions for key analyses. This draft plan will focus on the analysis of the primary endpoint, radiographic Progression-Free Survival (rPFS), and the secondary endpoints, PSA response rate, objective response rate (ORR), time to PSA progression, quality of life, and safety according to the NCI CTC, version 4. The primary analyses of these endpoints will be described, the populations for analysis defined, and specific rules for "data handling" relevant to undertaking the key analyses will be described.

2.0 Study Overview, Objectives and Endpoints

2.1 Overview of Study

This study will treat asymptomatic men with progressive metastatic castration-resistant prostate cancer (CRPC) after treatment with abiraterone acetate). This study is a randomized, multi-institutional, open label study to determine if treatment with intramuscular supraphysiologic testosterone (T) given on a dose/schedule designed to result in rapid cycling from the polar extremes of supraphysiologic to near castrate levels [i.e. Bipolar Androgen Therapy (BAT)] will improve primary and secondary objectives vs. enzalutamide.

2.1.1 Stratification and Randomization Scheme

At randomization, patients will be stratified based on duration of prior abiraterone acetate therapy (6 months or less or greater than 6 months). Only patients determined to be eligible following completion of the Randomization eCRF will be randomized. A dynamic balancing randomization algorithm will be used to ensure assignment of the treatments is balanced across the stratification factor. This procedure balances the marginal distribution of the

stratification factor between the treatment regimens. The approach is based on the method described by Pocock and Simon (Pocock and Simon, 1975).

2.2 Study Objectives: Primary

The primary objective of the study is to determine if treatment with supraphysiologic testosterone (T) will improve PFS compared to Enzalutamide in men with CRPC post-treatment with abiraterone.

2.3 Study Objectives: Secondary

Secondary objectives are to investigate:

- the safety of cyclical parenteral testosterone in asymptomatic men with recurrent castrate resistant prostate cancer,
- PSA response rate prior to crossover,
- PSA response rate in patients that choose to crossover,
- Radiographic Progression-free Survival (rPFS)
- Objective response rate,
- Time to PSA progression,
- Time to initiation of docetaxel chemotherapy
- Comparison of quality of life (QoL) as assessed by multiple validated instruments.
- The effect of BAT vs. Enzalutamide on expression of AR- V7 and frequency of AR-mutation.

2.4 Clinical Trial Endpoints

2.4.1 Radiographic Progression-Free Survival (rPFS)

Radiographic progression-free survival (rPFS) is defined as the time from the date of randomization to the date of first documented radiological confirmation of progression (per RECIST 1.1 for soft tissue or Prostate Cancer Working Group 2 (PCWG2) for bone lesions) or death, whichever occurs first, before the

patient begins alternate anti-cancer treatment or crossover treatment. The standard RECIST 1.1 criteria for soft tissue are described in detail in the Protocol Appendix 1. On bone scan, radiographic progression will be defined by PCWG2 criteria as ≥2 new bone lesions (Appendix 7 of the Protocol). Specifically, for the first reassessment scan on treatment only (Cycle 3 of treatment), the patient is kept on study for 3 more cycles and have a confirmatory scan performed 12 weeks (3 cycles) later. If this confirmatory scan shows 2 or more additional new lesions, this defines progression; the progression date is the date of the first reassessment scan. Subjects who have unconfirmed progression will continue the treatment. If progression is observed on subsequent scans, then the date of that scan will be the date of progression and subsequent radiologic confirmation is not required. If there are inconsistencies between the progression date reported on the crossover treatment form and the disease assessment form, the progression date from the disease assessment form prior to the crossover treatment date will be used. If a patient has not had an event at the date of the analysis cut-off, initiation of crossover treatment, or alternate treatment, rPFS will be censored at the time of the last tumor assessment before the cut-off date. Subjects who die without a reported radiographic progression while on study drug (and before crossover) will be considered to have progressed on the date of their death. Subjects who do not progress or die while on study drug or before crossover will be censored on the date of their last evaluable tumor assessment prior to the initiation of subsequent anti-cancer therapy or crossover. Subjects who do not have any on-study tumor assessments and did not die will be censored on their date of randomization.

2.4.2 **Progression-Free Survival (PFS)**

Patients who do not meet the criteria of radiographic progression who are removed from study for worsening symptoms that are attributable to prostate cancer progression will be considered to have clinical progression.

Clinical progression will be defined as documentation in the CRF of any of the following (whichever occurs earlier):

- Cancer pain requiring initiation of chronic administration of opiate analgesia (oral opiate use for ≥3 weeks; parenteral opiate use for ≥7 days. Patients with cancer pain requiring opiate analgesia for relief

should also be assessed by the investigator for the need for initiating systemic chemotherapy or palliative radiation.

• Development of a skeletal-related event (SRE): pathologic fracture, spinal cord compression, or need for surgical intervention or radiation therapy to the bone.

• Development of clinically significant symptoms due to loco-regional tumor progression (e.g. urinary obstruction) requiring surgical intervention or radiation therapy.

PFS is defined as the time from randomization to the earliest of clinical progression, radiographic progression or death, before the patient starts alternate anti-cancer treatment or crossover treatment. Patients alive without clinical or radiographic progression will be censored at the date of their last tumor assessment prior to initiation of alternate anti-cancer therapy and the cut-off date.

2.4.3 Time to initiation of Docetaxel chemotherapy

Time to Initiation of Docetaxel is defined as the time from the date of randomization to the date of initiation of Docetaxel therapy. If a patient is known to be alive at the time of analysis cut-off without previously starting docetaxel therapy, this endpoint will be censored at the last date the patient is known to be alive. Patients will be censored at the date of randomization if no additional data are obtained.

2.4.4 Objective Response Rate (ORR)

Objective response rate (ORR) will be measured in patients with measurable disease using RECIST criteria. The number of patients in each arm achieving at least partial response (PR) prior to removal from the study, radiographic progression, or death will be divided by the number randomized to each arm. Patients with no response assessments will be analyzed as non-responders.

2.4.5 PSA Response Rate

The PSA response rate will be calculated in each arm as the number of subjects with a PSA response (decrease > 50% from baseline, confirmed with a second measurement at least 4 weeks after the first measurement) prior to

radiographic progression or crossover divided by the number of subjects randomized.

For patients that crossover following radiographic progression, the PSA response rate after crossover will be calculated as the number of subjects in each crossover arm with a PSA response after initiation of crossover treatment divided by the number of subjects in each crossover arm.

2.4.6 Time to PSA Progression

The time to PSA progression will be calculated as the number of days from randomization to the earliest of PSA progression, death from any cause, crossover treatment or removal from the study prior to completion of randomized study treatment. PSA progression will be defined per PCWG2 criteria as PSA increasing to 25% above baseline level and a minimum of 2 ng/ml, confirmed with another measurement at least 4 weeks later (see Protocol Section 6.2). Death, treatment crossover and removal from the study will be treated as competing events. Patients without a PSA progression or competing event at the time of analysis will be censored at the last PSA assessment date. Patients with no post-baseline PSA assessments will be censored at the date of randomization.

2.4.7 Safety Endpoints

The overall safety profile and toleration of BAT and enzalutamide will be characterized by incidence of adverse events categorized by type, frequency, severity, timing and relationship to study therapy as well as laboratory abnormalities. Adverse events will be summarized by the frequency of patients experiencing treatment emergent adverse events by body system and preferred term, by worst NCI CTCAE grade (version 4.0).

2.4.8 Quality of Life (QoL) Endpoints

QoL will be assessed using RAND-SF36 Quality of Life Survey, FACIT-F Version 4, I-PANAS-SF. Sexual capacity, functional capacity and pain will be assessed using IIEF, Global Assessment of Change and BPI, respectively. For each instrument, summary statistics of the scores will be reported at baseline randomization, one month post-randomization, three months post-randomization, six months, and 12 months post-randomization. In each arm changes in quality of life scores pre- versus post-treatment will be evaluated at each follow-up time by paired-sample t-tests or Wilcoxon signed rank tests as

appropriate. In addition, mixture effect models may be fitted for accessing the quality of life changes over time.

2.4.9 Correlative Endpoints

AR variant status (detectable vs. not detectable) will be determined at baseline.

3.0 Populations For Analysis

Figure 3-1 describes the respective patient populations being defined for the study in a flow chart format.

3.1 Screened Population

The screened population is defined as all patients who have signed an informed consent and participated in screening procedures to assess eligibility.

3.2 Intent to Treat (ITT) Population

The Intent-to-Treat population includes all patients who have been randomized to receive study treatment; this is the analysis population for all efficacy endpoints as well as patient characteristic summaries. The treatment group used in the analyses will be the treatment assigned by randomization.

3.3 Safety Population

The safety population is defined as all ITT patients who receive at least one dose of study treatment. This population will be used for all safety analyses. The treatment group used in the analyses will be the actual treatment received.

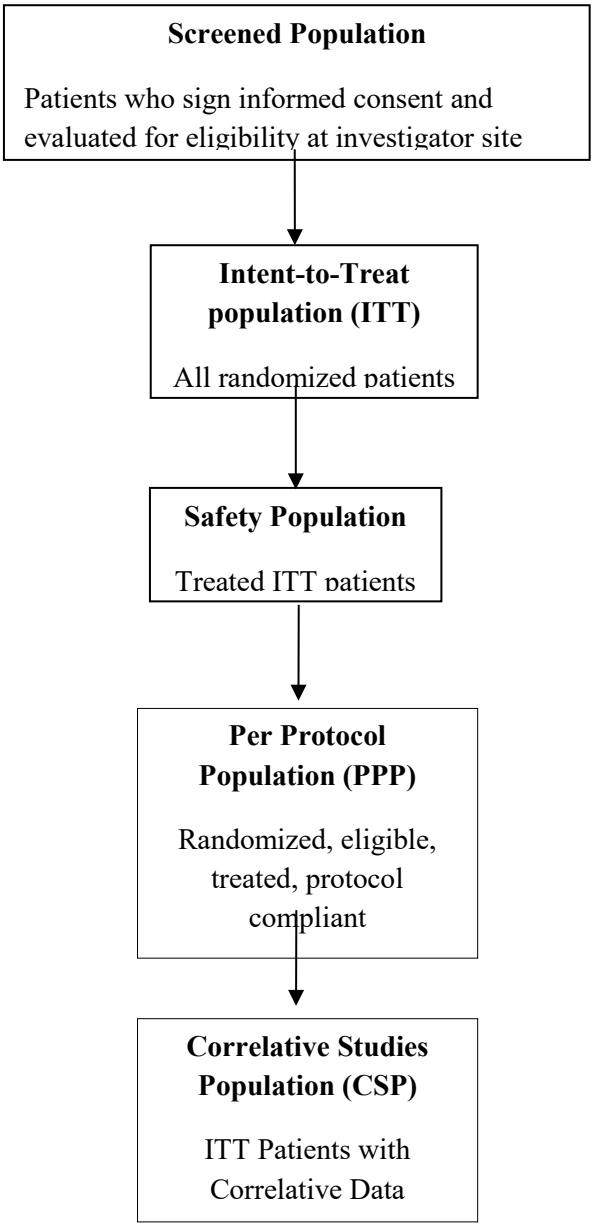
3.4 Per Protocol Population (PPP)

The Per Protocol Population (PPP) is defined as the subset of ITT patients who meet all inclusion and exclusion criteria (i.e. eligible patients). and are treated according to their randomization assignment, and protocol compliant. Patients with major protocol violations will be excluded from this population. All major violations will be determined before treatment group unblinding.

3.5 Correlative Studies Population (CSP)

The Correlative Studies Population (CSP) is defined as the subset of ITT patients who have non-missing correlative data.

Figure 3-1: Patient Disposition for Protocol Defined Populations



4.0 Statistical Methods and Determination of Sample Size

4.1 Determination of Sample Size

This trial is designed to detect a 50% improvement in median PFS in the BAT group compared to enzalutamide, from 6 to 9 months (corresponding to a hazard ratio (HR) of 0.667 for BAT vs. enzalutamide), at one-sided significance level of 0.05. Allowing for the interim analysis plan discussed below, the study requires at least 156 events to ensure a sequential test procedure will have 80% power. With an expected recruitment period of 24 months and an additional 12 months of follow-up, and accounting for 15% loss to follow-up, a total of 194 patients (97 per arm) will be randomized to observe 156 events in the 36 total months of this study. This calculation was performed using EAST® software.

4.2 Interim Analyses

Two interim analyses for PFS are planned that allow the study to stop early for efficacy. An α-spending function method of Lan and DeMets with an O'Brien-Fleming-type stopping boundary (1983), will be used to construct the stopping boundaries for the efficacy analyses. At each interim analysis, a one-sided stratified log-rank test will be used to compare the experimental arm (BAT) with the control arm (enzalutamide) and a hazard ratio will be calculated. The study will also be monitoring for early stopping for futility using Jennison-Turnbull repeated confidence interval methodology (1993). At each interim analysis the nominal (1- 2 × α) confidence interval on the PFS hazard ratio comparing the BAT arm to the enzalutamide arm (arm A versus arm B) will be computed, where α is the nominal one-sided significance level of the boundary from the error spending function at the information fraction for the particular analysis time. If the confidence interval does not contain the target alternative HR of 0.667, then the Data Monitoring Committee may consider terminating the study early for lack of treatment differences. The monitoring plan has negligible impact on the power of the study.

4.2.1 First Interim Analysis for PFS

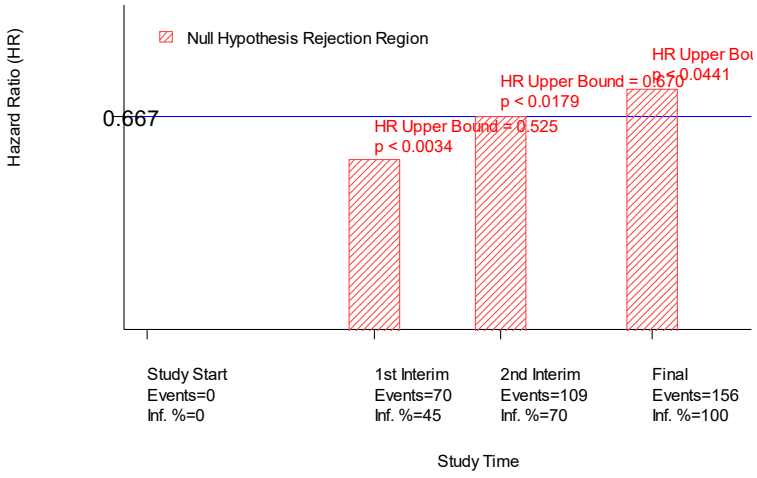
The first interim analysis will be conducted after 70 events (approximately 45% of the information) have been entered in the database. Based on 70 events, a one-sided p-value < 0.0034 from a stratified log-rank test will be sufficient to declare superiority of BAT compared to enzalutamide. Under the assumption of exponentially distributed survival times, this p-value corresponds to a hazard ratio less than or equal to 0.525. Additionally, a two-sided nominal 99.3%

confidence interval for the hazard ratio will be computed. If the lower bound of this interval is greater than the target alternative hazard ratio of 0.667, then the data monitoring committee may consider terminating the study early for futility.

4.2.2 Second Interim Analysis for PFS

The second interim analysis will be conducted after 109 events (approximately 70% of the information) have been entered in the database. Based on 109 events, a one-sided p-value < 0.0179 from a stratified log-rank test will be sufficient to declare superiority of BAT compared to enzalutamide. Under the assumption of exponentially distributed survival times, this p-values corresponds to a hazard ratio less than or equal to 0.670. Additionally, a two-sided nominal 96.4% confidence interval for the hazard ratio will be computed. If the lower bound of this interval is greater than the target alternative hazard ratio of 0.667, then the data monitoring committee may consider terminating the study early for futility.

Figure 4.1: Interim Analyses and Final Analysis Details for PFS



4.2.3 Toxicity Monitoring

The adverse event of pain on the BAT arm will be monitored after every cohort of 15 patients has been treated for 12 weeks. If the toxicity of grade 3 or higher pain appears

in over 20% of the BAT-treated patients, the trial will temporarily be halted pending safety evaluations. Specifically, we will apply a Bayesian toxicity monitoring rule that suspends the enrollment if the posterior probability of risk of pain being larger than that threshold is at least 75%. Based on the data of previous studies the monitoring rule uses Beta (1, 10) for the prior distribution. This means that our prior estimate of proportion of severe pain is 9%, and there is 90% chance that this proportion is 0.5%-26%.

The decision rule for safety stopping is as follows:

Stop if:

# patients with severe pain is at least:	6	10	13	16	19	23
Out of:	15	30	45	60	75	90

4.3 Data Monitoring Committee (DMC)

The DMC will be responsible for safeguarding the interests of trial participants, assessing the safety and efficacy of the clinical trial intervention during the trial, and for monitoring the overall conduct of the trial. The DMC membership will be comprised of clinicians and a Biostatistician who have experience in the management of patients in oncology and in the conduct and monitoring of randomized clinical trials. The DMC membership will be restricted to individuals free of apparent significant conflicts of interest. The DMC will be advisory to the study sponsor.

The sponsor will be responsible for promptly reviewing the DMC recommendations, for deciding whether to continue or terminate the trial, and for determining whether amendments to the protocol or changes in the study conduct are required.

4.4 Statistical Methods

The total counts and percentages of patients will be presented for categorical variables, both overall and by treatment arm. The mean, median, standard deviation, and range, will be presented for continuous variables, both overall and by treatment arm. All statistical outputs will present results by treatment arm and over both

treatment arms except where noted below. All listings will include treatment group. See Appendix 1 for a proposed list of tables, listings and figures.

4.4.1 Demographic and Baseline Characteristics

Demographics

Demographic characteristics will be summarized for all analysis populations as defined in section 3.0 (ITT population, Safety population, and Per-Protocol population). Age in years will be calculated as the date of randomization minus the date of birth divided by 365.25 truncated to the lowest integer.

Baseline Patient Characteristics

The following patient characteristics collected at the Pre-Study visit and/or pre-dose will be summarized by treatment arm and overall: height; weight; BSA; ECOG Performance Status; vital signs; laboratory evaluations, and physical examination. Additional by-treatment arm summaries will reflect prior therapies and duration of prior abiraterone acetate therapy. Baseline characteristics will be summarized for all analysis populations as defined in section 3.0 (ITT population, Safety population, and Per-Protocol population).

Baseline values are defined as the last recorded value prior to the first dose of study drug. For physical examination findings, baseline is defined as the last value for each body system prior to receiving the first dose of study drug.

4.4.2 Patient Disposition

The numbers and percentages of patients who were registered and who are included in the analysis populations will be summarized.

The number and percentages of patients who discontinued from the study and the reason for termination will be presented.

The number of patients that crossover will be summarized, based on crossover for any reason and, in separate analyses, based on crossover after radiographic progression. The null hypothesis of equal crossover rates in each arm will be tested using Fisher's Exact Test.

The date of randomization, date of first dose, date of last dose, date of termination, and reason for termination will be listed for each patient.

Length of time on study will be analyzed using Kaplan-Meier techniques, with time computed from date of randomization to the latest of the last on-study

date, date of last contact or date of death. Length of time on study will be computed for prior to crossover, after crossover and overall.

4.4.3 Efficacy Analysis

All efficacy analyses will be performed using the ITT analysis population unless otherwise noted. Efficacy analyses will be presented by treatment arm and overall and for each randomization stratum; results by randomization stratum will not present p-values. For all treatment comparisons involving odds ratios or hazard ratios, these ratios will be computed using BAT values in the numerator and the control arm values in the denominator. The following sections assume crossover will only occur following radiographic progression as described in the study protocol. If crossover occurs without prior radiographic progression, follow-up will be censored at the date of crossover.

4.4.3.1 Primary Endpoint Analysis

Primary efficacy analyses of PFS will be performed using the ITT population. For the primary endpoint analysis, a stratified log-rank test will be used to compare the experimental arm (BAT) to the control arm (enzalutamide). The test will be stratified by the stratification factor defined in section 2.1.1. The analysis for the primary endpoint will be performed after 156 events have occurred. A one-sided p-value less than 0.0441 from a stratified log-rank test will be sufficient to declare superiority of BAT compared to enzalutamide based on 156 events. Under the assumption of exponentially distributed failure times, this corresponds to a hazard ratio less than or equal to 0.761. Additionally, a two-sided nominal 91.18% confidence interval for the hazard ratio will be computed in accordance with the critical values derived from the α-spending function method described above.

This analysis will also be repeated in the Per-Protocol Population to assess the sensitivity of the results to eligibility and protocol compliance.

Exploratory subgroup analyses of PFS will stratify the above analysis by ECOG (0 vs 1+2), age group (<65 versus ≥ 65), duration of response to prior abiraterone (< 6 months versus ≥ 6 months) and race (White versus Non-White).

4.4.3.2 Secondary Endpoint Analyses

4.4.3.2.1 Radiographic Progression-Free Survival (rPFS)

rPFS will be analyzed using the ITT population. A stratified log-rank test will be used to compare the experimental arm (BAT) to the control arm (enzalutamide). The test will be stratified by the stratification factor defined in section 2.1.1.

4.4.3.2.2 PSA response rate from randomization to crossover

The PSA response rate prior to crossover will be estimated for each arm. A 95% confidence interval will be calculated for each estimate using the Clopper-Pearson method. The null hypothesis of equal PSA response rates in each arm will be tested using Fisher's Exact Test.

4.4.3.2.3 PSA response rate following crossover

The PSA response rate after crossover will be estimated for each crossover arm. A 95% confidence interval will be calculated using the Clopper-Pearson method. The null hypothesis of equal PSA response rates in each arm will be tested using Fisher's Exact Test.

4.4.3.2.4 Objective response rate (ORR)

Objective response rate will be summarized for each arm. A 95% confidence interval will be calculated for each arm using the Clopper-Pearson method. The null hypothesis of equal response rates in each arm will be tested using Chi-Square test. If the assumptions of the Chi-Sqare test are not met, Fisher's Exact Test will be used.

4.4.3.2.5 Time to PSA progression

The probability of observing a PSA progression will be calculated using the cumulative incidence method described by Gooley, et al; the equivalence between

arms will be tested using a log-rank test. Death, radiographic progression and discontinuation from the study will be treated as competing events. Patients without a PSA progression or competing event at the time of analysis will be censored at the last PSA assessment date. Note that this analysis differs from the one described in the protocol.

4.4.3.2.6 Time to Initiation of Docetaxel

Time to initiation of Docetaxel will be analyzed using Kaplan-Meier techniques; the treatment arms will be compared using the stratified log-rank test.

4.4.3.2.7 Quality of Life and Metabolic Endpoints

Quality of life (QoL) will be assessed using the RAND-SF36 Quality of Life Survey, Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) Version 4, and International Positive and Negative Affect Schedule Short-Form (I-Panas-SF). Sexual capacity, functional capacity and pain will be assessed using the International Index of Erectile Function (IIEF), Global Assessment of Change and the Brief Pain Inventory (BPI), respectively. For each instrument, summary statistics of the scores and change scores will be reported at baseline and post-randomization at 1, 3, 6 and 12 months. The null hypothesis of equal change between arms will be tested at each time point using a Wilcoxon signed rank test with a multiple comparison adjusted significance level of 0.0125. In addition, repeated measures models will be used to access changes in QoL summary measures over time.

The change from baseline in metabolic measures (e.g. bone density, lipid profile, metabolic panel, steroid panel, inflammation/cytokines, BMI, lean mass) will be evaluated using the same analysis methods as for QoL.

Metabolic parameters will be compared between the two treatment arms using regression models.

4.4.3.3 Exploratory Endpoint Analyses

4.4.3.3.1 AR-V7 Correlative Analyses

The association between BAT-induced decrease in AR-V7 expression and restoration of sensitivity to ADT and enzalutamide will be assessed.

The association of AR-V7 at baseline with clinical outcomes of rPFS, objective response rate (ORR) and PSA response will be examined by treatment arm. For rPFS, Kaplan-Meier curves will be generated for each treatment arm stratified by baseline AR-V7 status. Cox proportional hazards models will be used to estimate the hazard ratio with 95% Confidence Intervals for each arm to quantify the effect of AR-V7 presence/absence on the risk of progression. The binary outcomes of ORR and PSA response will be compared between baseline AR-V7(+) and AR-V7(–) groups using Chi-square tests (or Fisher's Exact test if the assumptions of the Chi-square test are not met). The interaction of treatment and AR-V7 status may also be explored using Cox Proportional Hazards modelling for rPFS.

4.4.4 Safety Analysis

Safety analyses will be performed using the Safety Analysis Population unless otherwise specified. Safety analyses will be presented separately for the pre- and post-crossover treatment periods. Safety analyses pre-crossover will include all events from the date of first treatment to the earliest of removal from the study, radiographic progression or crossover. Safety analyses post-crossover will include all events from the date of first crossover treatment until removal from the study.

4.4.4.1 Adverse Events

All Adverse Event summaries will be presented by treatment period (randomization versus crossover). An overall summary of adverse events will be presented by NCI CTCAE grade. Tabulations of events by maximum grade, system organ class and preferred term will also be presented. The number and proportion of patients experiencing each event will be presented according to the maximum grade experienced. Separate summaries of drug-related adverse events will be generated using drug-related attribution corresponding to the actual drug received in the treatment period reflected in the analysis (randomization versus crossover). For example, for patients on BAT in a given treatment period, the BAT drug-relationship response will be used to ascribe relationship in AE summaries for that treatment period.

4.4.4.2 Serious Adverse Events

Serious adverse events will be tabulated by system organ class and preferred term. The number and proportion of patients experiencing each event will be presented. A separate summary of serious drug-related adverse events will be generated.

Serious adverse event data will also be reflected in a listing.

4.4.4.3 Clinical Laboratory Toxicities

Selected clinical laboratory assessments will be graded according to CTC version 4. The following lab values will be summarized according to maximum grade when available: WBC, Neutrophils, Lymphocytes, Platelets, Hemoglobin, Sodium, Serum Potassium, Calcium, Creatinine, and Total Bilirubin. An overall summary will be presented for the treatment periods before and after crossover. A shift table will be generated indicating change from baseline to maximum grade observed during the pre-crossover treatment period.

If local lab normal ranges cannot be determined from data entered in the study database, standard lab ranges will be used. When there are potential conflicts between local lab normal ranges and ranges used in CTC grading, CTC normal ranges will be used. See Appendix 2 for the lab toxicity grade definitions.

4.4.4.4 Toxicity Monitoring for Pain

Continuous monitoring of adverse events of pain on the BAT arm will be carried out by the statistician named in the protocol.

4.5 Other Issues and Further Details

4.5.1 Imputation Rules for Primary Endpoint (PFS)

If a radiographic progression date or death date must be imputed to calculate the primary endpoint, the following algorithm will be used:

- If the date is missing month and day and the subject was last known to be progression free and alive in the same year, the last radiographic assessment date will be used as the progression date.
- If the subject's last assessment date was in the previous year, the month and day will be imputed as the first day of the year (JAN01).
- If the year and month are known, the date will be imputed as the maximum of the last assessment date and the first day of the known year and month.

4.5.2 General Imputation Rules for Partial Dates

When imputation of partial dates is required for calculation of durations, a method like the one described in Section 4.5.1 of this document will be utilized.

4.5.3 Conversions from Days to Years, Months or Weeks

Years = # of days / 365.25
Months = # of days / 30.4375 (i.e. 365.25/12)
Weeks = # of days / 7

Summaries based on the above computations will be rounded to tenths.

4.5.4 Computation of Duration

Duration for time variables based on two dates, e.g., Start Date and End Date, will be calculated as (End Date – Start Date + 1) (in days) unless otherwise specified.

4.5.5 Missing normal ranges for laboratory parameters

When either the lower limit of normal, the upper limit of normal or both are missing or are not machine readable, a standardized reference range will be used.

4.5.6 Non-Numeric Laboratory Results and Calculation of Normal Ranges

Laboratory values including symbols ("<" or ">", for example) will not be used in summary analyses. These values will be reflected in listings of the data.

4.5.7 Procedures for Reporting Deviations to Original Statistical Analysis Plan

Any deviations from this plan will be described in an appendix to the final study report.

4.6 Statistical Software used in data analysis

All statistical analyses will be performed using SAS (version 9.4 or greater) or R (version 3.1 or greater).

4.7 Additional analyses not referenced in this document

All additional analyses not referenced in this document will be considered exploratory.

5.0 References

- Pocock, S., and Simon, R. (1975). Sequential treatment assignments with balancing for prognostic factors in the controlled clinical trial. *Biometrics 31: 103-115.*
- Lan, K. K. G., and DeMets, D. L. (1983) Discrete sequential boundaries for clinical trials. *Biometrika* 70 (3): 659-663
- Jennison C, Turnbull BW (1993) Sequential equivalence testing and repeated confidence intervals, with applications to normal and binary responses. *Biometrics*.49:31-43.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, *50:163-170*.
- Kaplan, E.L., and Meier P. (1958). Nonparametric estimation from incomplete observations. *J Am Stat Assoc, 53:457-481*.
- Agresti, Alan. (1992) A survey of Exact Inference for Contingency Tables, *Statistical Science* 7(1): 131-153.
- Clopper, CJ and Pearson, ES, (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26:404-413.
- Gooley TA, Leisenring W, Crowley J, et al. (1999) Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med* 18:695-706.
- Hosmer, D.W., and Lemeshow, S. (2000). *Applied logistic regression, second edition.* Hoboken: John Wiley & Sons, Inc.
- Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, B34:187-220*.

6.0 Appendices

Appendix 1. Proposed list of outputs

Tables

Subject Demographics - ITT Analysis Population

Subject Demographics - Safety Analysis Population

Subject Demographics - PP Analysis Population

Subject Demographics – Correlative Studies Population

Summary of Prior Therapy – ITT Analysis Population

Baseline Physical Examination – ITT Analysis Population

Baseline Vital Signs – ITT Analysis Population

Baseline Laboratory Values – ITT Analysis Population

Patient Disposition – ITT Analysis Population

Summary of Inclusion Criteria – ITT Analysis Population

Summary of Exclusion Criteria – ITT Analysis Population

Length of Follow-up – ITT Analysis Population

Summary of rPFS – ITT Analysis Population

Summary of rPFS – PP Analysis Population

Summary of rPFS by Subgroup – ITT Analysis Population

Summary of PSA Response – ITT Analysis Population, Randomization to Cross-over

Summary of PSA Response – ITT Analysis Population, post-Crossover

Summary of Secondary Endpoint, Time to PSA Progression – ITT Analysis Population

Summary of Quality of Life Measures: RAND-SF36 – ITT Analysis Population

Summary of Quality of Life Measures: FACIT-F– ITT Analysis Population

Summary of Quality of Life Measures: I-PANAS-SF – ITT Analysis Population

Summary of Quality of Life Measures: IIEF – ITT Analysis Population

Summary of Quality of Life Measures BPI – ITT Analysis Population

Overall Summary of AEs - Safety Analysis Population, Randomization

Overall Summary of AEs - Safety Analysis Population, Crossover

Adverse Events by NCI CTC System Organ Class and Preferred Term – Safety Analysis Population, Randomization

Adverse Events by NCI CTC System Organ Class and Preferred Term – Safety Analysis Population, Crossover

Serious Adverse Events by NCI CTC System Organ Class and Preferred Term – Safety Analysis Population, Randomization

Serious Adverse Events by NCI CTC System Organ Class and Preferred Term – Safety Analysis Population, Crossover

Adverse Events by NCI CTC System Organ Class, Preferred Term and Maximum Grade– Safety Analysis Population, Randomization

Adverse Events by NCI CTC System Organ Class, Preferred Term and Maximum Grade– Safety Analysis Population, Crossover

Drug-Related Adverse Events by NCI CTC System Organ Class, Preferred Term and Maximum Grade– Safety Analysis Population, Randomization

Drug-Related Adverse Events by NCI CTC System Organ Class, Preferred Term and Maximum Grade– Safety Analysis Population, Crossover

Clinical Laboratory Events by NCI CTC Preferred Term and Maximum Grade – Safety Analysis Population, Randomization

Clinical Laboratory Events by NCI CTC Preferred Term and Maximum Grade – Safety Analysis Population, Crossover

AR-V7 Correlative Analysis: rPFS

AR-V7 Correlative Analysis: ORR

AR-V7 Correlative Analysis: PSA Response

Figures

rPFS by Treatment Arm- ITT Analysis Population

rPFS by Treatment Arm - PP Analysis Population

Time to PSA Progression by Treatment Arm- ITT Analysis Population

rPFS Forest Plot for Pre-Specified Subgroups

rPFS by Treatment Arm and AR-V7 -

Listings

Subject Demographics and Randomization

Patient Disposition

Inclusion and Exclusion Criteria

Adverse Events

Death Information

Radiographic Progression

Metabolic Measurements

PSA Measurements

Tumor Response

Treatment after Radiographic Progression

Appendix 2. Toxicity Grading Definitions for Clinical Lab Results

Form	Lab variable	CTCAE v 4.0 Term	Grade 1	Grade 2	Grade 3	Grade 4	Reporting units	Reference Range Source	Young & Huth Page #	LLN	ULN
Hematology	Platelets	Platelet count decreased	<LLN - 75,000/mm3; <LLN - 75.0 x 10e9 /L	<75,000 - 50,000/mm3; <75.0 - 50.0 x 10e9 /L	<50,000 - 25,000/mm3; <50.0 - 25.0 x 10e9 /L	<25,000/mm3; <25.0 x 10e9 /L	10^3/uL	*	209	150	450
	WBC	White blood cell decreased	<LLN - 3000/mm3; <LLN - 3.0 x 10e9 /L	<3000 - 2000/mm3; <3.0 - 2.0 x 10e9 /L	<2000 - 1000/mm3; <2.0 - 1.0 x 10e9 /L	<1000/mm3; <1.0 x 10e9 /L	10^3/uL	*	175	3.2	9.8
		Leukocytosis	-	-	>100,000/mm3		10^3/uL	*	175	3.2	9.8
	Neutrophil %	Neutrophil count decreased	<LLN - 1500/mm3; <LLN - 1.5 x 10e9 /L	<1500 - 1000/mm3; <1.5 - 1.0 x 10e9 /L	<1000 - 500/mm3; <1.0 - 0.5 x 10e9 /L	<500/mm3; <0.5 x 10e9 /L	10^3/uL	*	193	3	5.8
	Lymphocyte %	Lymphocyte count decreased	<LLN - 800/mm3; <LLN - 0.8 x 10e9/L	<800 - 500/mm3; <0.8 - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200/mm3; <0.2 x 10e9 /L	10^3/uL	*	178	1.5	3
		Lymphocyte count increased	-	>4000/mm3 - 20,000/mm3	>20,000/mm3	-	10^3/uL	*	178	1.5	3

	Hemoglobin	Anemia	(Hgb) <LLN - 10.0 g/dL; <LLN - 6.2 mmol/L; <LLN - 100 g/L	Hgb <10.0 - 8.0 g/dL; <6.2 - 4.9 mmol/L; <100 - 80g/L	Hgb <8.0 g/dL; <4.9 mmol/L; <80 g/L		g/dL	*	151	13.5	17.5
		Hemoglobin increased	Increase in >0 - 2 gm/dL above ULN or above baseline if baseline is above ULN	Increase in >2 - 4 gm/dL above ULN or above baseline if baseline is above ULN	Increase in >4 gm/dL above ULN or above baseline if baseline is above ULN	-	g/dL	*	151	13.5	17.5
Chemistry	Creatinine	Creatinine increased	>ULN - 1.5 x ULN; >1 - 1.5 x baseline;	>1.5 - 3.0 x ULN; >1.5 - 3.0 x baseline	>3.0 - 6.0 x ULN; >3.0 baseline	>6.0 x ULN	mg/dL	*	114	0.6	1.2
	Calcium	Hypocalcemia	Corrected serum calcium of <LLN - 8.0 mg/dL; <LLN - 2.0 mmol/L	Corrected serum calcium of <8.0 - 7.0 mg/dL; <2.0 - 1.75 mmol/L	Corrected serum calcium of <7.0 - 6.0 mg/dL; <1.75 - 1.5 mmol/L	Corrected serum calcium of <6.0 mg/dL; <1.5 mmol/L	mg/dL	*	92	8.4	10.2
		Hypercalcemia	Corrected serum calcium of >ULN - 11.5 mg/dL; >ULN - 2.9 mmol/L	Corrected serum calcium of >11.5 - 12.5 mg/dL; >2.9 - 3.1 mmol/L	Corrected serum calcium of >12.5 - 13.5 mg/dL;>3.1 - 3.4 mmol/L	Corrected serum calcium of >13.5 mg/dL; >3.4 mmol/L	mg/dL	*	92	8.4	10.2
	Total bilirubin	Blood bilirubin increased	>ULN - 1.5 x ULN	>1.5 - 3.0 x ULN	>3.0 - 10.0 x ULN	>10.0 x ULN	mg/dL	*	87	0.2	1.3
	Sodium	Hyponatremia	<LLN - 130 mmol/L	-	<130 - 120 mmol/L	<120 mmol/L	mmol/L	*	228	137	145

		Hypernatremia	>ULN - 150 mmol/L	>150 - 155 mmol/L	>155 - 160 mmol/L	>160 mmol/L	mmol/L	*	228	137	145
	Potassium	Hypokalemia		<LLN - 3.0 mmol/L	<3.0 - 2.5 mmol/L	<2.5 mmol/L	mmol/L	**		3.5	5
		Hyperkalemia	>ULN - 5.5 mmol/L	>5.5 - 6.0 mmol/L	>6.0 - 7.0 mmol/L	>7.0 mmol/L	mmol/L	**		3.5	5

* E, Young D and Huth. SI Units for Clinical Measurement. Philadelphia, Pennsylvania : Americal College of Physicians, 1998. ISBN 0-943126-51-7.

** Merck & Co, Inc. Normal Laboratory Values. The Merck Manual Professional Edition. [Online] April 2013. http://www.merckmanuals.com/professional/appendixes/normal_laboratory_values/normal_laboratory_values.html.

7.0 Approvals

Printed Name	Signature	Date
Antje Hoering		
Lynn Shemanski		
Sam Denmeade		
Hao Wang		
Grace Powell (CRAB QA)		