Efficacy and Safety of Trabectedin or Dacarbazine for Metastatic Liposarcoma or Leiomyosarcoma After Failure of Conventional Chemotherapy: Results of a Phase III Randomized Multicenter Clinical Trial

Demetri, et al

Statistical Analysis Plan

A Randomized Controlled Study of YONDELIS® (Trabectedin) or Dacarbazine for the Treatment of Advanced Liposarcoma or Leiomyosarcoma

Protocol ET743-SAR-3007; Phase 3

R279741 (trabectedin)

Issue/Report Date: 17 October, 2013
Document No.: EDMS-ERI-65507261: 2.2

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

TABLE OF CONTENTS

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis populations, derived variables and statistical methods for the analysis of subject information, efficacy and safety. This SAP specifies the preplanned analysis and serves as the basis for the ET743-SAR-3007 clinical study report. Trabectedin (R279741, YONDELIS®) is formerly known as ET-743 (ecteinascidin).

1.1. Trial Objectives

1.1.1. Primary Objectives

The primary study objective is to evaluate whether the overall survival (OS) for the trabectedin group is superior to the dacarbazine group for subjects with advanced L-sarcoma (liposarcoma or leiomyosarcoma) who were previously treated (in any order) with at least: a) an anthracycline and ifosfamide containing regime, or b) an anthracycline containing regimen and 1 additional cytotoxic chemotherapy regimen.

1.1.2. Secondary Objectives

Secondary objectives are to evaluate progression-free survival (PFS), time to progression (TTP), objective response rate (ORR), symptom severity, and safety in the trabectedin group and dacarbazine group.

1.2. Trial Design

This is a randomized, open-label, active-controlled, parallel group, multicenter, phase 3 study. Subjects will be randomized in a 2:1 ratio to receive trabectedin or dacarbazine using one of the following treatment schedules:

- Trabectedin Group: 21-day cycle, 1.5 mg/m^2 as a 24-hr i.v. infusion once every 3 weeks (q3wk 24-h). A central venous catheter must be used to administer study drug to subjects in the trabectedin group. These subjects will be pretreated with 20 mg of dexamethasone i.v. (or an equivalent i.v. corticosteroid) on Day 1 of each treatment cycle, 30 minutes before study drug.
- Dacarbazine Group: 21-day cycle, 1 g/m^2 as a 20- 120 minutes i.v. infusion once every 3 weeks (q3wk).

A total of 570 subjects with unresectable, measurable, locally advanced or metastatic liposarcoma (dedifferentiated, myxoid round cell, or pleomorphic) or leiomyosarcoma previously treated (in any order) with at least a) an anthracycline and ifosfamide containing regimen, or b) an anthracycline containing regimen and 1 additional cytotoxic chemotherapy regimen will be randomized. Subjects are randomized to either of the 2 treatment groups in a 2:1 ratio. The randomization will be stratified by screening

European Clinical Oncology Group (ECOG) Performance Status (0 versus 1), the number of lines of prior chemotherapy (1 versus 2 or more), and L-sarcoma subtype (liposarcoma versus leiomyosarcoma).

Subjects will be evaluated for overall survival, progression-free survival, tumor response, and time to progression. Disease response will be assessed according to Response Evaluation Criteria in Solid Tumors (RECIST Version 1.1). Radiographic assessment of disease, including radiographic imaging of the chest (with lung views), abdomen, and pelvis, will be performed up to 30 days before randomization and every 6 weeks for the first 36 weeks on study and every 9 weeks thereafter, until disease progression occurs, the subject begins subsequent anticancer therapy, the study ends, or the subject dies. Subsequent therapy should only be started after disease progression has been documented. For both treatment groups, tumor assessment schedules will be the same. Additional tumor assessments may be scheduled at the discretion of the investigator. The same method of tumor measurement should be used at each assessment. The investigator will do the evaluation of disease status. Subjects may continue to receive therapy until there is evidence of disease progression or unacceptable toxicity.

One interim analysis of OS will be performed when approximately 188 deaths have occurred. The purpose of the interim analyses will be to terminate the study early if superiority of the trabectedin group will be demonstrated.

The efficacy of each treatment group will be evaluated on the basis of OS, PFS, TTP, objective response (CR + PR) rate, and duration of response.

Safety will be evaluated by adverse events (AEs) and hematology and clinical chemistry test results.

1.3. Sample Size Justification

It was assumed that the hazards for the 2 treatment groups will follow a proportional hazards model for OS. The test to detect a difference between a median OS of 10 months in the dacarbazine group and a median OS of 13.5 months in the trabectedin group (HR=0.74) at an overall 2-sided significance level of 0.05 with a power of 80% requires 376 events. Assuming an enrollment rate of 25 subjects per month over 23 months, a sample size of approximately 570 subjects is planned for the study. One interim analysis will be performed when approximately 50% of the required number of events, 188 death events, is observed. The final OS analysis will be conducted when approximately 376 deaths are observed. It is anticipated that the final analysis will be performed at approximately at 32 months from the start of randomization.

1.4. Randomization and Blinding

Randomization will be performed via a centralized, interactive voice response system (IVRS). In order to assure treatment balance (in a 2:1 ratio between the 2 treatment groups) for number of lines of prior chemotherapy (1 versus 2 or more), ECOG performance status (0 versus 1), and L-sarcoma subtype (liposarcoma versus leiomyosarcoma), subject allocation to a treatment group will be performed using the permuted block randomization method with number of lines of prior chemotherapy (1 versus 2 or more), ECOG performance status (0 versus 1), and L-sarcoma subtype (liposarcoma versus leiomyosarcoma) as the 3 randomization stratification factors.

The study is an open-label trial.

2. ANALYSES PLANNED

The following definitions, study populations, and methods of analyses are planned.

2.1. General Analysis Definitions

2.1.1. Definition of cycle, treatment phase and follow-up phase

Cycle: The duration of cycles may not consistently be the planned 21 days. The duration will depend on the date of the first dose of study drug given in the next cycle. If a subject experiences toxicity at the planned start of a cycle that leads to a delay in study drug administration, this new cycle will be delayed, and the previous cycle will be prolonged (without additional trabectedin/dacarbazine treatment).

The start of a cycle is the date that of first administration of study drug in a cycle. The end of a cycle is the day before the first infusion of the immediate subsequent cycle recorded on the CRF.

The last cycle ends 3 weeks after the start of this cycle.

Treatment phase: Treatment phase starts at the first date of study medication administration and ends at the end-of-treatment visit about 30 days after the last study medication administration. The assessments performed during the end-of-treatment visit will be included in the last cycle for analysis.

All lab samples taken up to 30 days after the last study medication administration will be captured in the last cycle.

Follow-up phase: Follow-up phase begins immediately after end-of-treatment phase. Subjects will be followed-up for the collection of survival status and the use of subsequent anticancer therapy every 60 days for the first 2 years after the last dose of

study drug, and then every 90 days thereafter. Collection of survival status will continue until 376 deaths have been observed, at which time clinical cutoff will occur. Subjects who discontinue study drug before disease progression occurs (e.g., subjects who discontinue treatment due to unacceptable toxicity) will continue to have radiographic assessments of disease. The study end date will be the clinical cutoff date for the clinical study report or 30 days after the last dose of study drug has been administered, whichever is later.

Reference start date:

The reference start date is used to calculate relative study day. It is defined as the date of the first study drug administration. If the first administration date is missing or the administration is never done, then the randomization date will be used.

Baseline:

Baseline is defined as the closest measurement taken prior to or at the time of the first study medication administration. For randomized patients who did not receive study medication administration, the baseline is the closest measurement taken prior to or on the reference start date.

Relative day:

Assessments/events will be presented chronologically by cycle day or study day or both, which are defined in the following:

Cycle day = assessment/event date - date of first study drug administration for the cycle + 1;

Study day =assessment/event date – reference start date + 1 if an assessment/event date is equal to or later than the reference start date.;

Study day = assessment/event date – reference start date if an assessment/event date is earlier than the reference start date..

In addition to the cycle day and study day defined above, the laboratory assessment done on Day 1 of a cycle is considered as the baseline for that cycle as well as the last laboratory assessment for the previous cycle.

2.1.2. Pooling Algorithm for Analysis Centers

For this study, center effect will not be included in the analysis.

2.1.3. Analysis Sets

Three analysis sets are defined and used for analyses.

All randomized subjects analysis set is defined as all subjects who are randomized, independent of whether they received study medication or not.

All treated subjects analysis set is defined as all randomized subjects who receive at least 1 dose of study medication. Subjects who receive dexamethasone as premedication but do not receive trabectedin will not be included in the all treated subject analysis set.

All evaluable subjects analysis set is defined as all randomized subjects with measureable disease who received at least one dose of study drug and for whom at least one post baseline response evaluation (scheduled or unscheduled) is available prior to or on the start date of subsequent anticancer therapy or death. The reference date for evaluable subject analysis will be Cycle 1 Day 1. Measurable disease is defined as having at least one measurable lesion/lymph node at baseline. A measurable lesion is defined as having a diameter of ≥20 mm by chest x-ray or of ≥10 mm by a CT scan, MRI or caliper measurement. A measurable lymph node is defined as having a short axis of ≥15 mm.

2.1.3.1. Efficacy Analysis Set(s)

2.1.3.1.1. Primary Efficacy Analysis Set

Efficacy analysis will be based on the all randomized subjects analysis set.

2.1.3.1.2. Secondary Efficacy Analysis Set

As a sensitivity analysis, the all evaluable subjects analysis set will be used for ORR and best overall response summary analyses.

2.1.3.2. Safety Analysis Set

Safety analysis is based on the all treated subject analysis set.

2.1.3.3. Definition of Subgroups

In addition to the main analysis, the efficacy analysis will be performed, if appropriate, by baseline ECOG performance status (0, 1), the number of lines of prior chemotherapy (1, 2 or more), L-sarcoma subtype (liposarcoma, leiomyosarcoma), race, sex, age and other appropriate subgroups.

2.2. Methods of Analysis

2.2.1. Statistical Hypotheses for Trial Objectives

The primary study objective is to evaluate whether the overall survival for the trabectedin group is superior to the dacarbazine group for subjects with unresectable, measurable, locally advanced or metastatic liposarcoma (dedifferentiated, myxoid round cell, or pleomorphic) or leiomyosarcoma previously treated (in any order) with at least a) an anthracycline and ifosfamide containing regimen, or b) an anthracycline containing regimen and 1 additional cytotoxic chemotherapy regimen.

The general hypotheses used to address this objective are as follows:

H_0: the survival distributions of the trabectedin group, St(t), and the dacarbazine group, Sd(t), are equal at all time points t:

$St(t) = Sd(t)$, for all $t > 0$

versus

H_1: The survival distributions are not equal for at least one time point t:
$St(t) \neq Sd(t)$, for some $t > 0$

These hypotheses will be tested using an unstratified log-rank test. The statistical inference of OS will be carried out within the context of a group sequential testing design as described in Section 2.2.2,

OS will be compared between treatment groups using an unstratified 2-sided log-rank test. One treatment group will be declared better than the other group if the p-value is less than or equal to the significance level as specified by the alpha spending function. The overall 2-sided 0.05 significance level will be spread over two OS analyses, the interim analysis when approximately 188 events are seen, and the final analysis when approximately 376 events are seen. The exact significance levels will be calculated based on the O'Brien-Fleming boundaries as implemented by Lan-DeMets alpha spending method once the exact number of events is known.

2.2.2. OS Interim Analysis

One interim analysis and one final analysis are planned after approximately 50% and 100% of the total required OS events have occurred, respectively. The East® software will be used to obtain stopping boundaries that control the overall 2-sided significance level of 0.05 and provide 80% power to detect a hazard ratio of 0.74 (corresponding to a 35% improvement in median OS for the trabectedin group compared with the dacarbazine group).

The O'Brien-Fleming boundaries, as implemented by Lan-DeMets alpha spending function will be used for the efficacy boundary. Operating characteristics for these boundaries are presented in the following table.

Stopping Boundaries for OS

	Analyses	
Variable	**Interim**	**Final**
Observed OS Events	188	376
Anticipated Time to Analysis (months)	20	32
Anticipated Enrollment (n)	500	570
Efficacy Boundary (HR)	0.639	0.810
Boundary Crossing Prob. (H_0)	0.003	0.047
Cumulative Stop Prob. Under (Ho)	0.003	0.050

HR=Hazard ratio; Ho = 0% improvement; H_1 = 35% improvement

The purpose of the interim analysis will be to terminate the study early if superiority of the trabectedin group is demonstrated. An Independent Data Monitoring Committee (IDMC) has been formed to monitor the safety at regular intervals, and to evaluate efficacy at the interim analysis. The decision to terminate the study will be based on a review from the IDMC who will adhere to the prescribed statistical guideline. The IDMC will review the interim analysis data to monitor the safety of the subjects, to determine if there is early evidence of a difference in efficacy or safety between the two treatment groups, and to provide the sponsor with advice regarding on the conduct of the study. The IDMC activities and responsibilities are provided in a separate IDMC Charter.

PFS analysis

Only one analysis will be conducted for PFS for this study and it will be performed at the time of OS interim analysis. It is anticipated that there will be approximately 331 PFS events and approximately 500 subjects of the 570 planned subjects would have been enrolled in the study at the time of the OS interim analysis. It is estimated that 331 PFS events will provide at least 90% power in detecting a HR of 0.667 (median PFS of 2.5 months for the dacarbazine group versus 3.75 months for the trabectedin) with 2-sided significance level of 0.05.

2.2.3. General Analysis Specifications

All subject characteristics and efficacy analyses will be performed by randomized treatment group. All safety analyses and exposure summaries will be performed by treatment group as treated in the first cycle.

Continuous variables will be summarized and presented with summary statistics, i.e., mean, standard deviation, median and range.

Categorical variables will be summarized in frequency tables. Percentages in the summary tables will be rounded to 1 decimal place.

In case of pretreatment characteristics with multiple measurements per subject before the start of treatment (laboratory assessments, vital signs) the baseline measurement will be considered the last value prior to or on the first day of treatment.

For time-to-event variables, the data will be summarized by the Kaplan-Meier method (Kaplan and Meier, 1958). The treatment group comparison will be presented by the hazard ratio and its 95% CI using Cox proportional hazards model (Cox, 1972) and by p-value using the log-rank test.

2.2.4. Demographics and Baseline Characteristics

Demographics and baseline characteristics will be summarized for all randomized subjects.

Age, baseline weight, height, and body surface area (BSA) will be summarized descriptively. Baseline weight, height, and body surface area (BSA) are recorded on the vital signs CRF page. Baseline body mass index (BMI) will be calculated, using the following formula:

$$\text{BMI} = \text{baseline weight(kg)}/\text{height } (\text{m}^2).$$

BMI will be summarized, and the frequency counts of the different categories (<20, 20 - <25, 25 - <30, ≥30) will be displayed.

Age category (<18, 18 - <65, 65 - <75, ≥75) and race will be summarized by frequency counts.

Baseline ECOG and ECG (normal, abnormal clinical insignificant, abnormal clinical significant) will be summarized.

For the sarcoma history, the tumor subtype will be summarized. Time from initial diagnosis to randomization and time from the latest disease progression to randomization will be calculated in months and summarized descriptively.

Stratification factors will be tabulated.

A frequency tabulation of the number of subjects with a previous surgery for malignancy, radiotherapy for malignancy, or chemotherapy for malignancy will be given. A summary of medical history will be presented per body system.

2.2.5. Disposition Information

Number of subjects randomized and subject disposition will be summarized. A listing of subjects randomized by country/site will also be presented. Subject disposition includes the number of subjects in each of the following categories:

- Randomized
  - Not Treated (subjects who are randomized but not treated with study drug)
  - Treated (subjects who are treated with study drug)
- Evaluable

Reasons for treatment discontinuation will be collected on the CRF and will be summarized overall and by cycle for all randomized subjects by the categories recorded on the CRF page.

Reasons for treatment discontinuation due to adverse event are further grouped into drug-related and non-drug-related adverse events.

2.2.6. Extent of Exposure

Cycle duration:

Cycle duration (weeks) is defined as: [(The first study drug dosing date of the next cycle) - (the first study drug dosing date of the current cycle)]/7.

For the final cycle, cycle duration is equal to the pre-planned duration, i.e. 3 weeks.

Treatment duration:

The treatment duration (weeks) is defined as: [(The last cycle end date) – (the first study drug dosing date) + 1]/7.

Cumulative dose:

Cumulative dose during the treatment is defined as:

sum of (dose infused at a cycle (mg) / body surface area (m^2) from eCRF for trabectedin group for that cycle) for all cycles; and

sum of (dose infused at a cycle (g) / body surface area (m^2) from eCRF for dacarbazine group for that cycle) for all cycles.

Dose intensity:

Dose intensity during the treatment is defined as:

Cumulative dose/(treatment duration (in weeks)/3)

The dose intensity is expressed in mg/m² per 3 weeks for trabectedin group, and in g/m² per 3 weeks for dacarbazine group.

Relative dose intensity:

Relative dose intensity is the dose intensity divided by the planned dose intensity.

For the trabectedin group, the planned dose intensity is 1.5mg/m² per 3 weeks.

For the dacarbazine group, the planned dose intensity is 1.0g/m² per 3 weeks.

Number of cycles:

For each subject, total number of cycles received will be calculated.

Descriptive statistics of treatment duration, cumulative dose, dose intensity, relative dose intensity, and the number of cycles will be presented.

Cycle delay, temporarily infusion interruption and dose reduction:

The number of subjects with a cycle delay, a temporarily infusion interruption, or a dose reduction will be summarized. The number of and reasons for these events will also be tabulated. The reasons for these events are further classified into drug-related AE, non-drug-related AE, or other reason.

2.2.7. Protocol Deviations

Major protocol deviations will be summarized and listed for all randomized subjects.

A summary table with the number of inclusion/exclusion violators will be presented per criterion.

Important protocol deviations are documented during the study execution based on predefined criteria. The detailed criteria are documented in the data management plan and are agreed to by the Janssen R&D project physician. The classification of such protocol deviations will also be reviewed by the Janssen R&D project physician and finalized prior to the database lock.

Protocol deviations will be summarized by categories provided in final database.

2.2.8. Concomitant Medications

Concomitant therapies will be coded using the WHO drug dictionary. The number of subjects receiving each type of therapy during the treatment phase will be tabulated in 2 separate tables: a frequency tabulation of the different therapies that started prior to baseline, and a frequency tabulation of the different therapies that started on or post baseline.

2.2.9. Efficacy

2.2.9.1. Analysis Specifications

2.2.9.1.1. Level of Significance

In general, all hypotheses testing will be performed at a 2-sided 0.05 significance level.

For OS, the overall 2-sided significance level is 0.05. This 0.05 will be spread over 2 analyses by an alpha spending function with O'Brien-Fleming boundaries as implemented by Lan-DeMets alpha spending method. The significance of efficacy will be claimed if the p-value is less than or equal to the significance level as calculated based on the above specified alpha spending function and the observed number of events. The significance of efficacy will be claimed if the p-value is less than or equal to the significance level as calculated based on the above specified alpha spending function and the observed number of events.

Secondary efficacy endpoints are given in Section 2.2.9.4. Among these secondary efficacy endpoints, PFS and ORR are considered the major secondary endpoints. Analyses of PFS and ORR will only be performed at the time of OS interim analysis. These two secondary endpoints will be tested using the Hochberg test procedure to control overall type I error rate at the 2-tailed 0.05 level. Testing will begin with p-value $P_{(2)}$ with corresponding hypothesis $H_{(2)}$, where $P_{(2)}$ is the larger p-value between these two secondary endpoints and $H_{(2)}$ is the corresponding hypothesis. If $P_{(2)} \leq \alpha$, then both hypotheses are rejected. If not, then $P_{(1)}$, the smaller p-value, is compared with $\alpha/2$. If $P_{(1)} \leq \alpha/2$, then the hypothesis $H_{(1)}$ is rejected. Analyses of other secondary efficacy endpoints will not be adjusted for multiple testing.

2.2.9.2. Tumor Assessment

Tumor response will be assessed by the investigators, according to the RECIST Version 1.1 response criteria.

Complete tumor assessments will be performed approximately every 6 weeks for the first 36 weeks of study and every 9 weeks thereafter, until disease progression occurs, the study ends, the subject begins subsequent anticancer therapy, or the subject dies. Using

RECIST Version 1.1, tumor responses are evaluated separately by target lesions, non-target lesions, and new lesions during each assessment. Then the combined results of target, non-target, and new lesions will provide an overall tumor response for this assessment. The response criteria for these are detailed in Section 2.2.9.2.2.

2.2.9.2.1. Response Criteria (RECIST Version 1.1)

Response evaluation for the target lesions, non-target lesions, and new lesions, as assessed by the investigator, will be recorded on the CRF and included in the database.

For details on the RECIST Version 1.1 criteria, see Attachment 1 of the protocol.

2.2.9.2.2. Evaluation of Overall Response for Each Assessment

The overall response of the subject at an assessment cycle will be determined based on the investigator's assessments of target, non-target, and new lesions according to the guidelines below:

Target lesions	Non-target lesions	New lesions	Overall response
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	Not evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR
SD	Non-PD or not all evaluated	No	SD
Not all evaluated	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease; NE = Not evaluable.

For subjects without non-target lesions at baseline, the overall response will equal the target lesion evaluation, except in the case where a new lesion appears and the overall response thus becomes PD. Every effort should be made to document the objective progression even after discontinuation of treatment.

2.2.9.2.3. Date of Disease Progression

For analysis based on radiographic assessment, the date of disease progression will be the earliest date between the date(s) of PD as overall response on the Evaluation of Response eCRF page and the date(s) of PD on the Disease Progression eCRF page based on radiographic criteria.

For analysis based on both radiographic assessment and clinical assessment, the date of disease progression will be the earliest date between the date(s) of PD as overall response on the Evaluation of Response eCRF page and the date(s) of PD based on radiographic assessment or clinical assessment on the Disease Progression eCRF page.

2.2.9.2.4. Best Overall Response

Best overall response is the best response recorded from the start of the treatment until disease progression or the start of subsequent anticancer therapy. The best overall response will be summarized per treatment group in a frequency table with categories: complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), not evaluable (NE).

2.2.9.3. Primary Efficacy Endpoint: OS

2.2.9.3.1. Definition

OS is defined as the time between randomization and death. Subjects who die, regardless of the cause of death, will be considered to have had an event. All subjects who are lost to follow-up prior to the end of the study or who are withdrawn from the study will be censored at the time of the last contact date where the subject is known to be alive. Subjects who are still being treated will be censored at the last available date where the subject is known to be alive.

2.2.9.3.2. Analysis Methods

The "All Randomized Subjects" analysis set will be used for primary analysis. OS will be compared between treatment groups using unstratified 2-sided log-rank test. The Kaplan-Meier method will be used to estimate the distribution functions of OS for each treatment group. The number of events, subjects censored, the estimate of medians, and 95% confidence interval for the medians will be presented. A plot of OS using the Kaplan-Meier method will be presented.

Three-month, 6-month, 9-month, 12-month, 15-month, and 18-month survival rates will be calculated using the Kaplan-Meier method.

As a sensitivity analysis, OS will be compared between the treatment groups by a 2-sided stratified log-rank test using the three stratification factors.

A Cox proportional hazards model will examine the effect of prognostic factors. Prognostic factors will include the following variables as covariates whenever appropriate: age, race, sex, and stratification factors: ECOG (0 vs. 1), L-sarcoma subtype (leiomyosarcoma vs. liposarcoma) and number of lines of prior chemotherapy (1 vs. 2 or more), baseline BMI (<30 vs. ≥30), and time from initial diagnosis to randomization.

These prognostic factors will be analyzed in separate univariate analyses and also in one multivariate analysis.

From the Cox proportional hazards regression, hazard ratio (HR) estimates and their 95% CIs will be estimated for treatment and for the prognostic factors.

To check the proportional hazards assumption, a plot of log(-log (survival)) vs. log(time) will be examined.

2.2.9.4. Secondary Endpoints

2.2.9.4.1. PFS

Definition

PFS is defined as the time between randomization and disease progression (either radiographic progression or clinical progression) or death regardless of the cause of death, whichever occurs first.

Analysis method

For PFS, methods similar to those used to evaluate OS will be used for analysis.

Two-month, 4-month, 6-month, 8-month, 10-month, and 12-month PFS rates will be calculated using the Kaplan-Meier method.

Subjects who progressed or died will be considered to have had an event, except if this event occurs after the start of subsequent anticancer therapy, in which case the subject is censored at the time of last response assessment prior to the initiation of the first subsequent anticancer therapy. Subjects who did not progress nor died (withdrawn consent, lost to follow-up or still being treated without disease progression, or started subsequent anticancer therapy) will be censored at the date of the last response assessment (prior to or on the first day of subsequent anticancer therapy).

PFS based on radiographic and/or clinical assessments will both be analyzed by 2-sided log-rank test. Radiographic PFS (rPFS) based on only radiographic progression will also be analyzed. In the rPFS analysis, disease progression based only on clinical assessment will not be considered as an event. All other censoring rules are the same.

As a sensitivity analysis, PFS will be compared between treatment groups using the all randomized subjects analysis set, by a 2-sided stratified log-rank test.

Handling of missing assessment

Tumor assessments are mandated every 6 weeks for the first 36 weeks and every 9 weeks thereafter until disease progression. Despite this, assessments are sometimes missing (not

performed, not all lesions measured, lost, technically inadequate). If there is a time interval between the last Non-PD assessment and the date of death or the date of progression, the following algorithm will be applied:

> If a subject has an interval of >14 weeks during the first 36 weeks or >20 weeks after 36 weeks, this is an indication that more than one assessments is missing. Then the PFS is censored at the last Non-PD assessment prior to this interval.

2.2.9.4.2. Time to Progression (TTP)

Definition

TTP is defined as the time between randomization and disease progression. Subjects who progressed or died due to disease progression will be considered to have had an event. Subjects who died without evidence of disease progression will be considered censored at time of the last response assessment before death. All other censuring rules will be the same as that for PFS.

Analysis method

For TTP, analysis methods similar to those used to analyze PFS will be used.

2.2.9.4.3. ORR

Definition

Objective response (OR) is defined as having a "CR" or "PR" as best overall response. Detail of the tumor response evaluation is described in Section 2.2.9.2. ORR is calculated as the number of objective responders divided by the number of subjects in the all randomized subjects analysis set (for the primary analysis of ORR) or by the number of subjects in the all evaluable subject analysis set (for the sensitivity analysis of ORR).

Analysis method

ORR will be compared between the 2 treatment groups using the Fisher's exact test. Response rate and the associated 95% confidence interval (CI) will be provided for each treatment group.

Best overall response will be summarized per treatment group in a frequency table with the categories: complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), and not evaluable (NE).

2.2.9.4.4. Clinical Benefit Rate (CBR)

Definition

CBR is defined as proportion of subjects who achieve a best overall response of CR or PR, or SD (with duration ≥18 weeks).

Analysis method

Same analysis method for ORR will be used for CBR.

2.2.9.5. Other Secondary Endpoints

2.2.9.5.1. Duration of Response (DR)

Definition

DR is defined only for subjects who have CR or PR as Best Overall Response. Duration of response is calculated from the date of first documentation of response to the date of disease progression or death due to PD. Subjects who have not progressed nor died due to PD will be censored at their last response assessment date. Any subject who starts subsequent therapy without prior disease progression will be censored at the last response assessment date prior to or on the initiation of the start of the first subsequent therapy.

Analysis method

DR will be compared between treatment groups using the log-rank test. The Kaplan-Meier method will be used to estimate the distribution of DR for each treatment group. The number of events, subjects censored, the estimate of medians, and 95% confidence interval for the medians will be presented.

2.2.9.5.2. Duration of Stable Disease

Duration of stable disease is defined, for subjects with best overall response of SD, as time from randomization to date of initial documented PD (either radiographic PD or clinical PD) or date of death due to PD, whichever occurred earlier.

Same analysis method for duration of PFS will be used for duration of stable disease.

2.2.9.6. Supportive Analyses

2.2.9.6.1. Symmetry of Tumor Assessment Schedules

Time to the first 6 tumor assessments

It is defined as the number of weeks between baseline (C1D1) and the recorded date of the first 6 post baseline radiographic assessments.

Time to the first 6 tumor assessments will be presented by box plot.

2.2.9.7. Patient Reported Outcomes (PRO) Analysis

M.D. Anderson Symptom Inventory (MDASI) scores will be used to assess subjects' perceived symptom burden and to determine the impact of treatment on symptom change or stability.

The endpoints for the PRO analysis will be the change from baseline in mean score of all symptom severity items, mean score of all symptom interference items and each individual item scores. The change in these PRO scores between baseline and post-baseline assessment will be summarized.

Mean score of the MDASI symptom severity items will be obtained when patients scored at least 7 of the 13 items using the formula: (sum of scores for all items answered) / (total number of items answered). The mean score of the interference items will be calculated in a similar way. The mean score of symptom interference items will be calculated if more than 50% (four of six items) were completed: (sum of items answered) / (total number of items answered).

2.2.9.8. ECOG

The ECOG value at the end-of-treatment visit will be cross-tabulated against the baseline value.

2.2.9.9. Subsequent Anticancer Therapy

Subsequent anticancer therapy will be coded using the WHO drug dictionary and summarized by ATC class and generic term. The number of lines of subsequent anticancer therapy will be summarized using descriptive statistics.

The accompanying listing will contain details on the type of follow-up therapy as well as the start date.

2.2.10. Safety

2.2.10.1. Adverse Events

A treatment emergent adverse event (TEAE) is defined as any AE occurring on or after the first treatment of study drug, and within 30 days after the last dose. For imputation rules for incomplete dates, see attachment Imputation Rules for Incomplete Dates. If after imputation the start date of an AE is still missing, this AE will be treated as a TEAE.

Adverse events are documented on the CRF together with their severity, according to the NCI CTC version 4.0, also referred to as NCI toxicity grading. For the categorization of the adverse events, the MedDRA dictionary (version 16.0) will be used. Adverse events are considered drug related when the relation to trial medication is considered to be possible, probable, or very likely, according to the investigator's opinion.

The summary of overall adverse events will be done by system organ class and preferred term, by severity (worst toxicity grade), by relationship to study drug, and by AE

outcome. Tables will be sorted by system organ class/preferred term and by the highest incidence in trabectedin treatment group.

A listing will be prepared on all treatment-emergent adverse events. If data are available, a listing for delayed adverse events (defined as new onset of AE occurring 30 days after last dose as judged by investigator to be related to trial medication) will be presented.

Grade 3 and 4 adverse events will be reported. The drug-related grade 3 and 4 adverse events will be summarized.

A frequency table will be made for the AEs leading to drug interruption, dose reduction, or withdrawal of study medication. Adverse events leading to permanent stop and AE with fatal outcome will also be presented by drug relatedness.

For all serious adverse events (SAEs), the investigator has to send an SAE form to the Janssen R&D Drug Safety and Surveillance department. The AEs that are indicated on the CRF as serious are used for the summary tables. A summary table by any grade SAEs will be presented. A similar table will be presented for the drug-related SAEs.

Adverse events will also be tabulated by following special AE groups: thrombocytopenia and bleeding, neutropenia and infections, sepsis or septic shock, CPK elevations or rhabdomyolysis, catheter related complications, hepatobiliary disorders, multi-organ failures, cardiac disorders, and renal disorders. Incidence tables of these different groups will be created for all AEs, for drug-related AEs, for serious AEs.

2.2.10.2. Clinical Laboratory Tests

2.2.10.2.1. Hematology

Laboratory results will be classified according to the NCI CTC version 4.0 . For absolute neutrophil count (ANC), platelet count, hemoglobin and white blood cell (WBC) count, the worst grade per subject will be tabulated overall during treatment and per cycle.

Overall cross tabulation will be presented for the worst grade during treatment versus the baseline toxicity grading.

A listing for those whose laboratory values reach grade 3 or 4 anytime during the study will be provided.

Time to and duration of the first occurrence of a grade 4 neutropenia and thrombocytopenia

Time to first occurrence of a grade 4 neutropenia will be calculated for all treated subjects who have a grade 4 ANC, as the time from the first treatment until the onset day of the first occurrence of grade 4 ANC. Similarly, time to first occurrence of a grade 4 thrombocytopenia is calculated based on grade 4 platelet count.

The duration of a grade 4 neutropenia is defined as the period from the date of first occurrence of a grade 4 ANC to the date of recovery from the grade 4 toxicity during the treatment phase. The date of recovery is defined as the first date that the ANC toxicity grade is smaller than 2. Duration of grade 4 thrombocytopenia is calculated in a similar fashion.

Descriptive statistics will be presented for the time to and duration of grade 4 neutropenia and thrombocytopenia.

Evolution of the ANC values over time will be analyzed. A summary table and a figure will be presented for all subjects with a grade 3 or 4 neutropenia value. A figure will also be created showing median of ANC nadir by cycle for all treated subjects.

Laboratory assessed on day 1 of each cycle is considered as the baseline value of that cycle as well as the last value for the previous cycle.

2.2.10.2.2. Serum Chemistry

Similar to hematology analysis, the worst grade during treatment will be cross tabulated to the baseline grade for creatinine, alkaline phosphatase, AST, ALT, total bilirubin, direct bilirubin, creatine phosphokinase, and albumin.

Overall cross tabulations will be presented for the worst grade during treatment versus the baseline toxicity grading.

A listing for those whose laboratory values reach grade 3 or 4 anytime during the study will be provided.

Time to and duration of the first occurrence of a grade 4 ALT/AST

Time to first occurrence of a grade 4 ALT will be calculated for all treated subjects who have a grade 4 ALT, as the time from the first study drug treatment until the onset day of the first occurrence of grade 4 ALT. Similarly, time to first occurrence of a grade 4 AST is calculated based on grade 4 AST.

The duration of a grade 4 ALT is defined as the period from the date of first occurrence of a grade 4 ALT the date of recovery from the grade 4 toxicity. The date of recovery is

defined as the first date that the ALT toxicity grade is smaller than 2. Duration of grade 4 AST is calculated in a similar fashion.

Descriptive statistics will be presented for the time to and duration of grade 4 ALT and AST.

Evolution of the ALT and AST values over time will be analyzed. A summary table will be presented for all subjects with a grade 3 or 4 ALT and AST. A figure showing median values of ALT for subjects with a grade 3 or 4 ALT and a figure showing median of ALT peak value by cycle for all treated subjects will also be created.

Laboratory assessed on day 1 of each cycle is considered as the baseline value of that cycle as well as the last value for the previous cycle.

2.2.10.2.3. Hepatotoxicity Cases Analysis

For subjects with any elevated transaminases (AST or ALT) of ≥3xULN, ALP (liver fraction) ≤2xULN, and associated with an increase in total bilirubin ≥2xULN, a listing for all subjects with all such records will be produced and a summary table of number of such subjects by treatment group will also be generated. The same listing will be provided for subjects meeting eDISH (Evaluation of Drug-Induced Serious Hepatotoxicity) criteria. The eDISH criteria is defined as elevated AT (AST or ALT) of ≥3xULN, and associated with an increase in total bilirubin ≥2xULN.

2.2.10.3. Vital Signs and Physical Examination Findings

Vital signs (pulse, respiration, blood pressure, as well as temperature) and body weight will be measured at screening visit only. A physical examination will also be performed at screening visit only. Document any clinically significant abnormal change in physical findings, including vital signs and ECOG Performance Status score, as an adverse event. No analysis will be performed on vital signs and physical examination results.

2.2.10.4. Deaths

Deaths during treatment or within 30 days from the last study drug administration will be tabulated, as well as the primary cause of death. A distinction will be made between drug related and non-drug-related AEs.

A similar table will be made, specifying all deaths with the corresponding causes. Separation will be made between subjects who die within 60 days after randomization, subjects who die during treatment or within 30 days after the last study drug treatment, or subjects who die more than 30 days after the last study drug treatment.

2.2.10.5. Safety Narratives

Safety narratives will be written based on the following 4 criteria:

- **Criterion 1:** Deaths for reasons other than disease progression that occurred within 30 days of the last dose of study medication.
- **Criterion 2:** Drug-related treatment-emergent serious adverse event(s).
- **Criterion 3:** Drug-related treatment-emergent adverse events that led to discontinuation of study treatment.
- **Criterion 4:** Any Grade 4 drug-related treatment-emergent adverse events(s) in trabectedin group.
- **Criterion 5:** Any Grade 3 or greater treatment-emergent adverse event(s) of sepsis or septic shock that occurred within 30 days of the last dose of study medication.
- **Criterion 6:** Any elevated transaminases (AST or ALT) of ≥3xULN, ALP (liver fraction) ≤2xULN, and associated with an increase in total bilirubin ≥2xULN that occurred within 30 days of the last dose of study medication.

REFERENCES

Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *J Am Stat Assoc*. 1958; 53:457–48.

Cox DR. Regression models and life-tables. *J Royal Statist Soc B*. 1972; 34(2): 187–220.