Statistical Analysis Plan

Randomized, Open-Label Study of Abiraterone Acetate (JNJ-212082) plus Prednisone with or without Exemestane in Postmenopausal Women with ER+ Metastatic Breast Cancer Progressing after Letrozole or Anastrozole Therapy

Protocol 212082BCA2001; Phase 2

JNJ-212082 (Abiraterone Acetate)

Issue/Report Date: 30-July-2013
Document No.: EDMS-ERI-68198934: 0.1

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

TABLE OF CONTENTS

AMENDMENT HISTORY

NA

ABBREVIATIONS

AA	abiraterone acetate
AAP	abiraterone acetate plus prednisone
AAPE	abiraterone acetate plus prednisone combined with exemestane
AE	adverse event
ALT	alanine aminotransferase (SGPT)
AR	androgen receptor
AST	aspartate aminotransferase (SGOT)
ATC	Anatomical Therapeutic Chemical
BPI-SF	Brief Pain Inventory – Short Form
CBC	complete blood count
CI	confidence interval
CT	computed tomography
CTC	circulating tumor cells
CR	complete response
CRF	case report form
DRC	Data Review Committee
E	exemestane
ECG	electrocardiogram
ECHO	echocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EMEA	European Medicines Agency
EORTC-C30	European Organization for Research and Treatment of Cancer core quality-of-life questionnaire
EQ-5D-5L	Euro-QoL quality-of-life questionnaire
EQ-VAS	EQ visual analogue scale
ER+	estrogen receptor positive
FDA	Food and Drug Administration
HDL	High density lipoprotein
ICH	International Conference on Harmonisation
INR	international normalized ratio
IPCW	inverse probability of censoring weighted
ITT	intent-to-treat
IVRS/IWRS	interactive voice/web response system
LFT	liver function test
LDL	Low density lipoprotein
LVEF	Left Ventricular Ejection Fraction
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
MUGA	multiple-gated acquisition scan
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events, version 4
OS	Overall survival
PFS	progression-free survival
PK	Pharmacokinetics
PR	Partial response
PR+	progesterone receptor positive
PRO	patient-reported outcome
PT	prothrombin time
PTT	partial thromboplastin time
Qol	Quality of life

RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SOC	System Organ Class
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of the analysis populations, derived variables and statistical methods to be used in the analyses of efficacy and safety for the study Protocol 212082BCA2001, titled "Randomized, Open-Label Study of Abiraterone Acetate (JNJ 212082) plus Prednisone/ Prednisolone with or without Exemestane in Postmenopausal Women with ER+ Metastatic Breast Cancer Progressing after Letrozole or Anastrozole Therapy".

1.1. Trial Objectives

1.1.1 Primary Objective

The primary objective is to assess the safety and efficacy of abiraterone acetate plus prednisone and abiraterone acetate plus prednisone combined with exemestane, each compared with exemestane alone, in postmenopausal women with estrogen receptor positive (ER+) metastatic breast cancer progressing after letrozole or anastrozole therapy.

1.1.2 Secondary Objectives

Secondary objectives are to assess abiraterone acetate plus prednisone (AAP) and abiraterone acetate plus prednisone combined with exemestane (AAPE), each compared with exemestane (E) alone, in postmenopausal women with ER+ metastatic breast cancer progressing after letrozole or anastrozole therapy, with respect to the following:

- Overall survival (OS);
- Overall response rate;
- Patient-reported outcome (PRO) from the following three instruments: European Organization for Research and Treatment of Cancer core quality-of-life questionnaire (EORTC-C30), Euro-QoL quality-of-life questionnaire (EQ-5D-5L), and Brief Pain Inventory-Short Form (BPI-SF) pain intensity scale;
- Endocrine markers estradiol, testosterone, estrone, and other biomarkers;
- Pharmacokinetics (PK) characterization of abiraterone and exemestane.

1.2. Study Design

Abiraterone, the active metabolite of abiraterone acetate (AA) (JNJ 212082), is a novel selective cytochrome P450 CYP17 enzyme inhibitor being developed for treatment of patients with advanced metastatic castration-resistant prostate cancer and breast cancer.

This is a randomized, open-label, parallel-group, multicenter Phase 2 breast cancer study. From approximately 65 sites, approximately 300 subjects will be randomized in a 1:1:1 ratio to one of the three treatment groups: AAP, AAPE, or E. Subjects are stratified by two factors: number of prior therapies in the metastatic setting (0 or 1 versus 2) and setting of prior letrozole or anastrozole treatment (adjuvant versus metastatic).

This study is divided into three phases: Screening, Treatment, and Follow-Up. During Screening, potential study subjects are assessed for study eligibility after providing signed informed consent.

The Treatment phase comprises a series of 28-day cycles with continuous study treatment. Determination of progression free survival (PFS) will use radiographic progression defined by Response Evaluation Criteria in Solid Tumors (RECIST) on measurable lesions captured by computed tomography (CT) or magnetic resonance imaging (MRI) at baseline and, using the same modality, repeated every 2 cycles initially (Cycle 3, 5, and 7) and every 3 cycles thereafter (e.g., Cycle 10 and 13) until, and including, the End-of-Treatment visit. Treatment will continue until the earliest of the following events: disease progression, unacceptable toxicity, or death. At disease progression, subjects randomized to exemestane alone may be switched to abiraterone acetate plus prednisone at the discretion of the investigator; if not switched, these subjects must be discontinued from study drug.

Subjects will enter the Follow-Up phase regardless of reason for study drug discontinuation and will be monitored every 3 months until death, loss to follow-up, consent withdrawal, or abiraterone acetate development in this indication is discontinued. Survival status and non study anti-breast-cancer therapies will be collected by telephone or chart review. Patient-reported outcomes also will be completed per protocol. The total study duration is for approximately 36 months (3 years).

1.3. Statistical Hypotheses for Trial Objectives

The primary efficacy endpoint is PFS. For the experimental treatment groups AAP and AAPE, each is compared with the control group (E) in postmenopausal women with ER+ metastatic breast cancer.

For each of the pair-wise comparison, the statistical hypotheses are as follows:

H_0: The PFS distributions of experimental treatment group, ST(t), and the control group, SC(t), are equal at all time points t:

$$ST(t) = SC(t), \text{ for all } t > 0$$

versus

H_1: The PFS distributions are not equal at least one time point t:

$$ST(t) \neq SC(t), \text{ for some } t > 0$$

These hypotheses will be tested using log-rank test and assessed within the context of a group sequential testing design as described in Sections 3. Each pair-wise testing will be conducted independently, no adjustment to the Type I error will be made.

1.4. Sample Size Justification

Recent publication in a similar patient population treated with exemestane (Chia 2008) suggests that the estimated median PFS for the control exemestane alone group (E) is approximately 4 months. For sample size calculation, it is assumed that PFS follows an exponential distribution with a constant hazard ratio.

Assuming an underlying hazard ratio of any pair-wise comparison (AAP or AAPE, each compared with E) is 0.65 (median PFS 6.2 and 4.0 months, respectively), the study has 80% power at a significance level of 0.10 (two-sided) to demonstrate a treatment difference with approximately 150 PFS events. Assuming an enrollment rate of 20 subjects/month for 15 months, sample size of approximately 300 subjects (100 subjects/group) is planned for this study to achieve 220 PFS events over a study duration of 19 months.

1.5. Randomization and Blinding

This is an open-label study. Eligible subjects will be randomized in a 1:1:1 ratio to one of three treatment groups: AAP, AAPE, or E. Randomization is stratified based on two factors: number of prior therapies in the metastatic setting (0 or 1 versus 2), and setting of prior letrozole or anastrozole treatment (adjuvant versus metastatic). Randomization will be implemented by an Interactive Voice/Web Response System (IVRS/IWRS).

2. GENERAL ANALYSIS DEFINITIONS

2.1. Population Analyzed

2.1.1. Intention-to-Treat (ITT) Population

The intent-to-treat (ITT) population is defined as all subjects randomized into the study and who will be classified according to assigned treatment group, regardless of the actual treatment received. This population will be used for all efficacy analyses including PFS and overall survival (OS), and all analyses of disposition, demographic, and baseline disease characteristics.

Analysis of select secondary efficacy endpoints will be based on the ITT principle with following modifications:

- Overall response rate analysis will include all ITT subjects with measurable disease at baseline. Subjects with missing post-randomization measures will be considered nonresponders.
- Analysis of endocrine markers will include all ITT subjects who have a valid baseline and at least 1 post-randomization value.
- PK analysis will include all ITT subjects with sufficient and interpretable PK assessments to estimate the noncompartmental PK parameters.
- Analysis population for PRO endpoints will be detailed in Section 7.

2.1.2. Safety Population

Safety population is defined as all randomized subjects who receive at least one dose of study drug. This population will be used for all safety analyses and all analyses of treatment compliance and exposure. All subjects will be analyzed according to the treatment which they actually received.

2.2. Baseline Definitions or Conventions

Unless specified otherwise, the baseline value is defined as the last available value collected on or prior to the first dose of study drug.

2.3. Study Day and Visit Windows

The protocol allows a 2 day window between randomization and first dose. For analyses on efficacy endpoints, e.g. time-to-event endpoints and PRO endpoints, the randomization day is considered as the Day 1. For analyses on safety, drug exposure and compliance, the first day of dosing is considered as Day 1. Subjects' time on study will be determined in study days. Study day is defined as follows:

Study Day = the current date – Day 1 + 1, if the current date ≥ date of Day 1;

Study Day = the current date – Day 1, if current date < date of Day 1.

The study date and corresponding study visit/cycle will be captured on each case report form (CRF). Visit windows will be created around the study day of each scheduled visit. It will be used to aggregate data, which are to be summarized by visit, e.g. clinical laboratory, vital signs, and PRO assessments. The visit windows and the targeted study day are indicated in Table 1. If more than one assessment falls within the same visit window, then the assessment that occurred first will be used in the analysis. If an assessment is not scheduled for every visit, windows will be combined so that the interval between targeted study days is split evenly and consistently between visits. Any cycles beyond Cycle 12 will have the windows similarly constructed.

Table 1. Visit Windows

Scheduled Study Day	Visit Window	Scheduled Study Day	Visit Window
Cycle 1 Day 1*	1, 7	Cycle 6 (Day 140)	127, 154
Cycle 1 Day 15	8, 22	Cycle 7 (Day 168)	155, 182
Cycle 2 Day 1 (Day 28)	23, 35	Cycle 8 (Day 196)	183, 210
Cycle 2 Day 15 (Day 42)	36, 49	Cycle 9 (Day 224)	211, 238
Cycle 3 Day 1 (Day 56)	50, 63	Cycle 10 (Day 252)	239, 266
Cycle 3 Day 15 (Day 71)	64, 78	Cycle 11 (Day 280)	267, 294
Cycle 4 (Day 84)	79, 98	Cycle 12 (Day 308)	295, 322
Cycle 5 (Day 112)	99, 126		

* For safety, drug exposure and compliance, Study Day 1 begins on the day of first dose. For efficacy endpoints, Study Day 1 begins on the day of randomization.
Each cycle consists of 28 days.
Baseline value is defined as the last available value collected on or prior to the first dose of study drug.

2.4. Missing and Partial Dates

For missing and incomplete dates during the study, the following imputation rule serves as a general guideline for imputation in the assessment of an event:

1. If the date is completely missing or the year is missing, no imputation will be made.

2. Missing Start Day, and only day is missing:

   - If it has the same month or year as the start of treatment, then Day = day of start of treatment.
   - Otherwise, Day = 01, e.g. November 1990 is imputed as 01NOV1990.

3. Missing Start Month and Day:
   - If it has the same year as the start of treatment, then Month and Day = month and day of start of treatment.
   - Otherwise, Month = 01 and Day=01, e.g. 1990 is imputed as 01JAN1990.

4. Missing Stop Day, and only day is missing:
   - If it has the same month or year as the end of treatment, then Day = day of end of treatment.
   - Otherwise, Day = 30 or 31 (or 28 or 29 for February), depending on the month and the year, e.g. November 1990 is imputed as 30NOV1990.

5. Missing Stop Month and Day:
   - If it has the same year as the end of treatment, then Month and Day = month and day of end of treatment.
   - Otherwise, Month = 12 and Day= 31, e.g. 1990 is imputed as 31DEC1990

3. INTERIM ANALYSIS AND DATA REVIEW COMMITTEE

3.1. Interim Analysis

One interim analysis is planned. Key efficacy analyses as indicated in Table 4 below will be performed together with primary safety analyses. It is scheduled when 50% of the PFS events have occurred, or approximately 75 PFS events have occurred in any of the two pair-wise comparisons: AAP versus E, and AAPE versus E. This is anticipated to be approximately 110 PFS events across all the three treatment groups, and projected to be approximately 12 months after first subject has been randomized.

3.1.1 Stopping Boundary

The purpose of the interim analysis is to investigate if AAP or AAPE demonstrate efficacy improvement over E based on PFS. A significant efficacy finding in PFS at interim analysis will allow planning of a Phase 3 confirmatory study. The study may drop one or both experimental treatment groups for futility or for superiority.

Gamma-family spending functions (Huang 1990) were used to control the Type I error of 0.05 (one-sided) and Type II error of 0.20 for each comparison. Figure 1 shows the shape of the Type I error (alpha) and Type II error (beta) spending functions for the PFS endpoint.

- Alpha spending: a Gamma-family spending function with parameter rho=1 is used to be similar to Pocock spending function. Figure 1 shows that it allocates cumulative alpha 0.031 (62% of total spending) at the interim (50% of information), and cumulative alpha 0.05 (one-sided) at the final analysis.

- Beta spending: a Gamma-family spending function with parameter rho=-2 is used. Figure 1 shows it allocates cumulative beta 0.054 (27% of total spending) at the interim (50% of information), and cumulative beta 0.20 at the final analysis.

The stopping boundaries for PFS-based futility and efficacy were calculated using East® software v5.2. Table 2 summarizes the efficacy and futility monitoring plan for each pair-wise comparison. The corresponding estimated hazard ratio for cutoff values are approximates.

Figure 1. Cumulative Error Spending Function for Alpha and Beta

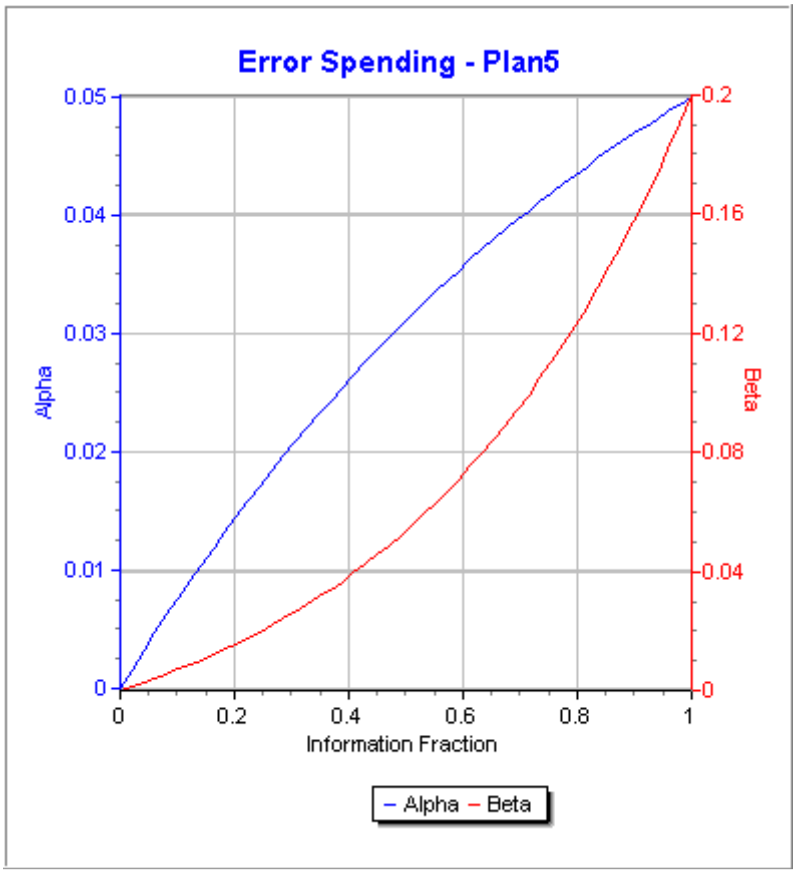

Table 2. Stopping Boundaries and Corresponding Minimal Observed Difference for Each Pair-wise Comparison

Analysis	PFS Events (information)	Cumulative Error		Stopping Boundary				Observed Hazard Ratio	
		Alpha	Beta	Z score		p-value (one-sided)			
		Efficacy	Futility	Efficacy	Futility	Efficacy	Futility	Efficacy	Futility
Interim	75 (50%)	0.031	0.054	1.865	0.286	<0.031	≥ 0.388	<0.650	≥ 0.936
Final	150 (100%)	0.05	0.20	1.887	1.887	<0.030	≥ 0.030	<0.735	≥ 0.735

Type I error of 0.05 (one-sided), efficacy boundary is based on Gamma family spending function with rho=1, similar to Pocock.
Power of 80%, futility boundary is based on Gamma family spending function with rho=-2
Observed hazard ratio are approximates.

3.1.2 Conditional Power

Conditional power is the probability that the study will demonstrate statistical significance at the end of the study (i.e. final analysis to claim superiority on PFS), conditioning on the data observed in the study thus far, and an assumption about the trend of the data to be observed in the remainder of the study. Two assumptions about the trend of the data were presented below:

The futility boundary corresponds to a conditional power of approximately 39% if the original hazard ratio assumption is true, while only 4% conditional power will be achieved if the observed hazard ratio at interim is true for the remainder of the study.

The efficacy boundary corresponds to a conditional power of approximately 90% if the original hazard ratio assumption is true, and 92% conditional power will be achieved if the observed hazard ratio at interim is true for the remainder of the study.

The conditional power of stopping boundaries was computed using method of Lan (2009).

3.2. Data Review Committee

A Data Review Committee (DRC) will be formed to monitor data on a regular basis to ensure the safety of the subjects in this study, assess the evidence of benefit or adverse effects of abiraterone acetate, and to monitor the overall conduct of the study. The DRC is comprised of experts in breast cancer, medical safety, and biostatistical aspects of clinical studies. Details regarding DRC responsibilities and activities will be provided in a separate DRC Charter.

In addition to the planned interim analysis, the DRC will meet periodically to review the cumulative safety data. The first safety review meeting will be held approximately after the 10th subject in each treatment group (~30 subjects total) has received study drug for 28 days (Cycle 1). The DRC will review unblinded reports for safety, to identify any potential added toxicity when abiraterone acetate is combined with exemestane. Subject accrual will continue only if no safety concerns arise. The second safety review meeting will be held when baseline and at least one post-treatment ECG are available for approximately the first 25 subjects in each treatment group (~75 subjects total). The DRC will review unblinded reports for safety and determine if additional ECGs are needed for these and subsequent subjects.

3.3. Protocol Amendment and Final Analysis

The DRC reviewed the unblinded efficacy, safety, and biomarker outcomes of the interim analysis after the clinical cutoff (110 [50%] of progression or death events) occurred. The clinical cutoff date was December 14, 2012. On 08 March 2013, the DRC recommended discontinuing the enrollment to the AAP group because the interim analysis criterion for futility was met. The study protocol was amended accordingly and the following changes were implemented with the new protocol (Amendment INT-5):

- The randomization ratio will now be 1:1 with randomization to the AAPE group or the E only group. There will no longer be randomization to the AAP group.

- No crossover is permitted. Subjects with disease progression must discontinue study drug.
- Subjects who were randomized to the AAP group before implementation of Amendment INT-5 can decide whether or not to continue study drug.

The final analysis will be conducted when 150 progression or death events occurred in the AAPE group and the E alone group. Although the enrollment to AAP group has been discontinued per DRC recommendation, the data display will present all the 3 treatment groups. However, the statistical inference will be focused on the comparison between the AAPE group and the E only group.

4. SUBJECT AND TREATMENT INFORMATION

All statistical analyses will be performed using SAS®. Analyses of disposition, demographic, baseline disease characteristics, and prior and concomitant therapy will be conducted on the ITT population. Analyses on treatment compliance and extent of exposure will be conducted on the Safety Population. No statistical testing is planned.

Unless otherwise specified, all continuous endpoints will be summarized using descriptive statistics, which will include the number of subjects with a valid measurement (n), mean, standard deviation (SD), median, minimum, and maximum. All categorical endpoints will be summarized using frequencies and percentages. Percentages will be calculated by dividing the number of subjects with the characteristic of interest by the number of subjects with non-missing value in the analysis population, unless otherwise specified.

4.1. Subject Disposition

Distribution of subjects by treatment group for each of the analysis population will be provided. The number of subjects enrolled, dosed, and discontinued will be summarized. Treatment discontinuation will be summarized according to reasons of discontinuation as documented in the CRF.

In addition, study enrollment by country and study site will be summarized for each randomized treatment group and for the 3 groups combined.

4.2. Demographics and Baseline Characteristics

Subject demographics and baseline disease characteristics as described in Table 3 will be summarized using descriptive statistics.

Table 3. Demographics and Baseline Characteristics

Variable	Description
Demographics	
Continuous Variables: • Age (yr)	(n, mean, SD, median, minimum, and maximum)
Categorical Variables: • Age group (<65, 65 to 69, 70 to 74, ≥ 75) • Sex (Female) • Race • Ethnicity	(n, %)
Baseline Vital Signs	
Continuous Variables: • Weight (kg) • Height (cm) • Systolic and Diastolic Blood Pressure (mmHg) • Heart Rate (beats/min) • Respiratory Rate (breaths/min) • Body Temperature (^{o}C)	(n, mean, SD, median, minimum, and maximum)
Baseline Disease Characteristics	
Continuous Variables: • Time from initial diagnosis to first dose • Time from initial breast cancer diagnosis to diagnosis of metastatic disease • Time from diagnosis of metastatic disease to first dose • Time from end of letrozole/anastrazole to first dose • Time from most recent progression to first dose • Time from surgical removal of primary tumor to diagnosis of metastatic disease • Time from end of last endocrine therapy to first dose	(n, mean, SD, median, minimum, and maximum)
Categorical Variables • Subject status at initial diagnosis (post-menopausal, pre-menopausal) • Sites of metastatic disease (visceral, visceral only , lymph nodes, bone, bone only, skin, muscle, bone and lymph nodes without visceral, other) • Number of sites of metastatic disease (1,2,3,>=4) • Determination of recent progression (radiological, clinical) • ECOG PS Score (0, 1) • PR status (positive, negative)	(n, %)

• ER status (positive, negative) Baseline LDH (high, normal or low)	
Prior Therapies • Type of prior cancer therapy: surgery, radiotherapy, hormonal, chemotherapy, other • Subjects with adjuvant prior chemotherapy (Yes/No) • Best response of prior therapy (complete response, partial response, stable disease) • Subjects with measurable disease at baseline (Yes/No)	(n, %)
Stratification Factors • Prior therapies in the metastatic setting (0 or 1 versus 2) • Prior letrozole or anastrozole treatment (adjuvant versus metastatic)	(n, %)
ER=estrogen receptor; PR=progesterone receptor; ECOG PS=Eastern Cooperative Oncology Group Performance Status; LDH=lactate dehydrogenase; ,	

4.3. Prior and Concomitant Medications

For summarization purposes, medications will be coded to a generic term based on the World Health Organization (WHO) dictionary. Medications administered prior to the first dose of study drug will be considered prior medications. Concomitant therapies include those taken on or after first dose date through 30 days after the last dose.

Incidence of prior and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) class and drug generic term. Summary of subjects' prior anti-cancer therapy including radiotherapy, surgery, hormone therapy, and chemotherapy will each be presented.

4.4. Protocol Deviations

In general, subjects who do not meet the inclusion/exclusion criteria are not allowed to be randomized into the study.

Protocol deviations will be captured in the CRF and reviewed by a medical monitor. Major protocol deviations will be summarized by the following categories:

- Subjects entering study although entrance criteria not met;
- Subjects developing withdrawal criteria during the study, but who were not withdrawn;
- Subjects receiving the wrong treatment or incorrect dose;
- Subjects receiving an excluded concomitant treatment;
- Abiraterone acetate not taken in the fasted state;
- Other major non-compliance.

Individual subjects with protocol deviations will be listed.

4.5. Treatment Compliance

Drug compliance is estimated by the percent of number of tablets actually taken during the course of study while the subject is on active treatment. Specifically, for subject *i*

$$Compliance_i = \left(\frac{\text{Actual number of tablets taken}}{\text{Total number of tablets supposed to be taken}} \right) \times 100$$

For each drug, the number and percentage of subjects who completed the specified range of compliance rate, starting at ≤ 75% compliance to a maximum of 100% in increments of 5% will be summarized by treatment as appropriate. The period after discontinuation (of study medication or from the study itself) will not be considered in the compliance rate calculation.

The Medication Kit form in the CRF will serve as the source data for drug amount actually taken, based on tablet counts. The Treatment Exposure form of the CRF will serve as the source data for drug amount that should be taken.

4.6. Extent of Exposure

Treatment duration is defined as number of weeks between the first dose date and last dose date. Subjects will be classified into the following categories by the amount of time on study drug: 0 to 3 months, 3 to 6 months, 6-9 months, 9 to 12 months, 12 to 15 months, 15 to 18 months, and ≥ 18 months (the cutoff values are subject to change if warranted to better represent the exposure data). The number of subjects in each category will be reported. Number of weeks on study drug and the number of cycles the subject started treatment will be summarized as well. The cumulative distribution of treatment duration by cycle or time will also be summarized and presented graphically.

The number and percentage of subjects with dose increases, dose decreases, and dose interruption will be summarized. In addition, subjects with dose modifications and reasons for dose modifications will be listed.

5. EFFICACY

Analysis of efficacy endpoints will be conducted on the ITT Population. Table 4 summarizes the efficacy endpoints and analysis methods to be performed. The analyses on PROs and PK characterization abiraterone acetate and exemestane are detailed in Section 7 and 8, respectively.

Table 4. Summary of Efficacy Analyses to be Performed

Endpoint	Analysis	Analysis Method	Population
Primary			
Progression-Free-Survival	Primary* Sensitivity Subgroup*	Stratified log rank test, Cox regression model Non-stratified log rank test; Multivariate Cox regression model; Alternate censoring rules Non-stratified log rank test, Cox regression model within each subgroup	ITT
Secondary			
Overall Survival	Primary* Sensitivity Subgroup	Stratified log rank test, Cox regression model Alternate censoring rules Non-stratified log rank test, Cox regression model within each subgroup	ITT
Overall Response Rate*		Chi-square test or Fisher's Exact test as appropriate	ITT with measurable disease at baseline
Duration of Response*		Stratified log rank test, Cox regression model	ITT with CR or PR
Change from baseline in endocrine markers		Descriptive summary	ITT with a valid baseline and at least 1 post-randomization value
Exploratory			
Clinical Benefit Rate		Chi-square test or Fisher's Exact test as appropriate	ITT with measurable disease at baseline
Subsequent Therapy		Descriptive summary	ITT
ECOG PS Score*		Descriptive summary, shift table	ITT

CR=complete response; ECOG PS=Eastern Cooperative Oncology Group Performance Status; ITT=intent-to-treat; PR=partial response

Abiraterone acetate plus prednisone (AAP) or abiraterone acetate plus prednisone plus exemestane (AAPE) will be compared with exemestane (E). Nominal p-values will be presented for these comparisons.

*Performed during interim analysis

.

5.1. Analysis Specifications

For each of the pair-wise comparison, the statistical tests of treatment effects on primary endpoint of PFS will be conducted independently at the two-sided 0.10 level of significance. No multiplicity adjustment to the Type I error will be made. The alpha (Type I error) spending at interim and final analyses is specified in Section 3.1.1.

No adjustment for multiplicity testing will be performed for secondary efficacy endpoints. All interval estimations will be reported using two-sided 95% confidence intervals.

5.2. Primary Efficacy Endpoint

5.2.1. Progression Free Survival (PFS)

The primary efficacy endpoint, PFS, is defined as the time from randomization to first occurrence of disease progression (radiographic or clinical), or death from any cause.

Radiographic disease progression is defined by Response Evaluation Criteria in Solid Tumors (RECIST) on measurable lesions captured by CT or MRI. Clinical disease progression is considered only when disease progression cannot be confirmed by CT or MRI, such as when the disease site is skin, bone marrow, or central nervous system.

5.2.2. Analysis Methods

The PFS distribution and median PFS with its 95% confidence interval will be estimated using the Kaplan-Meier product-limit method. The stratified log-rank test accounting for number of prior therapies in the metastatic setting (0 or 1 versus 2) and setting of prior letrozole or anastrozole treatment (adjuvant versus metastatic) will be used as the primary analysis for treatment comparison. A stratified Cox proportional hazards model will be used to assess the magnitude of treatment effect. The model will include stratification factors and will provide estimates of hazard ratio with 95% confidence intervals. The validity of the proportional hazards assumptions will be explored. The following sample SAS code will be used for stratified log-rank test and Cox regression model in the primary analysis.

```
PROC LIFETEST DATA=dataset;
  TIME pfstime*censor(0);
  STRATA stratum1 stratum2;
  TEST trt;
RUN;

PROC PHREG DATA=dataset;
   STRATA stratum1 stratum2;
   MODEL pfstime*censor(0)=trt / RL;
RUN;
```

The non-stratified log-rank test will be conducted as sensitivity analysis. Multivariate Cox proportional hazards model including an independent variable for treatment groups and covariates for the 2 stratification factors will also be conducted. Estimates of hazard ratio with 95% confidence intervals for each covariate will be provided. This analysis is used as sensitivity analysis and to confirm the prognostic effect of stratification factors on PFS. The following sample SAS code will be used for non-stratified log-rank test and multivariate Cox regression model in the sensitivity analysis.

```
PROC LIFETEST DATA=dataset ;
  TIME pfstime*censor(0);
```

```
  TEST trt;
RUN;

PROC PHREG DATA=dataset;
  MODEL pfstime*censor(0)=trt  statum1 stratum2/ RL;
RUN;
```

5.2.3. Disease Progression and Censoring Rules

The date of progression is the earliest date of documented progressive disease based on radiographic assessment, clinical evaluation, and discontinuation due to disease progression. Subjects who were discontinued due to progressive disease will be considered to have disease progression at the date of last dose of study treatment. Death from any cause at any time will also be considered as an event. A listing of tumor assessment will be provided.

The data for subjects who were alive and did not have progression will be censored at the date of the last adequate radiologic assessment which has sufficient evidence to correctly indicate that progression has or has not occurred. The censoring rules for primary analysis of PFS are summarized in Table 5.

Table 5. Censoring Rules for Primary PFS Analysis

Situation	Outcome	Date
Progression	PFS event	Earliest date of documented progression
Death from any cause at any time	PFS event	Date of death
No on-study tumor assessment	Censored	Date of randomization
No progression	Censored	Date of last adequate radiologic assessment
New anti breast cancer treatment started before progression	Censored	Date of last adequate radiologic assessment before subsequent therapy
PFS=progression-free survival		

5.2.4. Impact of Potential Factors on PFS Analysis

Adequacy of Follow-Up

Adequacy of follow-up in each treatment group will be evaluated. The number (%) of subjects who have missed two or more follow-up visits will be tabulated by treatment group. An imbalance in deaths in such subjects could bias the measurement of PFS, because PFS will be artificially prolonged on the group with less adequate follow-up. The number of deaths occurring after 2 or more missing follow-up visits will be tabulated by treatment group.

In case of a large number of subjects having 2 or more missing follow-up visits, a supportive PFS analysis may be preformed to include only well-documented radiographic progression events and deaths for subjects with scheduled follow-up visits. Specifically, the deaths occurring after 2 or more missed follow-up visits will be censored at the last visit when the subjects are still

alive. The deaths occurring before the first progressive disease (PD) assessment planned, or between adequate assessment visits, or after 1 missed visit will be considered as PFS events.

Missing Tumor Assessments

Number (%) of subjects with missing lesion measurements (target lesion or non-target lesion) among the ITT population will be tabulated by cycles for each treatment group. In case of substantial missing data (e.g. 30% or more of lesions), the impact of the missing data on the analysis of PFS will be explored.

Assessment Bias

Tumor assessments will be performed by CT or MRI at baseline and repeated every 2 cycles initially (Day 1 Cycle 3, Day 1 Cycle 5, and Day 1 Cycle 7), and every 3 cycles thereafter (e.g. Day 1 Cycle 10, Day 1 Cycle 13) until, and including, the End-of-Treatment visit. Unscheduled assessments can occur for many reasons (including tumor-related symptoms, drug toxicity). In this open-label study, differences between study groups in the frequency or reason for unscheduled assessments are likely to introduce bias. On the other hand, scheduled visits and radiological assessments in the study are symmetric across the three groups, systematic bias is prevented. To evaluate the effects of bias due to unscheduled assessments, number of subjects with disease progression will be tabulated by assessment (scheduled vs. unscheduled) and treatment groups.

In case of substantial assessment bias due to missing data and unscheduled visit, it may not be appropriate to treat PFS data solely as right censored data. The data for PFS can be better presented and analyzed as interval-censored data. A likelihood based score test developed by Finkelstein (1986) for interval-censored data may be explored to test the difference between treatment groups.

5.2.5. Subgroup Analysis

Subgroup analysis is planned for the primary endpoint PFS to investigate whether treatment effects are consistent within subgroups. Subgroups are formed based on the following subject characteristics and baseline variables for those subjects whose data are available:

1. Number of prior therapies in the metastatic setting (0 or 1 versus 2);
2. Setting of prior letrozole or anastrozole treatment (adjuvant versus metastatic);
3. ECOG PS score (0 versus 1);
4. Number of lines of prior chemotherapy in the metastatic setting (0 versus 1);
5. Age (below median versus above median);
6. Geographic regions (North America, Western Europe, Eastern Europe, others);
7. Sites of metastatic disease (visceral, non-visceral, bone only);
8. PR status (positive vs negative);

9. Time from initial diagnosis to first dose (< median versus >= median);
10. Time from diagnosis of metastatic disease to first dose (< median versus >= median);
11. Time from surgery of removing primary tumor to diagnosis of metastatic disease (< median versus >= median)

The non-stratified log rank test and Cox proportional hazard regression model for PFS will be used for each of the subgroup analysis. Median PFS with 95% confidence interval and hazard ratio with 95% confidence interval (CI) will be provided by treatment group within each subgroup. Subgroup analysis will be presented graphically in forest plot. No formal testing for treatment-by-subgroup interaction will be performed.

5.3. Secondary Endpoints

5.3.1. Overall Survival

5.3.1.1. Definition

Overall survival is defined as from date of randomization to date of death from any cause.

5.3.1.2. Analysis Method

Overall survival will be analyzed using the same analysis methods as used for PFS. The OS distribution and median OS with its 95% confidence interval will be estimated using the Kaplan-Meier product-limit method. The stratified log-rank test and estimates of hazard ratio with 95% confidence intervals from stratified Cox proportional hazards model will be provided.

The same subgroup analysis for PFS may be performed for overall survival if the number of events within each subgroup is sufficient to provide a meaningful interpretation.

5.3.1.3. Censoring and Impact of Crossover

In this study, at disease progression, subjects randomized to the E group may be switched to AAP at the discretion of the investigator.

For primary analysis of OS, survival time of living subjects will be censored on the last date a subject is known to be alive or lost to follow-up. In this ITT estimate of OS, subjects are categorized by their assigned treatment, and subjects' survival after crossover from E to AAP is credited to the E treatment.

As a sensitivity analysis, data from subjects in the E group who crossover to receive AAP will be censored at the date of crossover.

In case of a large amount of treatment crossover and a substantial number of subjects have data that are censored, the following analysis may be explored to correct for dependent censoring:

- Iterative Parameter Estimate (IPE) method as described by Branson (2001) under an accelerated failure time model with the Weibull distribution;
- Rank preserving failure time model described by Robins(1991);

- Weighted log-rank test by the inverse probability of censoring weighted (IPCW) method (Robins 2000, Finkelstein 2011).

5.3.2. Overall Response Rate

5.3.2.1. Definition

Overall response rate is defined as the proportion of subjects with measurable disease achieving a best overall response of either complete response (CR) or partial response (PR) based on RECIST.

5.3.2.2. Analysis Method

All randomized subjects who have a valid baseline value will be included in this analysis. Subjects with missing post-randomization data are considered non-responders. Overall response rate will be summarized using descriptive statistics for categorical data by treatment group. For each pair-wise comparison, the relative risk (AAP vs E, or AAPE vs E) will be reported along with the associated 95% confidence interval. Statistical inference will be evaluated using Chi-square statistic, or Fisher's exact test if the expected counts in some cells are small. The following sample SAS code format will be used:

```
PROC FREQ DATA=dataset;
TABLE trt* response / CHISQ RELRISK;
RUN;
```

Note: CMH option may be used to replace RELRISK option if any cell in the 2 by 2 table has zero count to get the relative risk.

5.3.3. Duration of Response

This analysis will include all randomized subjects with a CR or PR. Duration of Response is measured from the first time that the CR or PR is achieved to the first observation of disease progression based on the RECIST criteria. The same analysis methods used for PFS will be used for duration of response.

5.3.4. Endocrine Markers

Change from baseline in serum endocrine biomarkers (progesterone, estradiol, testosterone, and estrone) will be summarized by treatment at each scheduled post-baseline assessment visit. The analysis will include all ITT subjects who have a valid baseline and at least 1 post-randomization value. Correlation of biomarker levels with clinical response to abiraterone acetate may be explored.

Circulating tumor cells (CTC) and other tumor biopsy biomarkers are optional data collected at some of the investigational sites. Exploratory analysis may be performed separately if there is sufficient data.

Except for the descriptive summary, all the exploratory biomarker analysis will be performed by the J&J Biomarkers Oncology Function and will be reported separately.

5.4. Other Endpoints

Clinical benefit rate is defined as the proportion of subjects with measurable disease achieving a best overall response of a complete response (CR), partial response (PR), or stable disease (SD) based on RECIST. The same analysis methods used for overall response rate will be used for clinical benefit rate.

Number of subjects receiving subsequent anti-cancer therapy will be summarized by therapy type. Listing of subjects with subsequent anti-cancer therapy will also be provided.

Mean values in ECOG values will be summarized by descriptive statistics at each scheduled visit. Minimum and maximum shifts in ECOG performance status score during the study will also be presented.

6. SAFETY

Analysis of safety data will be conducted on the Safety Population, which includes subjects randomized who receive any study medication. The safety variables to be analyzed include adverse events (AEs), electrocardiograms (ECGs), multiple uptake gated acquisition scan (MUGA)/echocardiogram (ECHO), routine hematology such as the complete blood count (CBC) with differential and platelet count, coagulation factors, serum chemistry panel, serum lipids, vital sign measurements, and deaths. Safety variables are to be tabulated by descriptive statistics (n, mean, median, standard deviation, minimum, and maximum; or n and percent). No formal statistical testing is planned.

6.1. Clinical Adverse Events

Adverse events are coded to System Organ Class (SOC) and preferred terms using the Medical Dictionary for Regulatory Activities (MedDRA)® coding system, version 13.1 or greater. For summary purposes, AEs refer to treatment emergent AEs, which is defined as all reported adverse events with a start date between first dose date and 30 days after last dose date.

The severity of AEs will be graded on a scale of 1 to 5 according to the adult NCI Common Terminology Criteria for Adverse Events (NCI CTCAE version 4.0) where higher grades indicate events of higher severity. Adverse events will be summarized by grade 1-4 according to the worst grade experienced. Grade 5 AEs (deaths) will be summarized separately.

All AEs will have their relationship to study drug recorded as not related, doubtful, possible, probable, or very likely on eCRF. Adverse events will be categorized and summarized according

to their highest relationship to study drug. Adverse events reported as possible, probable, or very likely will be classified as drug-related AEs.

Table 7 describes the summaries of AEs that will be provided as incidence tables (number of subjects experiencing an event). A summary of reasons for discontinuation will be provided summarizing the number and percentage of subjects for each reason indicated. Listings of premature discontinuations due to AEs, SAEs, and deaths will be provided. Listings of SAEs and deaths will also be provided.

Table 7. Summary of Adverse Events

Summary	**Sorted By**
AEs	SOC, PT
	SOC, PT, toxicity grade (1 to 4)
Drug-related AEs	SOC, PT
SAE	SOC, PT
	SOC, PT, toxicity grade (1 to 4)
Drug-related SAE	SOC, PT
Grade 3 or 4 AEs	SOC, PT
Drug-related Grade 3 or 4 AEs	SOC, PT
Most common AEs (≥ 5%[a] of the subjects in any of the 3 treatment groups)	SOC, PT
AE leading to death	SOC, PT
AEs leading to dose modification or interruption for abiraterone acetate	SOC, PT
AEs leading to dose modification or interruption for prednisone/prednisolone	SOC, PT
AEs leading to dose modification or interruption for exemestane	SOC, PT
AEs leading to hospitalization	SOC, PT
AEs leading to treatment discontinuation	SOC, PT

AEs of special interest (Selected AEs including osteoporosis and osteoporosis-related fractures, cataract formation, hypokalemia, hypertension, fluid retention/edema, hepatotoxity, anemia, and cardiac Disorders)	PT, severity grade (1 to 4)
Deaths during treatment or within 30 days of last dose	Reason for death
[a] the 5% cutoff is subject to change if warranted to better represent the safety data AE=adverse event; SAE=serious adverse event; PT=preferred term; SOC = System organ class	

6.2. Clinical Laboratory Tests

Laboratory tests will be summarized by hematologic, non-hematologic, and urinalysis, separately.

Selected hematologic laboratory parameters include: hemoglobin, platelet count, white blood cell count (neutrophil, lymphocyte, eosinophil), and coagulation factors (PT, PTT, INR).

Selected non-hematologic lab parameters include: albumin, amylase, blood urea nitrogen, calcium, creatinine, glucose (fasting), lactate dehydrogenase, liver function tests (alkaline phosphatase, AST, ALT, bilirubin), lipids (cholesterol, LDL, HDL, triglycerides), magnesium, phosphorus, potassium, sodium, and total protein.

Descriptive statistics will be provided for the values of selected clinical laboratory tests at each scheduled on-treatment evaluation including the final value (defined as the last observation obtained in the treatment phase) by treatment group. Changes from baseline to each scheduled on-treatment evaluation and to the final value will also be summarized.

A summary of the shifts in selected laboratory hematology and serum chemistry parameters from baseline to the worst toxicity grade during the study will be provided.

6.3. Vital Signs and Physical Examination Findings

Vital sign measurements collected will include blood pressure, heart rate, respirations, weight and body temperature. Mean values and change from baseline in vital signs will be summarized by descriptive statistics at each scheduled visit. The course of blood pressure over time will be presented graphically.

Incidence of abnormalities in vital signs, as categorized in Table 8, will be summarized for safety population who had a baseline assessment and at least one post-baseline assessment for that vital sign.

Table 8. Vital signs abnormality categories

Vital sign	Criteria
Temperature	>38 °C and with >=1 °C increase from baseline
Pulse	>120 bpm and with >30 bpm increase from baseline
	<50 bpm and with >20 bpm decrease from baseline
Systolic blood pressure	>180 mmHg and with >40 mmHg increase from baseline
	<90 mmHg and with >30 mmHg decrease from baseline
Diastolic blood pressure	>105 mmHg and with >30 mmHg increase from baseline
	<50 mmHg and with >20 mmHg decrease from baseline
Respiratory rate	<8 breaths/min and >25% decrease from baseline
	>22 breaths/min and >50% increase from baseline
	>100% increase from baseline
	>100% decrease from baseline

No quantitative post baseline physical examination data are collected in this study. A listing by subject for abnormal physical examination data will be presented with subject ID, treatment, time point, body system, and verbatim examination findings.

6.4. Electrocardiogram

Selected ECG variables include but may not be limited to QTc, PR interval, and QRS. Descriptive statistics will be presented at each scheduled evaluation. A shift table will be generated to indicate the number of subjects who shift categories (normal, clinically non-significant abnormal, and clinically significant abnormal) from baseline to post-baseline.

Subjects will be classified into the following categories by the maximum change from baseline of QTc interval: ≤ 30 msec, 30 to 60 msec, and > 60 msec. Subjects will also be classified into the categories by the maximum absolute values of QTc interval: ≤ 450 msec , 450 to 480 msec, 480-500 msec, and > 500 msec. The number and percentage of subjects in each category will be summarized by treatment groups. Subjects whose QTc increases > 60 msec, or who have QTc > 500 msec on study will be listed. For all tables and summaries on QTc, both the Fredericia and Bazett corrections will be reported.

6.5. Echocardiogram/MUGA (ECHO/MUGA)

Left Ventricular Ejection Fraction (LVEF) will be measured at screening and the End-of Treatment visit. Mean values and change from baseline will be descriptively summarized. A summary table will be generated to indicate the number of subjects in following categories: no change, improved from baseline evaluation, and worsening from baseline evaluation. Number of subjects with absolute decrease from baseline ≥15% and number of subjects with LVEF<50% at the end of treatment will be summarized by treatment group.

7. PATIENT REPORTED OUTCOMES

The following three PRO instruments that assess subject-perceived disease burden, including pain will be collected throughout the study and analyzed at the end of the study.

EORTC-QLQ C30 is a 30-item questionnaire and incorporates a global health status / QoL scale, 5 functional scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, pain, nausea and vomiting), and 6 single items assessing additional symptoms and perceived financial impact of the disease. Scores will be derived using validated scoring algorithms according to EORTC QLQ-C30 Scoring Manual.

Pain intensity domain of BPI-SF consists of 4 items that report the severity dimension of pain over time: pain at its 'worst', 'least', 'average', and 'now' (current pain). Each item is assessed in a 0–10 numerical rating scale where 0 is no pain and 10 is pain as bad as you can imagine.

EQ-5D-5L consists of the EQ-5D descriptive system and the EQ visual analogue scale (EQ-VAS). EQ-5D descriptive system assesses 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). Each dimension is measured on a 5-point ordinal scale ranging between 1 (no limitation) and 5 (greatest limitation) where a higher score corresponds to a worse health state. The EQ-VAS records the respondent's self-rated health on a 100 mm VAS where 0 is 'worst imaginable health state', and 100 is 'best imaginable health state'. The higher scores in VAS reflect a 'better' health state.

PROs will be collected before any other visit procedure within 2 days before randomization (Day 1 Cycle 1, defined to be baseline), repeated every cycle thereafter (e.g., Day 1 Cycle 2, Day 1 Cycle 3), End of Treatment, and every 3 months during Follow-Up.

The PRO data from all the visits including follow-up and cross-over visits will be included for analyses.

7.1. Primary PRO Analysis Endpoints

The PRO measures for primary analysis are listed below:

- BPI-SF: 'Worst' Pain Intensity
- EORTC: global health status/QoL scale; scales of physical, role, emotional, cognitive, and social

For each of the PRO scales, descriptive statistics (number of observations, mean, standard deviation, median, minimum, maximum) of scores at baseline and post-baseline assessments will be reported by treatment groups. The following three analyses will be conducted for each of the PRO scales.

7.1.1. Time to Progression or Deterioration Analysis

Using a pre-defined definition (see Table 9 below) for progression or deterioration specific to each PRO scale, time to progression or deterioration will be compared between treatment groups.

Distributions of time-to-event variables will be estimated using the Kaplan-Meier product-limit method. Median times to event with two-sided 95% confidence intervals will be estimated.

The stratified log-rank test accounting for number of prior therapies in the metastatic setting (0 or 1 versus 2) and setting of prior letrozole or anastrozole treatment (adjuvant versus metastatic) will be used for treatment comparison. A stratified Cox proportional-hazards model will provide estimates of hazard ratios with 95% confidence intervals.

7.1.2. Proportion of Subjects Experiencing Progression or Deterioration Analysis

Using the same threshold value to each PRO scale, the proportion of subjects experiencing progression or deterioration will be compared between treatment groups, using Chi-square test or Fisher's exact test if appropriate. Table 9 summarizes the definition of primary PRO endpoints and analysis methods to be used.

Table 9. Summary of Selected PRO Endpoints of Interest and Analysis Methods

PRO Endpoints	Definition	Analysis Method	Population
Time to pain progression	The time interval from randomization to the first date a subject experiences pain progression. Two definitions will be used separately: (1) the 'worst' pain intensity in the last 24 hours increased by ≥30% from baseline at 2 consecutive evaluations ≥4 weeks apart; (2): an increase from baseline of ≥2 points (in 11-point scale) in the BPI-SF worst pain intensity score in the last 24 hours observed at 2 consecutive visits ≥4 weeks apart. Subjects who have not experienced pain progression at the time of analysis will be censored on the last known date a subject was known to have not progressed. Subjects with no on-study assessment or no baseline assessment will be censored at date of randomization.	Stratified log rank test, Cox regression model	ITT
Proportion of subjects with pain progression	Proportion of subjects experiencing pain progression. Subjects with no post-baseline will be included in the denominator for the proportion calculation.	Chi-square test, or Fisher's exact test if appropriate	ITT subjects with a baseline value
Time to EORTC Deterioration (global QoL, physical, role, emotional, cognitive, and social)	The time interval from randomization to the first date a subject experiences a deterioration, defined as change from baseline ≤ -5 (decreased at least 5) in EORTC scale (in a 0-100 scale). Subjects who have not experienced EORTC deterioration at the time of analysis will be censored on the last known date a subject was known to have not deteriorated. Subjects with no on-study assessment or no baseline assessment will be censored at date of randomization.	Stratified log rank test, Cox regression model	ITT
Proportion of subjects with EORTC Deterioration (global QoL, physical, role, emotional, cognitive, and social)	Proportion of subjects experiencing a deterioration.	Chi-square test, or Fisher's exact test if appropriate	ITT subjects with a baseline value

7.2. Secondary PRO Analyses Endpoints

For all the other PRO measures in the three instruments (EORTC-QLQ C30, BPI-SF, and EQ-5D-5L), descriptive statistics (number of observations, mean, standard deviation, median, minimum, maximum) of scores at baseline and post-baseline assessments will be reported by treatment groups.

- Score values in EORTC-QLQ C30
    - 3 symptom scales: fatigue, nausea and vomiting, pain

- 6 single items: dyspnea, insomnia, appetite loss, constipation, diarrhea, financial difficulties

- Score values in BPI-SF:
  - Least pain
  - Average
  - Current pain

- Score values in EQ-5D-5L:
  - 5 dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression
  - EQ-VAS

7.3. Compliance with PRO Assessment Form

Compliance with PRO assessment form will be evaluated by summarizing the numbers of expected, received, and missing forms. Missing PRO assessments defined as the expected number of assessments for a particular visit minus the actual number of assessments for that visit may be tabulated by treatment group and overall for each visit (baseline and post-baseline including end of treatment visit). Expected number of assessments per visit will be determined by the number of expected subjects and forms to be completed in each group at each visit. Separate tables may be constructed for each instrument.

8. PHARMACOKINETICS (PK) Analyses

At selected study sites, twenty (20) subjects/treatment group will participate in PK assessments. The PK population will include all subjects with sufficient and interpretable PK assessments to estimate the noncompartmental PK parameters.

The plasma abiraterone and exemestane concentrations of at each time point will be summarized by descriptive statistics by treatment group. PK parameters to be estimated for abiraterone and exemestane include C_{max}, t_{max}, AUC_{last}, t_{last}, and AUC_{24h}, and will be descriptively summarized by treatment group. Individual and mean plasma concentration-time profiles of abiraterone and exemestane will be plotted by treatment group. All subjects and samples excluded from analysis will be clearly documented in the study report. Further exploratory PK analyses may be performed, as deemed appropriate.

The PK analyses will be performed by J&J Clinical Pharmacology Function.

REFERENCES

Branson M (2002) and Whitehead J Estimating a treatment effect in survival studies in which patients switch treatment. Statistics in Medicine 2002; 21:2449-2463.

Chia S (2008), Gradishar W, Mauriac L, et al. Double-blind, randomized placebo controlled trial of fulvestrant compared with exemestane after prior nonsteroida aromatase inhibitor therapy in postmenopausal women with hormone receptor-positive, advanced breast cancer: Results from EFFECT. J Clin Oncol 2008; 26:1664-1670.

Finkelstein DM (1986). A proportional hazards model for interval-censored failure time data, Biometrics 1986; 42:845–854.

Finkelstein DM (2011) and Schoenfeld DA. Correcting for discretionary treatment crossover in an analysis of survival in the Breast International Group BIG 1-98 trial by using the Inverse Probability of Censoring Weighted Method. J Clin Oncol 2011; 29:1093-1095.

Hwang IK (1990), Shih WJ, and De Cani JS. Group sequential designs using a family of type I error probability spending functions. Statistics in Medicine 1990; 9:1439-1445.

LAN KKG (2009), Hu P, and Proschan MA. A conditional power approach to the evaluation of predictive power. Statistics in Biopharmaceutical Research 2009; 1(2): 131-136.

Little RJ (2002) and Rubin DB. Statistical analysis with missing data. John Wiley & Sons (New York; Chichester), 2002.

Robins JM (2000) and Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with Inverse Probability of Censoring Weighted (IPCW) log-rank tests. Biometrics 2000; 56: 779-788.

Robins JM (1991) 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.

Farrara JT, Young Jr JP, LaMoreauxb L, 2001. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale, Pain 94:149–158

Logothetis CJ, Basch E, Molina A, 2012. Eff ect of abiraterone acetate and prednisone compared with placebo and prednisone on pain control and skeletal-related events in patients with metastatic castration-resistant prostate cancer: exploratory analysis of data from the COU-AA-301 randomised trial. Lancet Oncol 13: 1210–17

Osoba D, Rodrigues G, Myles J, et al. 1998. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 16:139–144.

Hamidou Z, Dabakuyo TS, and Mercier M. 2011. Time to Deterioration in Quality of Life Score as a Modality of Longitudinal Analysis in Patients with Breast Cancer. The Oncologist 16:1458–1468

ATTACHMENTS

NA