Polyphor Ltd

POL6326-009

An International, Phase 3, Multicenter, Randomized, Open-Label Trial Comparing Balixafortide in combination with Eribulin versus Eribulin alone in Patients with HER2 negative Locally Recurrent or Metastatic Breast Cancer

14MAY2021

Statistical Execution Plan

Version 1.0

Prepared by:

PPD
Granta Park, Great Abington
Cambridge, CB21 6GQ
UK

PPD	PPD Biostatistics and Programming
	Statistical Analysis Plan (SAP) Client Approval Form

Client:	**Polyphor Ltd**
Protocol Number:	**POL6326-009**

Document Description:	Final Statistical Execution Plan (FDA)
SAP Title:	An International, Phase 3, Multicenter, Randomized, Open-Label Trial Comparing Balixafortide in combination with Eribulin versus Eribulin alone in Patients with HER2 negative Locally Recurrent or Metastatic Breast Cancer
SAP Version Number:	1.0
Effective Date:	14 May 2021

Author(s):

For PPD: Gabriele Lapini, Biostatistician II

Approved by:

Stephane Lavigne
Principal Biostatistician
I approve this document
14 May 2021 11:46:28 +02:00
DocuSign

Stéphane Lavigne, Principal Biostatistician
Biostatistics, PPD

Date *(DD-MMM-YYYY)*

KJCarroll

20 May 2021

Kevin J Carroll, Statistician
KJC Statistics Ltd

Date *(DD-MMM-YYYY)*

Frank Weber, CMO
Polyphor Ltd.

21 S 2021

Date *(DD-MMM-YYYY)*

TABLE OF CONTENTS

List of Abbreviations

AE	Adverse Event
AEPI	Adverse Event of Particular Interest
AESI	Adverse Event of Special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANCOVA	analysis of covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BSA	Body Surface Area
CBR	Clinical Benefit Rate
CI	Confidence interval
CISH	Chromogenic In Situ Hybridization
CR	Complete Response
CSP	Colony Stimulating Factor
CT	Computed Tomography
CXCR4	C-X-C Chemokine Receptor Type 4
DCR	Disease Control Rate
DSMB	Data Safety Monitoring Board
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EORTC	European Organization for Research and Treatment of Cancer
EOT	End of Treatment
FACT-B	Functional Assessment of Cancer Therapy - Breast
FACT-G	Functional Assessment of Cancer Therapy - General
FAS	full analysis set
FDA	Food and Drug Administration
FISH	Fluorescence in Situ Hybridization
GCP	Good Clinical Practice
HEENT	head, ears, eyes, nose, throat
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hormone Receptor
HRQoL	Health-Related Quality of Life
ICF	Informed Consent Form
ICH	International Conference on Harmonization
IHC	Immunohistochemistry
IRC	Independent Review Committee
IRT	Interactive Response Technology
iRECIST	Response Evaluation Criteria in Solid Tumours (RECIST version 1.1) in cancer immunotherapy trials
ITT	Intent-To-Treat
IV	Intravenous
KM	Kaplan-Meier
MBC	Metastatic Breast Cancer
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model Repeated Measures

MRI	Magnetic Resonance Imaging
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NE	Not Evaluable
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease
PE	Physical Exam
PFS	Progression-Free Survival
PK	Pharmacokinetic
PP	Per Protocol
PR	Partial Response
PRO	Patient Reported Outcome
PT	preferred term
QoL	Quality of Life
RE	Response Evaluable
RECIST	Response Evaluation Criteria in Solid Tumors
REML	Restricted Maximum Likelihood
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Stable Disease
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
Tmax	Time to Maximum Concentration
TNBC	Triple-Negative Breast Cancer
ULN	Upper Limit of Normal
WHO-DDE	World Health Organization Drug Dictionary Enhanced

1. Introduction

This document describes the planned statistical analyses of the data captured in Polyphor Ltd. protocol POL6326-009 "An International, Phase III, Multicenter, Randomized, Open-Label Trial Comparing Balixafortide in combination with Eribulin versus Eribulin alone in Patients with HER2 negative Locally Recurrent or Metastatic Breast Cancer" Final Protocol Version 3.0, 15 July 2019. This Phase III trial is conducted in accordance with the protocol, Good Clinical Practice (GCP) and all applicable regulatory requirements.

The data from protocol POL6326-009 will serve a dual purpose: (i) to provide efficacy and safety data in the 3rd line+ patient population of patients with locally recurrent or metastatic breast cancer primarily for regulatory submissions in the US and jurisdictions in which the 3rd line+ eribulin label applies and (ii) to provide efficacy and safety data in the overall population of patients who receive study treatment in the 2nd line+ patient population , for regulatory submissions in the EU and jurisdictions in which the 2nd line+ eribulin label applies.

This execution plan describes the analyses and data summaries in relation with the 3rd line+ population.

2. Objectives

2.1. Primary Objective

The primary objective of the trial in the 3rd line+ population of patients with locally recurrent or Metastatic Breast Cancer (MBC) is to evaluate the efficacy of balixafortide + eribulin versus eribulin monotherapy on the co-primary endpoints of objective response rate (ORR) and progression free survival (PFS).

2.2. Secondary Objectives

In the 3rd line+ population of patients, secondary objectives are:

- To compare the overall survival (OS) between patients in the balixafortide + eribulin treatment arm versus eribulin monotherapy treatment arm.
- To compare measures of tumor response between patients in the balixafortide + eribulin treatment arm versus eribulin monotherapy treatment arm.
- To evaluate the safety and tolerability of balixafortide + eribulin versus eribulin monotherapy.

2.3. Exploratory Objectives

In the 3rd line+ population of patients, exploratory objectives are:

- To determine whether the response to treatment correlates with baseline estrogen receptor (ER), progesterone receptor (PgR) or C-X-C Chemokine Receptor Type 4 (CXCR4) expression level.
- To assess tumor tissue and circulating biomarkers and possible association with patient treatment outcome (i.e response or anti-tumor efficacy).
- To explore immune response, as assessed by Response Evaluation Criteria in Solid Tumours (RECIST version 1.1) in cancer immunotherapy trials (iRECIST).

- Measurement of plasma concentration of balixafortide and potential metabolites to integrate in a population PK model.
- To assess the Quality of Life (QoL) as reported by patients in the balixafortide + eribulin treatment arm versus eribulin monotherapy treatment arm using standard QoL assessments.

3. Investigational Plan

3.1. Overall Study Design and Plan

This is an open-label, 1:1 randomized, two-arm, multi-center, international pivotal Phase 3 trial designed to investigate the efficacy and safety of balixafortide in combination with eribulin relative to eribulin alone in patients with Human Epidermal Growth Factor Receptor 2 (HER2) negative, locally recurrent or metastatic BC that have previously been treated with one to four chemotherapeutic regimens for metastatic disease. Patients will have previously received an anthracycline and a taxane treatment.

Patients will be stratified according to:

- Line of therapy (2nd line versus 3rd line or later therapy) for locally recurrent or metastatic BC.
- Hormone receptor (HR) status (positive versus negative) based on estrogen receptor (ER) or progesterone receptor (PgR) status.
- Cyclin-dependent kinase (CDK) 4/6 inhibitor treatment received previously (received a CDK 4/6 inhibitor versus not received a CDK 4/6 inhibitor).
- Visceral versus non-visceral disease. Visceral metastases will include brain, pulmonary, pleural, hepatic and peritoneal involvement (including pleural effusions and ascites). Any other sites of metastases (e.g. bone, lymph nodes, and skin) will be considered to be non-visceral. Patients with visceral metastases, irrespective of the presence of any other metastatic sites (e.g. bone), will be categorized as visceral. All other patients without visceral metastases will be categorized as non-visceral.

A total of 384 patients will be randomized into this trial of which 320 will be 3rd line+. As mentioned above, this execution plan focusses on the analysis and reporting of the 3rd line+ population.

During the treatment, disease assessments will be performed every 6 weeks (±7 days) from the date of randomization for the first year and then every 12 weeks (±7 days) thereafter until Progressive Disease (PD) is documented by RECIST v1.1.

Each assessment will be performed as scheduled according to the calendar regardless of any dosing delay to prevent the introduction of bias into the assessment of efficacy. Tumor assessments will be performed until whichever of the following occurs first:

- radiographically and/or objective (i.e., for photographed or palpable lesions) documented as PD per RECIST v1.1,
- discontinuation of patient from overall trial participation (e.g. death, the patient withdraws consent, lost to follow-up).

The local Investigator will assess the presence of disease following CT/MRI scan performed at Screening, and those performed subsequently, and record the details in the patient's source data and eCRF relative to the previous assessment of disease status/metastatic burden.

A blinded Independent Review Committee (IRC) will perform a retrospective review of radiographic images and relevant clinical information collected on-study to assess the protocol defined endpoints of disease response and progression.

Following initial documentation of an objective response, a confirmatory CT/MRI scan will be performed at least 4 weeks later.

Patients discontinuing the study treatment for reasons other than death will attend an End of Treatment (EoT) visit as soon as possible, within 7 days after discontinuation and prior to initiation of any new anti-cancer therapy, regardless of the reason for discontinuation. A physical examination will be conducted, adverse events (AEs), Eastern Cooperative Oncology Group (ECOG) performance status, vital signs, clinical laboratory,12-lead Electrocardiogram (ECG), concomitant medications and Patient reported health outcome QoL questionnaires will be assessed and recorded for all patients at the EOT visit. Furthermore, tumor assessments should be conducted for patients who discontinued from study treatment in the absence of PD.

A 30-day safety follow-up visit will be performed within 30 days (and no later than 37 days) from the last dose of study treatment. Patients discontinuing the study treatment phase for any reason, without PD (i.e. due to unacceptable adverse events) will enter the long-term follow-up (PD follow-up and survival follow-up).

Patients who are discontinued from study treatment for any reason, in the absence of PD, will undergo repeat imaging and tumor response assessments every 8 weeks ±7 days (every 12 weeks ±7 days if the patient has been on the study for >= 1 year), regardless of regimen, until PD is documented, or death occurs, or patient is lost to follow-up, or the patient withdraws consent (whichever occurs first). If a patient stops study treatment and begins another anti-cancer therapy before PD is documented, the new anti-cancer treatment should be recorded in the patient's source data and eCRF and every effort should be made to perform tumor evaluation in these patients until disease progression.

All patients will be followed for OS status every 6 months until death, until the patient withdraws consent to follow-up for survival or until the patient is lost to follow-up (whichever occurs first).

Patients will complete the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC-QLQ-C30) Quality of Life (QoL), Functional Assessment of Cancer Therapy Breast (FACT-B) and EuroQol-5D (EQ-5D) questionnaires during the screening period, prior to first dose, and at the start of each subsequent cycle of study treatment during while still in receipt of study treatment.

An independent data safety monitoring board (DSMB) will review patient safety during the trial and will provide oversight of the planned, formal interim efficacy analyses.

The enrolment period is expected to be approximately 12 months, with an expected 12 months further follow-up prior to the analysis of PFS and an additional 12 months follow-up thereafter to the final analysis of OS. The overall trial duration (from first patient randomized until the end of the trial) is therefore projected to be approximately 36 months.

3.2. Study Endpoints

3.2.1. Co-Primary Efficacy Endpoint

In the 3rd line+ population intended for regulatory submissions in the US and jurisdictions on which the 3rd line+ eribulin label applies, the co-primary endpoints are:

- ORR (confirmed CR + confirmed PR) according to RECIST v1.1 guidelines, as assessed by the IRC.
- PFS according to RECIST v1.1 guidelines, as assessed by the IRC.

3.2.2. Key Secondary Efficacy Endpoint

The key secondary efficacy endpoint subject to Type I error control is OS.

3.2.3. Other Secondary Efficacy Endpoints

Other secondary efficacy endpoints are considered supportive and will not be subject to overall Type I Error control. These endpoints are:

- PFS according to RECIST v1.1 guidelines, as assessed by the local Investigator's review.
- ORR (confirmed CR + confirmed PR) according to RECIST v1.1 guidelines, as assessed by the local Investigator's review.
- Clinical Benefit Rate (CBR) [proportion of patients with confirmed CR, confirmed PR, or stable disease (SD) ≥6 months] according to RECIST v1.1 guidelines, as assessed by (i) the IRC and (ii) the local Investigator's review.
- Disease control rate (DCR; number of patients with confirmed CR, confirmed PR, or SD) according to RECIST v1.1 guidelines, as assessed by (i) the IRC and (ii) the local Investigator's review.
- Time to response, as assessed by (i) the IRC and (ii) the local Investigator's review.
- Duration of response, as assessed by (i) the IRC and (ii) the local Investigator's review.

3.2.4. Secondary Safety Endpoints

These will include:

- Type, frequency and severity of AEs (including serious adverse events [SAEs], adverse events of Special Interest [AESIs]).
- Laboratory abnormalities.
- Vital signs.

Additionally, signs and symptoms of anaphylaxis will be captured and summarized in accordance with guidance provided by the Second Symposium on the Definition of Anaphylaxis. Evaluation of liver parameters will be made according to the evaluation of Drug-induced Serious Hepatotoxicity (eDISH) criteria; this requires a log-log plots of alanine aminotransferase (ALT)/aspartate aminotransferase (AST) vs total bilirubin to identify potential cases of concern, followed by time course plots of all the identified potential cases.

3.2.5. Exploratory Efficacy Endpoints

- Efficacy as measured by PFS, ORR and OS in each exploratory subgroup (Demographic and Baseline Characteristics, Disease History and Progression Status).
- Response, as assessed by iRECIST (2017) in those patients who continue treatment despite progression.
- EORTC-QLQ-C30 QoL change from baseline to the end of each treatment cycle.
- The FACT-B and EQ-5D change from baseline to the end of each treatment cycle.

The following endpoints will be analyzed in a separate plan:

- Relationship between ORR, PFS, or OS and baseline biomarkers including ER status, PgR status or CXCR4 expression levels.
- Relationship between ORR, PFS, or OS and exploratory biomarkers (from tumor tissue or blood) including cytokines (e.g. IFN-gamma), immune cells profile, and ribonucleic acid (RNA) expression.
- Plasma concentration of balixafortide and potential metabolites (PK analysis set).

3.3. Treatments

Patients will be randomized as follows:

3.3.1. Balixafortide + eribulin

Patients will receive balixafortide + eribulin treatment. Balixafortide will be administered on Days 1−3 and Days 8−10 and eribulin will be administered on Days 2 and 9 of each 21-day cycle.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
POL6326	1	2	3	treatment break				8	9	10	treatment break										
Eribulin		2	treatment break						9	treatment break											

3.3.2. Eribulin monotherapy

Patients will receive eribulin monotherapy*;* Eribulin will be administered on Days 2 and 9 of each 21-day cycle.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Eribulin		2	treatment break						9	treatment break											

Balixafortide will be administered, at a dose of 5.5 mg/kg, intravenously over 2 hours ±10 min; however, the infusion time of balixafortide can be increased to a maximum of 3 hours at the discretion of the treating physician, for reasons of patient tolerability (i.e. to manage infusion related reactions). Eribulin should be administered intravenously as the ready to use solution at a dose of 1.4 mg/m^2 over 2 to 5 minutes on Days 2 and 9 of each 21-day cycle. **note:** this dose is expressed as the salt, eribulin mesylate; however, in some countries this dose is expressed as the

base, equivalent to eribulin 1.23 mg/m^2. Eribulin will be administered within 45 minutes after the end of balixafortide infusion.

3.4. Dose Adjustment/Modifications

3.4.1. Dose Modification Balixafortide

In the event of significant treatment-related toxicity, balixafortide + eribulin dosing may be modified in two ways:

- Within a cycle: dosing interruption during a given treatment cycle until there is adequate recovery;
- Between cycles: administration of the next cycle may be delayed due to persistent toxicity.

Patients that discontinue from balixafortide treatment due to treatment-related toxicity are to attend an EoT Visit within 7 days after discontinuation, the 30-day safety follow-up visit for assessments and recording of AEs and concomitant medications, and the Long-term Follow-up visits.

3.4.2. Dose Modification Eribulin

The dose of eribulin can be decreased or modified according to the US regulatory label for eribulin as detailed in the protocol.

3.4.3. Dosing Delays or Interruptions

Eribulin should not be given on Day 2 or Day 9, if any of the following occur:

- ANC <1000/mm^3.
- Platelets <75000/mm^3.
- Common Terminology Criteria for Adverse Events (CTCAE) Grade 3 or 4 non-hematological toxicities.

The Day 9 eribulin dose may also be delayed for a maximum of 1 week if toxicities:

3.5. Do not resolve or improve to CTCAE Grade ≤ 2 severity by Day 16, omit the dose.Subscale Scoring for the FACT-B

-
- Resolve or improve to CTCAE Grade ≤ 2 severity by Day 16, administer the eribulin at a reduced dose and initiate the next cycle no sooner than 2 weeks later.

If study treatment during a cycle is delayed for <14 days, then treatment can be resumed during that cycle. In these circumstances, the total duration of a cycle could potentially be extended up to 28 days. If study treatment during a cycle is delayed for ≥14 days, the treatment for that cycle is not resumed and the subsequent cycle is initiated as planned. If the next cycle cannot be initiated as planned, the patient should be discontinued from treatment unless discussed and agreed otherwise with the Sponsor.

The doses of balixafortide on Days 1, 2 and 3 are interdependent with the eribulin dose on Day 2. Similarly, the doses of balixafortide on Days 8, 9 and 10 are interdependent with the eribulin dose

on Day 9. Therefore, if an eribulin dose is delayed by a few days during a cycle, then the adjunctive balixafortide doses must also be delayed accordingly during that cycle.

Tumor assessments must continue to be performed according to the Study Schedule regardless of any dosing delay.

4. General Statistical Considerations

Summary statistics for continuous variables will include the mean, standard deviation, median, and range (minimum and maximum); the number of missing observations will also be displayed. The mean and median will be presented to one decimal place beyond which the data were captured. The standard deviation will be presented to two decimal places beyond which the data were captured. The minimum and maximum will be presented to the precision with which the data were captured.

Categorical variables will be presented as frequency counts and percentages. A row or column denoted 'missing' will be included in count tabulations where necessary to account for dropouts and missing values. Percentages will be rounded to 1 decimal place and the percent will be suppressed when the count is zero. The denominator will be the number of patients in that dose schedule within the analysis set of interest unless otherwise noted.

Time-to-event variables will be analyzed using the Kaplan-Meier (KM) method. The KM estimates for quartiles and the confidence interval (CI) for the median will be presented as described below. Kaplan-Meier curves will also be plotted.

Time-to-event variables will be also analyzed by a Cox proportional hazards model.

Data listings will be created to support each table and to present all data. Data listings will be presented by treatment arm and patient number.

Although this is an open label trial, grouped evaluation of randomized treatment arms will be performed only for the pre-planned interim efficacy analyses and DSMB safety reviews as described in Section 8.0 of this document and in the DSMB charter. In order to protect the integrity of the trial and the robustness of the results, project team members, both Polyphor Ltd. and CRO personnel, will not evaluate, formally or informally, data grouped by randomized treatment arm as data are accruing over time. Procedures relating to the execution of pre-planned interim efficacy analyses and DSMB safety reviews are described in the DSMB Charter.

Important study variables to be used in the analysis, defined as follows:

- Age: Recorded as "Age" in the eCRF instead of the Date of Birth.
- Baseline:
    - For ***efficacy evaluations***, the last available assessment before or at date of randomization is taken as "baseline" value or "baseline" assessment. For PRO endpoints, baseline is the mean of the PROs taken at screening and at last visit before first treatment administration with no imputation done if any of them is missing.
    - For ***safety evaluations*** (such as laboratory and vital signs), the last available assessment before first study treatment administration is taken as "baseline" value or "baseline" assessment. For patients who have not received any study treatment, the baseline will be the last assessment performed before randomization.
    - If patients have no value as defined above, the baseline result will be missing.

- Study day: For post-treatment events, study day is calculated as (event/assessment date – date of Day 1) + 1. For pre-treatment events, study day is calculated as event/assessment date – date of Day 1. There is no study day zero. Day 1 is the date of Cycle 1 Day 1 for balixafortide + eribulin arm, while it is the date of Cycle 1 Day 2 - 1 for eribulin arm:
    - Balixafortide + eribulin arm: Cycle 1 Day 1.
    - Eribulin arm: (Cycle 1 Day 2) – 1.
- Body Mass Index (BMI) (kg/m^2) is defined as [Weight(kg)] / $[Height(m)]^2$.
- Body Surface Area (BSA) (m^2) is defined as $\sqrt{Height(cm) * Weight(kg) / 3600}$.

4.1. Sample Size

The trial is powered for detecting superiority of balixafortide + eribulin versus eribulin monotherapy for the primary efficacy endpoint of PFS in the 3rd line+ population.

In the trial, patients will be randomized in a 1:1 ratio to the two treatment arms stratified by line of therapy (2nd line versus 3rd line+), hormone receptor status (positive versus negative), prior treatment with CDK 4/6 inhibitor (received a CDK 4/6 inhibitor versus not received a CDK 4/6 inhibitor) and visceral versus non-visceral disease.

The sample size calculations are event-driven. In the 3rd line+ population, approximately 320 patients are required in order that 286 PFS events occur within 24 months of randomizing the first patient. Patients who discontinue from the study will not be replaced.

The sample size calculations were based on the following assumptions:

- Overall 1-sided type I error rate: 0.020;
- In the 3rd line+ population, a median PFS on eribulin monotherapy of 3.9 months vs 5.8 months for balixafortide + eribulin to give a PFS hazard ratio 0.674;
- Power of 90%;
- Assuming a 12-month non-linear recruitment period (η=2; Carroll, 2009) and a minimum 12-month follow-up period after the last patient is randomized.
- Further, assuming a median OS on eribulin monotherapy of 14.4 months vs 17.4 months for balixafortide + eribulin to give an OS hazard ratio 0.828, a total of 226 deaths are expected in the 3rd line+ population in the final analysis of OS, planned to take place 12 months after the PFS analysis. Note, there is only one formal analysis of PFS to take place once 286 PFS events have accrued. No further formal analyses of PFS are planned.

4.2. Randomization, Stratification, and Blinding

This trial is an open-label, randomized, parallel, two-arm, multicenter, international Phase III trial. Patients will be randomized after the investigator has verified eligibility.

Patients will be randomized in a 1:1 ratio to one of the following treatment regimens:

- Balixafortide + eribulin

or

- Eribulin monotherapy

Patients will be stratified according to:

- Line of therapy (2nd line vs 3rd line+).
- HR receptor status (positive vs negative).
- Prior treatment with Cyclin-dependent kinase (CDK) 4/6 inhibitor (received a CDK 4/6 inhibitor vs not received a CDK 4/6 inhibitor).
- Visceral vs non-visceral disease.

For stratified analyses in the 3rd line+ population, the stratification factors (HR receptor status at randomization, prior CDK 4/6 inhibitor treatment and visceral disease) entered into the Interactive Response Technology (IRT) at the time of randomization will be used in analysis models. The randomization strata will be summarized by treatment arm for the Intention-to-treat (ITT) set, Line of therapy (only applicable in 2nd line population), HR receptor status, prior treatment with CDK 4/6 inhibitor (yes, no) and visceral disease (yes, no). Additionally, the exact line of therapy, as captured on the CRF, will also be summarized.

The analyses by line of therapy will be done on data collected in IRT system and imported as such in eCRF. The exact number of lines of therapies is collected in eCRF and will be used to summarize any discrepancy with the stratification factor.

The block size will be pre-specified and implemented in the IRT and will not be known to the Investigator. Polyphor Ltd. will determine the maximum number of patients that can be enrolled at each participating site and/or each geographic region.

4.3. Analysis Set

For the Food and Drug Administration (FDA) submission, the following analysis sets have been defined: Intent-to-Treat (3rd line+), Intent-to-Treat (3rd line+ confirmed), Per-Protocol (3rd line+ confirmed), Response Evaluable (3rd line+), Response Evaluable (3rd line+ confirmed), Patient Reported Outcome (3rd line+), Safety (3rd line+) in the 3rd line+ population. Note that the 3rd line+ population in these sets refers to those who were randomized as 3rd line+. 3rd line+ Confirmed refers to the subset of patients stratified as 3rd line+ who were confirmed as receiving the investigational medication as 3rd line of therapy or more in the eCRF data.

The efficacy endpoints analysis will be conducted in the ITT sets: Intent-to-Treat (3rd line+) for PFS and OS and Response Evaluable (3rd line+) for ORR. The other additional sub-sets defined in this section Intent-to-Treat (3rd line+ Confirmed), Per-Protocol (3rd line+ Confirmed) and Response Evaluable (3rd line+ Confirmed) will be used for sensitivity analysis.

In addition, the following analysis sets will be of interest for submissions to all regulators: Enrolled, Intent-to-Treat and Safety for the overall population.

4.3.1. Enrolled Set

Enrolled set is defined as all screened patients of the overall population who signed the informed consent.

4.3.2. Intent-to-Treat Set

The Intent-to-treat (ITT) set is defined as all randomized patients in the overall population. The co-primary endpoint PFS will be analyzed in this set if it is found to reach significance in the ITT (3rd line+) set as described in Section 8.3 and Figure 1.

4.3.3. Intent-to-Treat Set (3rd Line+)

The Intent-to-treat (ITT) set (3rd line+) is defined as all randomized patients in the 3rd line+ population. The 3rd line+ population is defined in the study as patients allocated to the 3rd line+ randomization strata. Efficacy data in this analysis set will be summarized by randomized treatment. The co-primary endpoint PFS, key secondary efficacy endpoint OS and exploratory efficacy endpoints will be analyzed by the treatment arm to which they are randomized using ITT (3rd line+).

4.3.4. Intent-to-Treat Set (3rd Line+ Confirmed)

The ITT set (3rd line+ confirmed) includes only those ITT (3rd line+) patients confirmed as receiving the investigational medication as 3rd line of therapy or more, as per eCRF data. Note that this analysis set therefore **excludes** patients randomized in 3rd line+ but receiving the investigational medication as only the 1st line or 2nd line of therapy as per eCRF.

A sensitivity analysis will be conducted on the co-primary endpoint PFS and the key secondary efficacy endpoint OS using the ITT set (3rd line+ Confirmed). Patients will be analyzed by the treatment arm to which they are randomized.

4.3.5. Per-Protocol Set (3rd Line+ Confirmed)

The Per-Protocol (PP) set (3rd line+ confirmed) is defined as a subset of the ITT set (3rd line+ confirmed), including only patients who took at least one dose of Eribulin and without any major protocol deviation that could affect the evaluability of the co-primary efficacy endpoints. Protocol deviations causing the exclusion of patient from PP set are listed in the Study Deviations Rules document.

A sensitivity analysis will be conducted on the PFS and the key secondary efficacy endpoint OS using the PP set (3rd line+ confirmed). Patients will be analyzed by the treatment arm to which they are randomized.

4.3.6. Response Evaluable Set (3rd Line+)

The Response Evaluable (RE) set (3rd line+) will be used to evaluate disease response. The RE set (3rd line+) includes all randomized patients in the 3rd line+ population with measurable disease by RECIST v1.1 as determined by the Investigator at screening during the evaluation of breast cancer history. The co-primary endpoint analysis of ORR and other secondary endpoint analyses of CBR, duration of response, DCR and time of response will be conducted using the RE (3rd line+) set. Patients will be analyzed by the treatment arm to which they are randomized.

4.3.7. Response Evaluable Set (3rd Line+ Confirmed)

The Response Evaluable (RE) set (3rd line+ confirmed) is a subset of RE set (3rd line+) including only those patients confirmed as receiving the investigational medication as 3rd line of therapy or more, as per eCRF data.

A sensitivity analysis will be conducted on the co-primary endpoint ORR using the RE set (3rd line+ confirmed)

4.3.8. Patient Reported Outcome Set (3rd Line+)

The Patient Reported Outcome (PRO) set (3rd line+) is defined as all randomized patients in the 3rd line+ population where the patient has at least one post-baseline PRO value (e.g. EORTC

questionnaire, EQ-5D or FACT-B). Patients will be analyzed by the treatment arm to which they are randomized.

4.3.9. Safety Set

The Safety set consists of all randomized patients in the overall population who receive at least one dose (or partial dose) of study treatment (i.e. a subset of ITT).Patients will be analyzed by the treatment received, based on the first dose of study drug, to summarize the safety data overall. They will be included in the balixafortide + eribulin arm if they received at least one dose of balixafortide during the study.

4.3.10. Safety Set (3rd Line+)

The Safety set (3rd line+) consists of all patients from the ITT set (3rd line+) who receive at least one dose (or partial dose) of study treatment. Patients will be analyzed by the treatment received, based on the first dose of study drug. They will be included in the balixafortide + eribulin arm if they received at least one dose of balixafortide during the study.

4.4. Handling of Missing or Partial Date Data

Missing date data will be handled as follows:

- For prior cancer therapies and medical history of cancer, the study day corresponding to the start and stop date will be imputed prior to calculating duration of the therapy/history using the following rules:
    - For partially missing start dates, missing day of the month will be imputed as the first of the month and missing month will be imputed as January. No imputation will be done if the year is missing.
    - For partially missing stop dates, missing day of the month will be imputed as the last day of the month and missing month will be imputed as December and at least 1 day after the start date of the therapy/history. Any imputed date occurring after the screening visit date will be imputed to the screening visit date. No imputation will be done if the year is missing.
- For prior radiotherapy and surgery, the relevant time to the initiation of the trial, etc. will be calculated using the recorded date of radiotherapy or date of procedure for surgery. The imputation rules described for prior cancer therapies will be applied to prior radiotherapy and surgery.
- For time since initial diagnosis to informed consent and time since diagnosis of locally recurrent or metastatic disease to informed consent, partial dates for initial diagnosis and diagnosis of metastatic disease, will be imputed as follows:
    - No imputation will be done if the year is missing.
    - If the year is before informed consent date, then missing days will be imputed as the first day of the month and missing months will be imputed as July.
    - If the year is the current year of informed consent date, then missing days will be imputed as the first day of the month and missing months will be imputed as January.
- For time since initial diagnosis to randomization, partial dates will be imputed as follows:

- No imputation will be done if the year is missing.
- If the year is before randomization date, then missing days will be imputed as the first day of the month and missing months will be imputed as July.
- If the year is the current year of randomization date, then missing days will be imputed as the first day of the month and missing months will be imputed as January.

- For Prior or Concomitant Medication: Partial start or stop dates (where UN and UNK indicate unknown day and month respectively) will be imputed as follows:
  - Start dates:
    - UN-MMM-YYYY: Impute missing day as 01, so that CMs are assigned to prior and concomitant if there's any doubt.
    - DD-UNK-YYYY/UN-UNK-YYYY: If the year is before the year of first dose of study treatment, assume 01-JUL-YYYY of the collected year. If the year is after the year of first dose of study treatment, assume 01-JAN-YYYY of the collected year. If the year is the same as the year of first dose of study treatment and the stop date (after any imputation) is missing or is on or after the first dose of study treatment, then assume 01-JAN-YYYY, so that CMs are assigned to prior and concomitant. If the year is the same as the year of first dose of study treatment and the stop date (after any imputation) is prior to the first dose of study treatment, then assume the stop date for the start date.
  - Stop dates:
    - UN-MMM-YYYY: Assume the last day of the month.
    - DD-UNK-YYYY/UN-UNK-YYYY: Assume 31-DEC-YYYY.
    - In case of the death/study termination of the patients and the imputed stop date is after the date of death/study termination, the stop date will be imputed as the date of death/date of study termination.
    - Completely missing dates will not be imputed.
- For Post Anti-Cancer Therapies: Rules are as per concomitant medications above except when the imputed start date is earlier than the date of last dosing date, then the start date will be imputed as the date of last dosing date of study treatment + 1.
- For Adverse Events (missing month is not expected; missing day is allowed): Partial start or stop dates will be imputed as follows:
  - Partial start dates (where UN indicates unknown day):
    - UN-MMM-YYYY: If the month and year are different from the month and year of the first dose of study treatment, or the year is the same as the year of the first dose of study treatment, but the month is different, assume 01-MMM-YYYY. If the month and year are the same as the first dose of study treatment month and year and the end date (after any imputation) is missing or is on or after the first dose of study treatment, then assume the date of the first dose of study treatment. If the month and year are the same as the first dose of study treatment month and year and the end date (after any imputation) is prior to the first dose of study treatment, then assume 01-MMM-YYYY.

- Partial end dates (where UN indicates unknown day):
  - UN-MMM-YYYY: Assume the last day of the month.
  - In case of the death/study termination of the patients and the imputed end date is after the date of death/study termination, the end date will be imputed as the date of death/date of study termination.

- For OS with partial death date:
  - Every effort will be made to collect accurate death date. In case that the day is missing, the later date of the following will be used to calculate time to death:
    - first day of the month 01-MMM-YYYY, or
    - last visit date + 1, or
    - last known alive date + 1.
  - If day and month are both missing, the later date of 1st January of that year 01-JAN-YYYY, last visit date + 1, or last known alive date + 1 will be assumed when calculating time to death.
  - If death date is completely missing then the later of last visit date + 1, or last known alive date + 1 will be assumed when calculating time to death.

Handling of missing data for analysis of PFS and QoL are described in the statistical analysis sections 8.1.1, 8.5.1 and 8.5.2.

4.5. Mapping of Unscheduled visits

All planned assessments will be reported without any remapping.

Missing scheduled assessments will not be replaced by unscheduled assessments.

4.6. COVID-19 Impact

Coronavirus (COVID-19) pandemic of 2019-20 poses risks to the safety of subjects enrolled in clinical trials, and the availability and interpretability of data from those trials. The FDA suggests sensitivity analyses to examine differences in baseline characteristics and post-baseline events between the originally enrolled participants and any additional participants.

The European Medicines Agency suggests to conduct risk assessment, primarily using aggregate blinded data, of the impact of COVID-19 on trial integrity and interpretability. This should consider implications on recruitment, loss of patients during the trial, ability to record data and ability to interpret the treatment effect (considering impacts on both trial participants and clinical trial conduct).

Any analysis at this stage would be premature considering data collection regarding COVID-19 is ongoing. A separate meeting will be held between Polyphor and PPD to evaluate the amount of data affected by COVID-19.

Potential follow-up considerations based on the risk assessment may include the following:

- Validation of outcomes that were measured differently before and after start of the pandemic;
- The need to adjust the trial sample size;

- Handling sources of bias such as missing values and newly identified intercurrent events (e.g. deaths or serious illness, missed or delayed doses due to missed visits, Discontinuation of study treatment due to logistical issues, Extension of treatment duration due to delayed end-of-study visit, Use of additional prohibited medications);
- Recommendations from a trial participant's safety perspective on how to stop, pause or re-start the trial.

In case the study conduct of a significant amount of patients will be affected by COVID-19, a sensitivity analysis will be proposed by PPD and agreed with Polyphor.

5. Patient Disposition

5.1. Disposition

A summary of patient disposition will display the number of patients who were randomized and who comprised each analysis set by treatment arm for the Enrolled set. In addition, the number of patients with status of completed, ongoing and discontinued study; reason for discontinuation from study will be summarized for the ITT (3rd line+) set. Also, the number of patients who discontinued from study treatment, overall and by reason; and the number of deaths and reason will be presented. The number of patients recruited by geographical region [USA, Europe, Asia and Latin America] will also be summarized for the ITT (3rd line+) set. Disposition data, including analysis sets to/from which patients are included/excluded, will be presented in a data listing.

The three randomization stratification factors relevant to the 3rd line+ population (HR receptor status, CDK 4/6 inhibitor treatment, and visceral disease) will be listed and summarized as recorded in the IRT system by treatment arm as randomized.

5.2. Protocol Deviations

The definition of protocol deviations will be made by the joint Polyphor/PPD trial team (clinical, medical and statisticians). The identification of protocol deviations will be conducted by clinical team.

All significant protocol deviations will be summarized by treatment arm, site and overall, importance (major/minor) and activity subtype for the ITT (3rd line+) set.

All protocol deviations will be listed.

5.3. Missed Visits due to COVID-19

Information about missed visits due to COVID-19 will be recorded on a dedicated page in eCRF. All information collected on this page will be presented in listing for ITT (3rd line+) set.

Treatment discontinuation and study termination due to COVID-19 will be reported in listing of patient disposition.

6. Demographics and Baseline Characteristics

The following baseline data will be summarized and listed in separate tables for the ITT (3rd line+) set: Age (years), age group (<65yr, >=65yr), sex, race, ethnicity, fertility status, birth control method, height (cm), weight (kg), BSA (m^2), BMI (kg/m^2), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), line of therapy(as collected in eCRF), ECOG performance status, smoking history (yes, no) and tobacco usage (current, previous) for smokers. All demographic data will be presented in a listing.

6.1. Medical History

6.1.1. General Medical History

Medical conditions collected at screening will be mapped by the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) and will be summarized by system organ class (SOC) and preferred term (PT) by treatment arm for the corresponding ITT (3rd line+) set. For summary tables, a patient will be counted only once per SOC and PT. All data will also be presented in a listing.

General Medical History will include significant medical conditions including other cancer history, prior surgeries and concurrent illnesses.

6.1.2. Breast Cancer History

Medical history of breast cancer will be summarized for the corresponding ITT (3rd line+) set. Breast Cancer History will include time since initial diagnosis of breast cancer (years), time since diagnosis of locally recurrent or metastatic disease (years), stage at initial diagnosis, stage at study entry, date of last tumor progression before study entry, histologic subtype, ER status, PgR status, HER2 status (FISH, CISH, IHC, other), site of disease, tumor marker status and Triple-Negative Breast Cancer (TNBC) subtype. For tissue profiling, if multiple assessments of breast cancer history data exist, only the last tumor assessment before randomization will be summarized. Time since initial diagnosis and since metastatic disease will be derived as follows:

- Time since initial diagnosis (years): (Date informed consent signed – Date of initial diagnosis) / 365.25.
- Time since diagnosis of locally recurrent or metastatic disease (years): (Date of informed consent date – Date of diagnosis of locally recurrent or metastatic disease) / 365.25.

See Section 4.4 for data handling rules for partial initial diagnosis and diagnosis of locally recurrent / metastatic disease dates.

All breast cancer data will also be presented in a listing.

6.1.3. Prior Anti-Cancer Therapy

The number and percent of patients receiving prior anthracycline and/or a taxane will be summarized. The number of prior hormonal therapies, immunotherapies, targeted therapies, cytotoxic chemotherapy agents for breast cancer, neoadjuvant agents, adjuvant agents, cytotoxic agents for locally recurrent and metastatic disease, will be presented based on prior cancer therapies reported on eCRFs and summarized for the corresponding ITT (3rd line+) set.

Duration of prior anthracycline and/or taxane use will be summarized descriptively.

Prior systemic cancer/oncology therapies will be tabulated for the corresponding ITT (3rd line+) set using the World Health Organization Drug Dictionary Enhanced (WHO-DDE) and Anatomical, Therapeutic, or Chemical (ATC) level-4 classifications and preferred term. If the ATC level-4 classification is missing, the next non-missing lower level of classification will be used (level 3). If a patient reports the same medication multiple times, then the frequency of that medication will be incremented by only one in the applicable arm. If a patient reports multiple medications within the same ATC level-4 classification, then the frequency of that ATC level-4 classification will be incremented by only one in the applicable arm. Percentages will be calculated using the number of patients in the corresponding ITT (3rd line+) set. The summary

table will be ordered by descending incidence of ATC class and by descending incidence of preferred drug name within each ATC class for all patients.

Data listings for prior systemic cancer/oncology therapies will include dose, unit, dose regimen frequency, best response to cancer therapy, and reason for discontinuation (toxicity, other or progressive disease). The study day corresponding to the starting and stopping date of the regimen will be calculated for display in the data listing.

The duration of last chemotherapy line before entering in the study will be summarized by treatment arm. The duration (months) is calculated by (Cancer Therapy End Date~~e~~ Cancer Therapy Start Date + 1) / 30.4375. Any missing or partial date will be imputed according to section 4.4.

6.1.4. Previous Radiotherapy and Surgical History

Previous radiotherapy and surgical history related to breast cancer will be listed for ITT (3rd line+) set.

Surgical history related to breast cancer and percentage of patients receiving prior radiotherapy will be summarized for the ITT (3rd line+) set.

6.2. Inclusion and Exclusion Criteria

Failed patient's' inclusion criteria and exclusion criteria will be listed for ITT (3rd line+) set.

7. Treatments and Medications

7.1. Prior and Concomitant Medications

Prior and concomitant medications will be recorded on the eCRF and coded to as per ATC and preferred drug name using the prevailing version of the WHO-DDE dictionary. The summary tables will be ordered by descending incidence of ATC class and by descending incidence of preferred drug name within each ATC class for all patients.

Concomitant medications are defined as medications with a start date on or after the date of the first dose in Cycle 1 up to 30 days following the last dose of study treatment in the study; it will also include medications initiated prior to the date of the first dose of study treatment and continued during treatment.

Concomitant medications will be summarized by treatment for the Safety set by WHO-DDE ATC classification and preferred term. If the ATC level-2 classification is missing, the next non-missing lower level of classification will be used (level 1). If a patient reports the same medication multiple times, then that medication will be counted only once. If a patient reports multiple medications within the same ATC level-2 classification, then the frequency for that ATC level-2 classification will be incremented by only one. Percentages will be calculated using the total number of patients in the Safety set.

Prior medications are defined as medications taken from 28 days prior to the first dose of study treatment in Cycle 1. Prior medications will be summarized similarly to concomitant medications for the ITT (3rd line+) set. Percentages will be calculated using the total number of patients in the ITT (3rd line+) set.

Prior and concomitant medications will be presented in a data listing.

7.2. Post Anti-Cancer Therapy

Post anti-cancer therapies (therapies applied after study treatment discontinuation for subjects that are continuing in the study for the FU phase) as reported in Post Anti-Cancer Therapy eCRF page will be summarized for the corresponding ITT (3rd line+) set, using the WHO-DDE and ATC level-4 classifications and preferred term. The number and percent of patients who took at least one post anti-cancer therapy will be calculated and presented by treatment arm, ATC level-4 classifications, and preferred term.

7.3. Study Treatments

Information about study treatments balixafortide and eribulin are provided in Section 3.3.

In the eCRF weight of eribulin is collected in both base and salt, this should be converted to salt by dividing by a factor of 0.88366.

Conversion factor is derived per weight as base and weight as salt:

- Weight as base / Weight as salt = 826.0 / 729.9 = 0.88366.

7.3.1. Extent of Exposure

Overall exposure to study treatment will be summarized for the Safety set.

The following exposure variables will be derived:

- Study treatment duration (days): date of last dose – date of first dose + 1.
- Number of cycles: Total number of complete or partial treatment cycles the patient received.
- Total cumulative dose (mg): The sum of actual doses (mg) received across all cycles.
- Average Duration of infusion: The mean duration of infusion across all cycles for each patient. The duration of infusion will be calculated as; completion time of infusion ~~n~~ start time of infusion; where infusions have been paused, the duration of the infusion will be computed as the sum of the start/stop and interrupted/restarted times as captured in the eCRF.

Total cumulative dose and average duration of infusion will be derived per drug.

7.3.2. Treatment Compliance and Modifications

The number of patients with dose reduction, dose interruption, missed dose, and dose delay will be summarized along with their reason. A dose interruption is defined as a dose stopped at least one time while being administered. A dose is missed when the administration at site does not occur. A dose delay occurs when the patient receives the treatment not as per scheduled date. All summaries will be performed for each cycle and for all cycles combined.

The following parameters will be calculated for balixafortide:

- Dose intensity (mg/kg/cycle): [Sum of actual doses received (mg/kg) across all cycles]/ (number of cycles received).
- Expected dose intensity (mg/kg/cycle): initial planned dose (5.5 mg/kg) [i.e. dose for each injection] × 6.

- Relative dose intensity (%): ((Dose intensity) / (expected dose intensity)) × 100.
- Average dose per infusion (mg/kg/infusion): [Sum of actual dose received (mg/kg) across all cycles] / (number of infusions received).

The following parameters will be calculated for eribulin:

- Dose intensity (mg/m^2/cycle): [Sum of actual doses received (mg/m^2 salt) across all cycles] / (number of cycles received).
- Expected dose intensity (mg/m^2/cycle) = initial planned dose (mg/m^2)[i.e. dose for each injection] × 2
- Relative dose intensity (%): ((Dose intensity) / (expected dose intensity)) × 100.
- Average dose per infusion (mg/m^2/infusion): [Sum of actual doses received (mg/m^2) across all cycles] / (number of infusions received).

8. Efficacy Analysis

Analysis of efficacy endpoints will be performed using the ITT set (3rd line+) unless otherwise noted.

As described in Section 3.1, a retrospective review of radiographic images and relevant clinical information (surgical history) collected on-study to verify the protocol defined endpoints of disease response and progression, as assessed by the Investigator will be performed. The IRC will be blind to treatment and will be able to provide consistent assessment over all the sites. Based on these blinded, central review of CT/MRI scans performed every 6 weeks (±7 days) during the first year, and then performed every 12 weeks (±7 days), until PD is documented by RECIST v1.1; the patients will be evaluated for Complete Response (CR), Partial Response (PR), Stable Disease (SD), Progressive Disease (PD) or Not Evaluable (NE).

8.1. Co-Primary Efficacy Endpoint

The sequence of testing efficacy and alpha recycling is explained in Section 8.3 and Figure 1.

8.1.1. Progression-Free Survival

PFS, as assessed by the IRC, is a co-primary efficacy endpoint in the 3rd line+ population. The RECIST v1.1 guidelines will be used to determine disease progression. The alpha allocation will be 0.040 2-sided. From the date of randomization, patients will be evaluated for PD based on blinded, central review of CT/MRI scans as described in Section 8. Patients who discontinue study treatment prior to PD will undergo repeat CT/MRI scans and tumor response assessments every 8 weeks (±7 days) for the first year post randomization and 12 weeks (±7 days) after the first year post randomization, regardless of regimen, until confirmed PD is documented, death occurs, patient is lost to follow-up, or the patient withdraws consent (whichever occurs first).

Date of Progression will be the date assessed by the IRC.

PFS is defined as the time from the date of randomization to the earliest evidence of documented PD or death from any cause. Patients who are alive without documented PD or are either lost to follow-up or withdrew consent, will be censored at the date of last tumor assessment without disease progression.

The duration of PFS will be calculated as follows:

- For patients who have PD or died, Duration of PFS = Date of death/PD – Date of randomization + 1 day;
- For patients who are alive without documented PD, Duration of PFS = Date of last tumor assessment (where response is not missing or unable to evaluate) – Date of randomization + 1 day.

Table 1 below specifies the date of disease progression or censoring rules that will be carried out for the PFS analysis on the ITT (3rd line+) set:

Table 1 - Date of Progression or Censoring for Progression-free Survival

Situation	Date of Disease Progression or Censoring	Outcome
No post-baseline assessments for tumor response and no death within first two tumor assessments	Date of randomization	Censored
Disease progression or death after two or more consecutive missed tumor response assessments	Date of last scan for tumor assessment showing no evidence of disease progression that is before the first missed visit	Censored
Not known to have progressed or died according to data in the database as of data-cut-off	Date of last scan for tumor assessment showing no evidence of disease progression	Censored
Disease progression reported on multiple response assessments	Earliest Date of Progression	Progressed
Disease progression or death after one missed tumor response assessment	Earliest Date of Progression	Progressed
Death without PD	Date of death	Progressed
Starting anti-cancer therapy prior to observing PD or in absence of PD	Date of last scan for tumor assessment showing no evidence of disease progression prior to start date of the anti-cancer therapy	Censored

8.1.2. Objective Response Rate

Objective response rate (ORR), defined as the proportion of patients with a confirmed complete response (CR) or a confirmed partial response (PR) according to RECIST v1.1 based upon the best response assessed, is a co-primary efficacy endpoint. The alpha allocation will be 0.001 2-sided. As described in section 8, patients will be evaluated for tumor response according to RECIST v1.1 for confirmed CR, confirmed PR, SD, or PD based on blinded, central review of CT/MRI scans. Response will be based on the IRC assessment with a supportive analysis based on the local Investigator's review. Following initial documentation of an objective response, a confirmatory CT/MRI scan will be performed at least 4 weeks later. Results based upon scans performed after start of new anti-cancer therapy will be excluded for analysis derivations.

For the secondary Efficacy Endpoints, the confirmed responses (best responses) determined from Overall responses based on the local investigator's review are not entered in the eCRF and will be derived from the algorithm described in Appendix 14.5.

These data from the CT/MRI scans will be presented in listings; listings for tumor assessment (target, non-target lesions, new lesions, sum of diameters, assessment of disease based on MRI and bone scan). Furthermore, the tumor assessments by the IRC will also be listed.

8.1.3. Primary Analysis

Progression-Free Survival

PFS, as assessed by the IRC, will be analyzed via a Cox proportional hazards model on the ITT (3rd line+) set and if found to be significant also in the ITT set as described in section 14.2.

Analysis of the ITT (3rd line+ confirmed) and PP (3rd line+ confirmed) sets is added as sensitivity and is not part of the alpha-control for this study. The analyses will be stratified for randomization stratification factors (excluding lines of therapy in the 3rd line+ sets) and a fixed effect term will be included for treatment arm.

The hazard ratio (balixafortide + eribulin vs eribulin monotherapy) will be estimated from the model along with the associated 96% CI and 2-sided p-value.

If PFS in the 3rd line+ population is met at the allocated alpha level, PFS in the overall population will also be formally tested at the same alpha level. (i.e. 0.040 2-sided).

The summary of reasons for PFS censoring will display the following categories (based on the end of treatment eCRF page, the study termination page, the survival follow-up eCRF page and relevant derivations of tumor assessments):

- Ongoing without event
- Lost to follow-up
- Withdrew consent
- Adequate assessment no longer available (No post-baseline assessments)
- New anti-cancer therapy added
- Event documented after two or more missing tumor assessments

As the Cox proportional hazards model assumes odds are proportional, a supportive analysis for PFS, as assessed by the IRC, will be conducted for the ITT (3rd line+) set using a stratified log-rank test which does not hold this same assumption; strata will be the randomization stratification factors for 3rd line+ population excluding line of therapy (see Section 4.2 for further details). The 2-sided p-value will be extracted from the stratified log-rank test and presented alongside the results of the Cox proportional hazards model analysis.

The number of patients experiencing events and the number of censored patients will be summarized by randomized treatment arm with reasons for an event/censoring. The median survival times will be estimated by the Kaplan-Meier method and presented along with their corresponding 95% CI. The 25th percentile and 75th percentile for the survival times along with the corresponding 95% CI for the percentiles will also be displayed, CI will be calculated using method of Brookmeyer and Crowley (1982). Additionally, PFS probabilities at fixed time points i.e. 3, 6, 9, 12, 15 months will be derived from the Kaplan-Meier estimates with 95% confidence intervals, CI will be calculated using the Greenwood formula. The time points will be displayed based on the data availability.

Corresponding plots of Kaplan-Meier curves by treatment arm will also be presented. The plots will display the number of patients at risk every 3 months. Alongside this the PFS data will be presented in a listing by treatment arm.

Objective Response Rate

The ORR, defined from the best response as assessed by the IRC, will be analyzed by an exact logistic regression, stratified for randomization stratification factors excluding line of therapy (see Section 4.2 for further details). The odds ratio will be estimated from the model along with the associated 99.9% CI (0.001 alpha) and 2-sided p-value. Exact Clopper-Pearson 2-sided 95% confidence limits will be calculated for the proportion of patients with objective response in each arm.

ORR analyses will be performed for the RE (3rd line+) set and RE (3rd line+ Confirmed) set. Analysis of the RE (3rd line+ Confirmed) set is added as sensitivity and is not part of the alpha-control for this study. In addition to presenting ORR, the best response as assessed by the IRC using response categories CR, PR, SD, PD and NE will be summarized. The number and percentage of patients in each response category will be presented by treatment arm. Patients who do not have any post-baseline tumor assessment will be counted under the category NE. SD requires at least one post-baseline scan performed at minimum of six (6) weeks (42 days) +/- seven (7) days, for a minimum time on-study of thirty-five (35) days or later. Patients who did not have any post baseline scans for the 6-week assessment or later and also do not have a documented PD will be counted as NE.

8.1.3.1. Sensitivity Analysis for PFS

A sensitivity analysis will be conducted for PFS, as assessed by the IRC on the ITT (3rd line+) set. The calculation of PFS is the same as described in section 8.1.1 where patients will not be censored at the start of a new anti-cancer therapy prior to observing documented PD. The same statistical analysis methods: Kaplan-Meier analysis, Cox proportional hazards model and supportive analysis using a stratified log-rank test will be used to estimate the effect of treatment irrespective of starting a new anti-cancer therapy.

8.2. Key Secondary Efficacy Endpoint Analysis

OS is the key secondary endpoint. The alpha allocation will be between 0.001 and 0.002 2-sided at the interim and between 0.008 and 0.050 at the final analysis depending on the results of previous analyses PFS and ORR analyses as illustrated in the flow diagram of section 14.2.

OS defined as the time from the date of randomization to death from any cause will be assessed in the ITT (3rd line+) set. The ITT (3rd line+ confirmed) and PP (3rd line+ confirmed) sets will also be analyzed as sensitivity. Patients will be followed until their date of death, lost to follow-up, withdrawal of consent, or end of study.

Patients who are lost to follow-up or are not known to have died at the time of data-cut-off for analysis will be censored at later of last visit date, or last known alive date.

Patients who do not have any follow up since randomization will be censored at the date of randomization.

The duration of overall survival will be calculated for all randomized patients as follows:

- For patients who have died: Duration of overall survival = Date of death – Date of randomization + 1 day;
- For patients who are lost-to-follow-up or not known to have died: Duration of overall survival = Later of Date of last visit date, last date known to be alive – Date of randomization + 1 day;

See section 4.4 for details on how to handle partial death date.

OS will be analyzed using the same Cox proportional hazards model as PFS. The CI coverage for OS at the planned interim and at the final analysis will be determined by the allocated alpha level.

A supportive analysis for OS will be conducted using a stratified log rank test for the ITT (3rd line+) set again as per the analysis of PFS; strata will be the randomization stratification factors as described in Section 8.1.3 for PFS.

The OS proportion at 6, 12 and 18 months, including 95% CIs, will also be presented.

8.3. Timing of Analyses of Primary and Key Secondary Efficacy Endpoints and Type I Error control

In the 3rd line+ population co-primary efficacy endpoints are ORR and PFS; the key secondary efficacy endpoint is OS. A formal analysis of ORR is planned once all 3rd line+ patients have been randomized and followed for a minimum of 6 months. The formal analysis of PFS will take place once 286 PFS events have accrued, being expected when all randomized patients in the 3rd line+ strata have been followed for a minimum of 12 months. The alpha allocation for the PFS analysis will be 0.040 2-sided; if the PFS in the 3rd line+ population meets the allocated alpha level, PFS in the overall population will also be formally tested at the same alpha level; further, at the time of the PFS analysis an interim analysis of OS will be performed with an alpha allocation of 0.001 to 0.002 2-sided, depending on the result of the previous ORR analysis. The final analysis of OS will take place once all randomized patients have been followed for a minimum of 24 months at which time 226 events are expected. The alpha allocation for this analysis will be between 0.008 and 0.05 2-sided depending on the results of previous analyses. Thus, there will be 4 formal efficacy analyses in the 3rd line+ population (ORR, PFS, interim OS and final OS). If PFS is met in the 3rd line+ population, PFS analysis will be performed in the overall population. Consequently, there will be a total of 5 formal efficacy analyses. The overall Type I error rate will be controlled at 0.05 across these formal analyses by means of alpha allocation and recycling as described in Figure 1 in Appendix 14.2.

8.4. Other Secondary Endpoints

Other secondary endpoints are PFS according to RECIST v1.1 guidelines, as assessed by the local Investigator's review, and ORR (confirmed CR + confirmed PR) according to RECIST v1.1 guidelines, as assessed by the local Investigator's review. Additionally CBR, DCR, time to response and duration of response will be analyzed. These analyses will be supportive in nature and, hence, will not be subject to Type I error control. Statistical tests will be 2-sided with a nominal Type I error rate of 0.05.

PFS and ORR as assessed by investigator, will be analyzed as described in Section 8.1.3 for the ITT (3rd line+) set. Date of Progression will be the date assessed by the investigator.

CBR and DCR will be analyzed in ITT (3rd line+) set and RE (3rd line+) set. Time to response and duration of response will be analyzed only in RE (3rd line+) set. These analyses will be performed for parameters defined from the best responses as assessed by (i) the IRC and (ii) the local Investigator's review.

8.4.1. PFS Assessed by Local Investigator's Review

PFS, as assessed by local Investigator's review, will be analyzed as described in section 8.1.3. The date of assessment used will be the date recorded in the eCRF page "Overall Response according to RECIST v1.1 assessment".

8.4.2.ORR Assessed by Local Investigator's Review

ORR, as assessed by local investigator's review, will be analyzed as described in section 8.1.3. It will be defined as the proportion of patients with a confirmed CR or a confirmed PR according to RECIST v1.1 based upon the best response assessed, as assess by local investigator's review. The confirmed responses (best responses) are not entered in the eCRF, the algorithm for the best response determination according to RECIST v1.1 guidelines, as assessed by the local Investigator's review, can be found in Appendix 14.5.

The Date of First Response is defined as the time point response (TPR) date when the criteria for CR or PR were first met, whichever status is recorded first, for subjects whose confirmed response (Best Response) is either a CR or a PR. The Date of First Response may or may not be the same as the Date of Best Response.

8.4.3. Clinical Benefit Rate

Clinical benefit rate (CBR) is defined as the proportion of patients with a confirmed CR, confirmed PR, or SD for at least 6 months (≥ 182 days).

The CBR, as assessed according to RECIST v1.1 guidelines by the IRC, will be analyzed for RE (3rd line+) and ITT (3rd line+) analysis sets using an exact logistic regression as described for ORR in Section 8.1.3.

The CBR, as assessed by local investigator's review, will also be analyzed will be analyzed for RE (3rd line+) and ITT (3rd line+) analysis sets using an exact logistic regression as described for ORR in Section 8.1.3.

8.4.4. Disease Control Rate

Disease control rate (DCR) is defined as the proportion of patients with a confirmed CR, confirmed PR, or SD of any duration. Patients with SD of any duration as the best response will be considered for this analysis and not only patients with SD of confirmed best response.

The DCR, as assessed by IRC, will be analyzed for RE (3rd line+) and ITT (3rd line+) analysis sets using an exact logistic regression as described for ORR in Section 8.1.3

The DCR, as assessed by local investigator's review, will also be analyzed will be analyzed for RE (3rd line+) and ITT (3rd line+) analysis sets using an exact logistic regression as described for ORR in Section 8.1.3.

8.4.5. Time to Response

Time to response is defined as the time from randomization to the first confirmed CR or PR as assessed by IRC and Local Investigator's Review. Time to response will be analyzed for the RE (3rd line+) set via Cox proportional hazards modelling with a fixed effect term for treatment arm and stratified by stratification factors. The hazard ratio (balixafortide + eribulin vs eribulin monotherapy) will be estimated from the model along with the associated 95% CI and 2-sided p-value. The hazard ratio is interpreted, in the context of response, as the odds a patient will respond faster with the combination therapy as compared to eribulin monotherapy so a value > 1 shows the addition of balixafortide to be beneficial.

The data will also be displayed using Kaplan-Meier curves; median time to response and the 95% CIs will be estimated as described for PFS in Section 8.1.3.

Time to response, as assessed by local investigator's review will also be analyzed for the RE (3rd line+) set as described for PFS in Section 8.1.3.

No censoring rule will apply since Time to Response will only be analyzed for patients achieving a confirmed response CR or PR.

This data will be presented in a listing for both IRC and local investigator's review.

8.4.6. Duration of Response

Duration of response is defined as the time from first confirmed CR or PR until the earliest evidence of disease progression or death from any cause, as assessed by IRC review. The algorithm for calculation of Date of Progression and censoring rules will be the same as defined in the analysis for PFS, as described for PFS in Section 8.1.3.

Duration of response will be analyzed for the RE (3rd line+) set via Cox proportional hazards model with a fixed effect term for treatment arm and stratified by stratification factors. The hazard ratio (balixafortide + eribulin vs eribulin monotherapy) will be estimated from the model along with the associated 95% CI and 2-sided p-value. The data will also be displayed using Kaplan-Meier curves; median duration of response and the 95% CIs will be estimated as described for PFS in Section 8.1.3.

Since duration of response is assessed in responding patients only, the comparison is biased and could be misleading if the ORR differs moderately between the treatment arms. Hence, a supportive analysis will be performed based on the expected duration of response as per Ellis et al (2008) as this analysis includes all randomized patients, including those who did not respond, and so provides a valid comparison of the treatment arms. The expected duration of response is a product of the estimated fraction of patients with a response (ORR) and the mean duration of response in responding patients.

Duration of response, as assessed by local investigator's review, will also be analyzed for the RE (3rd line+) set as described for PFS in Section 8.1.3.

This data will be presented in a listing for both IRC and local investigator's review.

8.5. Exploratory Efficacy Endpoints

Analysis of the other exploratory efficacy endpoints will not be subject to Type I error control. Statistical tests will be 2-sided with a nominal type I error rate of 0.05. Exploratory subgroup efficacy analyses of PFS, ORR and OS will be performed to qualitatively assess if the overall treatment effect is broadly consistent between clinically important subgroups.

The following subgroups will be analyzed for the ITT (3rd line+) set and for the RE (3rd line+) set as described below:

- Age category (<65yr vs ≥ 65yr)
- Race (White, Black, Asian)
- Visceral disease at baseline (Yes versus No)
- Number of sites of disease involved (≤2 versus > 2)
- ECOG status at baseline (0 versus 1 versus 2)
- Prior treatment with cyclin-dependent kinase (CDK) 4/6 inhibitor (received a CDK 4/6 inhibitor vs not received a CDK 4/6 inhibitor) only for HR+ patients

- Prior capecitabine (Yes versus No)
- Prior immune therapy (Yes versus No) only for TNBC patients
- Number of prior lines of therapy (≤2, 3, ≥4)
- Receptor status [TNBC versus HR+]
- Estrogen receptor status (Positive versus Negative)
- Progesterone receptor status (Positive versus Negative)
- Metastasis in liver (Yes versus No)
- Metastasis in lung (Yes versus No)
- Time from initial diagnosis to randomization (<2 years versus ≥2 years)
- Region (USA, Europe, Asia and Latin America)

The following subgroups will be analyzed for the ITT (3rd line+ Confirmed) set and for the RE (3rd line+ Confirmed) set as described below:

- Receptor status [TNBC versus HR+]
- Prior treatment with cyclin-dependent kinase (CDK) 4/6 inhibitor (received a CDK 4/6 inhibitor vs not received a CDK 4/6 inhibitor) only for HR+ patients
- Visceral versus non-visceral disease

Subgroups analyses for PFS, as assessed by IRC, and OS will be achieved by use of a Cox proportional hazards model with fixed effect terms for treatment arm, the subgroup as a class variable and treatment arm by subgroup interaction. The p-value for the interaction will be used as guide to the statistical plausibility of a true difference in the treatment effect between levels of the subgroup; an interaction p-value <0.05 will be suggestive of a true difference in the treatment effect between levels of the subgroup. The hazard ratio (balixafortide + eribulin vs eribulin monotherapy) together with its 95% CI and 2-sided p-value will be presented for each level of the subgroup. The hazard ratios and associated 95% CI will be displayed in a forest plot along with the interaction p-value. Subgroup analyses above may not be performed if the sample size is too small to provide accurate estimate.

Subgroup analyses for ORR will be achieved via an exact logistic regression, including an interaction term for treatment arm by subgroup interaction. The odds ratio will be estimated from the model along with the associated 95% CI and 2-sided p-value, additionally the p-value for the interaction term will be presented on the summary table. Exact Clopper-Pearson 2-sided 95% confidence limits will be calculated for the proportion of patients with objective response in each arm.

Additional supplementary Bayesian subgroup analyses may be performed by the sponsor. To guard against random high and random low results associated with the analysis of multiple subgroups for which the trial was not powered and some of which may be small, these Bayesian subgroups analyses will be performed using R based on methodology as per Dixon and Simon (1991).

8.5.1. EORTC-QLQ-C30 QoL

The European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 (version 3.0) is a validated instrument to measure quality of life and assess symptoms and side effects of treatment and their impact on everyday life. Analysis of the EORTC-QLQ-C30 QoL data will be performed in the PRO (3rd line+) set.

The instrument is composed of 5 multi-item functional scales (physical, role, social, emotional and cognitive functioning), a global health status/QoL scale, 3 symptom scales (fatigue, nausea/vomiting and pain), and 6 single items (financial difficulties, appetite loss, diarrhea, constipation, insomnia and dyspnea). Most items are scaled 1 to 4 except the items contributing to the global health status/QoL, which are 7-point questions. Raw scores will be transformed using a linear transformation to standardize the results so that scores range from 0 to 100.

The calculation for scoring these scales is the same in all cases:

1. Calculate the average of the items that contribute to the scale; this is the raw score.
2. Use a linear transformation to standardize the raw score, so that scores range from 0 to 100; a higher score represents a higher (i.e. better) level of functioning, or a higher (i.e. worse) level of symptoms.

Calculations for raw score and linear transformation are as follows:

Let items $I_1, I_2, \ldots I_k$ be included in a scale, where $k$is number of items in the scale, then:

$$Raw\ Score = RS = \frac{\sum_{j=1}^{k} I_j}{k}$$

To obtain the score for functional scales: $\left\{ = - \frac{RS-1}{range} \right\} \times 100$

To obtain the score for symptom scales/items and global health status/QoL: $= \left\{ \frac{RS-1}{range} \right\} \times 100$

Range is the difference between the maximum possible value of *RS* and the minimum possible value.

The structure of the EORTC-QLQ-C30 questionnaire is presented in Scoring for QLQ-C30 version 3.0

Table 4 in Appendix 14.3).

At each assessment point, summary statistics of absolute scores and changes from baseline will be calculated by treatment arm for each scale for both the raw score and the linear transformed score. The baseline result will be the mean of the screening visit score and the score at last assessment prior to dosing. Summary tables and boxplots of the score by visit and treatment arm will be presented. The best and worst changes from baseline and associated time points will also be summarized using descriptive statistics. For functional scales and global health status, best changes from baseline are positive changes; while for symptom scales/items, best changes are negative changes.

The percentage of patients missing the entire questionnaire will be summarized and reason it is missing will be listed by treatment arm.

If more than 50% of questions are unanswered within a subscale at a particular timepoint, the value for the subscale will be set to missing. If less than 50% of questions are unanswered, the

sum of the scores for the answered questions will be taken and scaled up in relation to the number of questions in the scale answered or not.

The change from baseline for the linear transformed score will be compared between treatment arms for the PRO (3rd line+) set using mixed model repeated measures (MMRM) analysis with fixed effect terms for treatment arm, visit, treatment arm by visit interaction, the randomization stratification factors and also with baseline score as a covariate. Patient will be included as a random effect. The default estimation method of Restricted Maximum Likelihood (REML) and the Kenward-Roger (KR) option for degrees of freedom will be used. An unstructured covariance matrix will be used to model the within-patient error. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, and autoregressive. A summary table will present number of subjects in each treatment group at each visit, LSMeans by treatment arm by visit, along with the difference in LSMeans, standard error, 95% CI and 2-sided p-value.

Patients will also be classified as improved, stable, or worsened. Minimal important difference (MID) thresholds, a threshold of 5 score points of the linear transformed score, will be used to categorize patients as improved, stable, or worsened. The first post-baseline assessment where either improved or worsened status is met will be used to assign patient status. The improved or worsened status has to be confirmed by the assessment at subsequent cycle.

The proportion of patients with each status will be calculated for each subscale based on patients who have at least two available change from baseline of the linear transformation scores in that scale. A proportional odds model will be used to test the difference in proportions of patient status between treatment arms for each scale; a fixed effect term for treatment will be included and the model will be stratified for randomization stratification factors. The odds ratio, 95% CI and 2-sided p-value will be extracted from the model and presented.

The individual questions and derived scales scores will be listed for the PRO (3rd line+) set.

8.5.2. EQ-5D

EQ-5D is a standardized measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal. Analysis of the EQ-5D data will be performed for the PRO (3rd line+) set.

The EQ-5D consists of two parts, the EQ-5D descriptive system and the EQ Visual Analogue Scale (EQ VAS).

EQ-5D Descriptive System

The EQ-5D descriptive system comprises 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). Each dimension has 5 levels (no problems, slight problems, moderate problems, severe problems, and extreme problems). The respondent is asked to indicate his/her health state or placing a cross in the box against the most appropriate level within each of the 5 dimensions. This decision results in a 1-digit number expressing the level selected for that dimension. The user guide emphasizes that the numerals 1-5 for each level have no arithmetic properties and should not be used as a cardinal score. Ambiguous item values (e.g. 2 levels are ticked for a single dimension) will be treated as missing values.

For the 5 dimensions within the descriptive, at each assessment point, the number and percentage of patients responding in each of the 5 levels will be presented in a table as well as displayed by treatment arm. The number of missing responses will also be displayed. The response profile for

each of the 5 dimensions will be compared between treatment arms over time using generalized estimating equations (GEE). The marginal model will take the form

$$E\left[Log[P(X \leq j)]\right] = \beta_{0j} + T \cdot \beta_T + V \cdot \beta_V + T \cdot V \cdot \beta_{TV}$$

where $E[\cdot]$ denotes expectation, X is one of the 5 dimensions, $j = 1, \dots, 5$ is the level of the dimension, $P(X \leq j)$ denotes the probability a patient's response is $\leq j$, β_{0j} is the intercept term associated with level $j = 1, \dots, 5$, $T = 0, 1$ is the treatment arm indicator, $V = 0, 1, \dots, v$ is the visit indicator, β_T are the regression parameters associated with treatment arm level $T = 0, 1$, β_V are the regression parameters associated with visit $V = 0, 1, \dots, v$ and β_{TV} are the regression parameters associated with treatment arm T and visit V.

Results will be displayed as well as presented in a table, in terms of the overall odds ratio for balixafortide + eribulin vs eribulin monotherapy along with the associated 95% CI and 2-sided p-value. Odds ratios by visit will be displayed also.

EQ VAS

The EQ VAS records the patient's self-rated health visual analogue scale with endpoints labelled 'the best health you can imagine' and 'the worst health you can imagine'. The patient is asked 'mark an X on the scale to indicate how your health is TODAY' and then to 'write the number you marked on the scale in the box below'. There will be no programming checks for a discrepancy between where the patient has placed the X and the number they have written in the box, the number in the box is the only value available on the database and to be used for analysis.

EQ VAS scores will be analyzed across visits using an MMRM approach in the same manner as described in Section 8.5.1 for the EORTC-QLQ-C30.

8.5.3. FACT-B

The FACT-B (version 4) is a self-administered questionnaire designed for patients suffering from breast cancer; FACT-B is a breast cancer subscale added to the Functional Assessment for Cancer Therapy –General (FACT-G). It consists of 37 questions grouped into 5 subscales being:

- Physical well-being (7 questions GP1 to GP7, min score=0, max score=28)
- Social/family well-being (7 questions GS1 to GS7, min score=0, max score=28)
- Emotional well-being (6 questions GE1 to GE6, min score=0, max score=24)
- Functional well-being (7 questions GF1 to GF7, min score=0, max score=28)
- Breast subscale (10 questions B1 to B9 and P2, min score=0, max score=40)

All questions are presented with answers on a five-point (0 to 4) Likert scale. As displayed in Appendix 14.4, subscale scores are calculated by first transforming all question scores within a subscale to have the same direction such that the higher the score, the better the QOL, and then summing the resulting scores. If more than 50% of questions are unanswered within a subscale at a particular timepoint, the value for the subscale will be set to missing. If less than 50% of questions are unanswered, the sum of the scores for the answered questions will be taken and scaled up in relation to the number of questions answered or not in the scale. The subscale scores will be combined as follows:

FACT-G total score: PWB score + **SWB** score + E**WB** score + **FWB** score

FACT-B total score: PWB score + **SWB** score + E**WB** score + **FWB** score + **BCS** score

No imputation will be done if any of the subscores is missing.

Analysis of the FACT-B data will be performed in the PRO set. The 5 subscale scores along with the FACT-G score and FACT-B total score will be summarized by treatment arm at each assessment point using mean, median, standard deviation, min, max and N.

Each of the subscale scores together with the FACT-G score, FACT-B total score will be analyzed over time using an MMRM approach in the same manner as described for the EORTC-QLQ-C30.

8.5.4. Relationship between biomarkers and clinical outcomes

Relation between biomarkers and clinical outcome will be treated in a separate analysis plan.

8.5.5. Response in patient who continue study treatment despite progression

Response in patients who continue study treatment despite progression will be assessed according to iRECIST (2017) by Investigator. A summary table of counts and percentage of patients experiencing each of CR or PR measured at an assessment date later than PD, by visit will be presented by treatment arm. Analyses on patients who continue study treatment after the progression will be descriptive only because these patients did not go through a second randomization process and can be affected by selection bias.

8.5.6. Baseline Tumor Samples

Archival Tumor Tissue and Fresh Tumor Biopsy information will be listed for the ITT (3rd line+) set.

9. Safety Analysis

Safety data will be assessed in the 3rd line+ and overall (2nd line+) populations. Safety data collected in this study includes AEs, serious adverse events (SAEs), clinical laboratory tests, vital signs, physical examination, electrocardiogram, and other safety data including pregnancy during exposure and medication errors. To maximize the ability to discern safety signals, safety and AE data will be summarized on the Safety and Safety (3rd line+) sets, i.e. the overall and 3rd line+ population and will be presented for balixafortide + eribulin and eribulin monotherapy treated patients. All safety data will be presented in data listings.

9.1. Adverse Events

Adverse events will be coded by SOC and preferred term using the latest MedDRA dictionary. Adverse event severity will be based on NCI CTCAE Grade (version 5.0).

All AEs reported from the first dose of study treatment (Cycle 1 Day 1 for balixafortide + eribulin or Cycle 1 Day 2 for eribulin monotherapy) until 30 days after the last dose of study treatment will be considered as treatment-emergent AEs (TEAEs). A TEAE is an AE that was not present prior to first dose of study treatment but appeared following treatment or was present prior to the start of study treatment but worsened thereafter. An AE that was present prior to the start of study treatment, resolved and then reappeared following first dose of study treatment, is a TEAE. If the start date of an AE is incomplete and cannot be categorized as occurring before or after the first

dose of study treatment, then the AE will be considered as a TEAE. If the start time of the AE is presented, then the AE start date and time will be used to classify the AE as TEAE or not. If the start time is not provided and the AE occurs on the same day of first dose of study treatment, then the AE will be classified as a TEAE.

9.1.1. Incidence of Adverse Events

AE data will be descriptively summarized by treatment arm. The incidence of TEAEs by MedDRA SOC, preferred term, and Investigator attributed relationship (related/not related) to balixafortide and/or eribulin, as relevant, will be summarized. AEs will be summarized in terms of frequency and percentage. For patient level of summaries, patients with multiple occurrences of events of the same preferred terms and SOC will be counted once for each preferred term or SOC (at the highest severity/grade and the closest relationship to study treatment as applicable). The summary tables will be ordered by descending incidence of SOC and by descending incidence of preferred term within each SOC class for all patients.

The following adverse events summaries will be provided:

- TEAEs by System Organ Class and Preferred Term
- TEAEs by System Organ Class and Preferred Term and CTCAE Grade
- Grade 3 or Higher TEAEs by System Organ Class and Preferred Term
- Grade 3 TEAEs by System Organ Class and Preferred Term
- Grade 4 TEAEs by System Organ Class and Preferred Term
- Serious TEAEs by System Organ Class and Preferred Term
- Serious TEAEs by System Organ Class and Preferred Term and CTCAE Grade
- TEAEs Related to Study Treatment by System Organ Class and Preferred Term
- TEAEs Related to Study Treatment by System Organ Class and Preferred Term and CTCAE Grade
- TEAEs Leading to Study Treatment Discontinuation by System Organ Class and Preferred Term
- TEAEs Leading to Study Treatment Discontinuation by System Organ Class and Preferred Term and CTCAE Grade
- TEAEs with fatal outcome by System Organ Class and Preferred Term, being AEs with the outcome 'Death Related to Adverse Event' regardless of the attributed relationship to study treatment
- TEAEs Related to Study treatment with outcome 'Death Related to Adverse Event' by System Organ Class and Preferred Term
- TEAEs of Special Interest by Special Interest Group
- TEAEs of Particular Interest by Particular Interest Group

All adverse events will be presented in a listing. Furthermore, listings will be presented specifically for SAEs, TEAEs leading to study treatment discontinuation and TEAEs leading to death.

In addition, the number of TEAEs leading to first eribulin dose reduction will be summarized by cycle and day of onset, in total and split by treatment arm.

9.1.2. Severity of Adverse Event

The severity that will be presented represents the most extreme severity captured on the eCRF page. The possible severities are measured on the CTCAE scale which goes from 1 to 5 where 1='mild', 2='moderate', 3='severe', 4='life-threatening' and 5='fatal'. Treatment-emergent AEs with a missing severity will be summarized in tables as Grade 3 but will be presented in the data listing with a missing severity.

In the TEAE severity table, if a patient reported multiple occurrences of the same TEAE, only the most severe will be presented. A further summary table of only those TEAEs with CTCAE grade 3 or higher will also be presented.

9.1.3. Serious Adverse Events

A SAE is defined in the protocol as any event that: results in death, is immediately life threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity, or is a congenital anomaly/birth defect. Important medical events that may not result in death, be life-threatening, or require hospitalization may be considered SAEs when, based upon medical judgment, they may jeopardize the patient and may require medical or surgical intervention to prevent one of the outcomes listed above. A summary table will be presented, together with a further summary table by CTCAE grade.

9.1.4. Relationship of Adverse Events to Study Treatment

The investigator will provide an assessment of the relationship of the event to each of balixafortide and eribulin. The possible relationships are 'related', and 'unrelated'.

Treatment-emergent AEs with a missing relationship will be summarized in tables as 'Related' but will be presented in the data listing with a missing relationship.

A summary table of related TEAEs will be presented, together with a further summary table by CTCAE grade.

9.1.5. Adverse Events Leading to Study Treatment Discontinuation

Following the AE, the investigator will record the action taken with each of the study treatments. If at least one of the drugs has an action taken equal to 'Drug Withdrawn', then the AE will be considered as leading to study treatment discontinuation. The summary table will present overall covering all treatment discontinuations, and separately for each of balixafortide and eribulin where Action Taken with respective treatment is 'Drug Withdrawn'. A summary table by CTCAE grade will also be presented.

9.1.6. Deaths

An AE is leading to death if the outcome is "Death Related to Adverse Event". All deaths and primary reason for patient death will be summarized by treatment arm; deaths within 30 days of the last dose of randomized treatment trial drug will be included. Additionally, deaths related to study treatment will be summarized by treatment arm.

9.1.7. Subgroup Evaluation of Adverse Events

AEs will be also summarized for the following subgroups:

- Patient age (<65yr vs ≥65yr)
- Bone metastases at baseline (Yes vs No)
- Liver metastases at baseline (Yes vs No)

9.1.8. Adverse Events of Special and Adverse Events of Particular Interest

Adverse Events of Special Interest (AESI) and Adverse Events of Particular Interest (AEPI) will be identified by the investigator and specified in the eCRF.

The list of AESI is: Hypersensitivity or IRRs (serious), Neutropenia (serious), Neutropenic infections, Febrile neutropenia, Renal impairment.

The list of AEPI is: Thrombocytopenia, Peripheral neuropathy, Hepatic impairment, QTc prolongation, Hypersensitivity/IRRs, Neutropenia (serious), Neutropenic infections, Febrile neutropenia and Renal impairment.

The number and percent of patients reporting an AESI or an AEPI will be summarized by treatment arm, MedDRA SOC and Preferred term. All AESI categories will be summarized, while only the following categories will be summarized for AEPI: Hypersensitivity/IRRs, Thrombocytopenia, QTc prolongation, Peripheral neuropathy and Hepatic impairment.

Number of patients reporting at least one event of Hypersensitivity/IRR, Neutropenia (serious) or Febrile neutropenia will also be summarized by maximum severity (Grade 1 to 5) and treatment arm.

The number of patients reporting Neutropenic infections will be summarized by treatment arm and maximum severity (Grade 1 to 5). This table will only be populated when at minimum 4 patients reported at least 1 event of Neutropenic infection and the number of such patients is more than 1% of the Safety population.

The number of Colony Stimulating Factor (CSP) medications will be summarized for all patients experiencing Neutropenia events.

The number of Hypersensitivity/IRRs and Neutropenia (serious) will be summarized by cycle, day of onset, in total and split by treatment arm. The number of all Hypersensitivity/IRR will also be summarized in terms of the event intensity allowing for multiple reactions within a patient total; the latter will be achieved by means of a Negative Binomial analysis with a fixed effect term for randomized treatment and time exposed to randomized treatment as the offset. The rate ratio and the corresponding 95% confidence interval will be obtained based on the Wald approach using the log as link function.

Neutropenia (serious) and Febrile neutropenia events will be listed reporting the cycle and day of onset and the duration of the event. The time spent experiencing Neutropenia (serious) and Febrile neutropenia will be summarized by treatment group and severity using the Probability of Being in Event Function, as described by Begg and Larsson (1982).

Separate Kaplan-Meier analysis will be conducted to describe the time to onset of the AESI and AEPI for both treatment arms. Time to onset (months) is defined and calculated as (the earliest date of event onset –date of first dose + 1)/30.4375. Patients who did not develop the event will be censored at their date of death, last trial visit, or the start date of new anti-cancer therapy,

whichever is earlier. Median time to onset and its 95% CI will be estimated for both treatment arms. If the median time to event cannot be estimated due to lack of events, the 25th percentile and its 95% CI will be reported.

9.2. Clinical Laboratory Evaluations

All safety laboratory evaluations will be done at local laboratories. Blood and urine samples for laboratory analysis are collected at Screening, on Day 1 and Day 8 of each cycle for balixafortide + eribulin, on Day 2 and Day 9 of each cycle for eribulin monotherapy, and at the end of treatment visit. The results of Day 1 for balixafortide + eribulin and Day 1 and Day 2 for eribulin monotherapy will be presented together. The results of Day 8 for balixafortide + eribulin and Day 8 and Day 9 for eribulin monotherapy will be presented together.

For hematology and biochemistry, absolute values and changes from baseline will be summarized by study treatment arm and cycle/day number. Change from baseline will be calculated as (value at a treatment cycle/day ~~v~~alue at baseline).

For those parameters where NCI CTCAE v5.0 grades are derivable, shift tables from baseline to each post-baseline time point will be presented by NCI CTCAE v5.0 grades. For all the other parameters, shift tables will be presented by analysis range (normal, high or low). For any shift table, percentages will be calculated using available data per cycle/day number for each laboratory parameter.

All safety laboratory test results will be summarized by study treatment arm, overall, by cycle and by day and also presented in data listings. Coagulation parameters (International Normalized Ratio [INR] and prothrombin time) will be listed only. Boxplots of mean ± standard error at each cycle/day number will be created for each parameter.

Urinalysis data will be listed only.

9.2.1. Hematology

The following parameters will be assessed: white blood cell (WBC) count with differential, red blood cell (RBC) count, hemoglobin (Hb), hematocrit (Ht), platelet count, Neutrophils, Lymphocytes, Eosinophils, Monocytes and Basophils.

9.2.2. Biochemistry

The following parameters will be assessed: albumin, total protein, alanine aminotransferase (ALT), aspartate transaminase (AST), total bilirubin (TBL), direct bilirubin, indirect bilirubin, gamma-glutamyl transferase (GGT), alkaline phosphatase (ALP), glucose, creatinine, blood urea nitrogen (BUN), urea, creatine kinase (CK), calcium (Ca), potassium (K), sodium (Na), and magnesium (Mg).

Potential cases of drug induced liver injury will be listed and summarized by treatment arm. Potential cases of drug induced liver injury are defined using concurrent ALT, AST and TBL as follows:

- For patients with all 3 parameters (ALT, AST and TBL) within the normal range at baseline:
    - ALT or AST ≥3 times the upper limit of normal (x ULN) with TBL ≥2 x ULN with no evidence of hemolysis and an ALP value ≤2x ULN or not available.;
- For patients with pre-existing ALT or AST values above the ULN at baseline:

- (ALT or AST) ≥2 x baseline and (ALT or AST) ≥3 x ULN and TBL ≥2 x ULN. Note that the double criterion on (ALT or AST) requires values ≥ maximum value of 2 x baseline and 3 x ULN which ensures an increase from baseline even in the case of a very large baseline value >3 x ULN;
- As above with additional more stringent criterion of (ALT or AST) ≥8 x ULN.

- For patients with pre-existing values of total bilirubin above the normal range at baseline:
  - (ALT or AST) ≥3 x ULN and (TBL>= baseline + ULN and TBL ≥3 x ULN);
  - As above with additional more stringent criterion of (ALT or AST) ≥8 x ULN.

A log-log plot of maximum ALT/AST/ALP vs maximum total bilirubin will be created to identify potential cases of concern. Additionally, an individual profile over time for each ULN-normalized liver function test (ALT, ALP, AST and TBL) will be plotted for all the potential cases.

9.2.3. Urinalysis

The following parameters will be assessed: appearance, color, dipstick (pH, Specific gravity, ketones, protein, glucose, blood, bilirubin, nitrite, leukocytes, bacteria, casts, crystals, epithelial cells, red blood cells, white blood cells and other results). Microscopic analysis if warranted by Dipstick results. Bacteria, casts, crystals, epithelial cells, red blood cells, white blood cells and other results come under the microscopic analysis on the eCRF.

9.3. Vital Sign Measurements

Weight, BSA, body temperature and resting blood pressure and respiratory and pulse rate will be measured in all patients at Screening, and at the EoT visit.

Weight and vital signs will also be assessed on Day 1 of each cycle for patients on the balixafortide + eribulin treatment arm and on Day 2 of each cycle for patients on the eribulin monotherapy treatment arm.

The vital sign results and changes from baseline will be summarized by cycle and treatment arm. A data listing will be produced as well.

9.4. Physical Examination

Physical examination will be performed at Screening, on Day 1 of each cycle for patients on the balixafortide + eribulin treatment arm and on Day 2 of each cycle for patients on the eribulin monotherapy treatment arm and in all patients at the EoT visit.

A summary table for the shift from baseline to post-baseline timepoints for physical exam will be presented by treatment arm. All physical examination results will be presented in a data listing.

9.5. Electrocardiogram

An electrocardiogram (ECG) will be performed at the clinical site during the visits and timepoints shown below.

Treatment Arm	Visit	Day	ECG Timepoint	ECG Type
All	Screening	-21 to -1	Anytime	12-lead

Treatment Arm	Visit	Day	ECG Timepoint	ECG Type
Balixafortide + eribulin	Cycle 1	1	Within 1 hour prior to administration of balixafortide	Duplicate 12-lead
Balixafortide + eribulin	Cycle 1	2	Immediately before starting the balixafortide infusion	12-lead
Eribulin monotherapy	Cycle 1	2	Immediately before starting the eribulin infusion	12-lead
All	Cycle 1	2	Within 1 hour after completing the eribulin infusion	12-lead
Balixafortide + eribulin	Cycle 1	9	Immediately before starting the balixafortide infusion	12-lead
Eribulin monotherapy	Cycle 1	9	Immediately before starting the eribulin infusion	12-lead
All	Cycle 1	9	Within 1 hour after completing the eribulin infusion	12-lead
Balixafortide + eribulin	Cycle 2 & subsequent cycles[a]	2	Immediately before starting the balixafortide infusion	12-lead
Eribulin monotherapy	Cycle 2 & subsequent cycles[a]	2	Immediately before starting the eribulin infusion	12-lead
All	Cycle 2 & subsequent cycles[a]	2	Within 1 hour after completing the eribulin infusion	12-lead
Balixafortide + eribulin	Cycle 2 & subsequent cycles[a]	9	Immediately before starting the balixafortide infusion	12-lead
Eribulin monotherapy	Cycle 2 & subsequent cycles[a]	9	Immediately before starting the eribulin infusion	12-lead
All	Cycle 2 & subsequent cycles[a]	9	Within 1 hour after completing the eribulin infusion	12-lead
All	EoT	-	Anytime	12-lead

a: After Cycle 2, the frequency of ECG recordings can be reduced to every 2 cycles if deemed clinically appropriate (i.e. absence of ECG findings in earlier assessments).

ECG measurements include Ventricular rate, PR interval, QT interval, QTcF interval, RR interval, and QRS duration. Summary tables of ECG results will be presented and changes from baseline will be summarized by treatment arm and by cycle, by day and by ECG timepoint. Any clinically significant changes will be recorded as AEs.

QTcF intervals between >480 ms and ≤500 ms (Grade 2) or between >500 ms or <600 ms (Grade 3) will be identified and summarized by study treatment arm, by cycle, by day and by ECG timepoint.

9.6. Other Safety Data

9.6.1. Medication Errors

The number of dosing errors will be summarized for each medication error event (Overdose, Consumption of the study treatment by the wrong patient, Other) by visit and overall. All the other variables collected for study medication errors will be listed only.

9.6.2. ECOG Performance Score

ECOG performance score at each timepoint and its change from baseline will be summarized by treatment arm. The data will also be listed.

9.6.3. Exposure during pregnancy

Pregnancy outcomes during exposure will be handled outside this execution plan, and relevant details will be documented in the Clinical Study Report.

10. Pharmacokinetics

Pharmacokinetics and Immunogenicity will be treated in a separate analysis plan.

11. Interim Analysis

A DSMB is established with the responsibility of safeguarding the safety of study participants. The DSMB will review accumulating safety data during the conduct of the study and the interim analyses. The DSMB will provide ongoing review and recommendations to the Sponsor as necessary in relation to study conduct and patient management.

The DSMB charter will describe the operational aspects of the DSMB and the analytical and monitoring plan for the study. The Charter will further state the primary responsibilities of the DSMB, its relationship with Polyphor, its membership, and the purpose and timing of its meetings, statistical monitoring guidelines to be implemented, and statistical analysis for open and closed sessions. The DSMB Charter and meeting minutes will be submitted as part of the final Clinical Study Report.

12. Changes in the Planned Analysis

The changes from planned analyses described in the protocol and in the Statistical Analysis Plan (SAP) version 1.4 dated 18Jan2019 to those presented in this execution plan are the followings:

- Added PP set definition and analysis of coprimary endpoints: PFS and OS in PP (3rd line+ confirmed) set. Clarified definition of 3rd line+ Confirmed in analysis sets is based on randomization strata and eCRF data.
- Added sensitivity analysis for PFS and OS in ITT 3rd line+ (ITT 3rd line+ Confirmed and PP 3rd line+ Confirmed).
- Added sensitivity analysis for ORR in RE 3rd line+ (RE 3rd line+ Confirmed).
- Data listings for prior systemic cancer/oncology therapies will include only dose, unit, dose regimen, frequency, best response to cancer therapy, and reason for discontinuation (toxicity, other or progressive disease).
- For this study, up to three (3) intervening NEs will be allowed to confirm a CR or PR. Subsequent documentation of a CR may provide confirmation of a previously identified CR with up to three (3) intervening non- evaluable (NE) assessments (e.g. CR NE CR, CR NE NE CR, or CR NE NE NE CR). Similarly, subsequent documentation of a PR may provide confirmation of a previously identified PR with up to three (3) intervening NEs (e.g., PR NE PR, PR NE NE PR, or PR NE NE NE PR). However, only one (1) intervening SD will be allowed between PRs for confirmation (i.e., PR SD PR).
- Listing of patients with clinically significant abnormalities in vital signs has been removed.
- All AESIs and only the non-overlapping categories for AEPIs will be summarized by treatment arm, MedDRA SOC, Preferred term, and toxicity grade.
- Prior anticancer therapies will not be summarized as regimens since regimens are not identified in eCRF. Individual therapies instead of regimens will be reported.
- Subgroups have been updated in the exploratory efficacy endpoints, section 8.5.
- COVID section added
- Medication errors will not be reported by MedDRA. Instead, they will be reported by medication error event overall and at each visit.

13. References

- Begg BB, Larson M. A study of the use of the Probability-of-being-in-response function as a summary of tumor response data. *Biometrics* 1982; 38:59–66.
- Ellis S, Carroll KJ and Pemberton K. Analysis of Duration of Response in Oncology Trials. Contemporary Clinical Trials, 2008; 29: 456-465.
- Carroll, KJ. Back to basics: explaining sample size in outcome trials, are statisticians doing a thorough job? Pharm Stats, 2009; 8: 333–345.
- Dixon D and Simon R. Bayesian Subset Analysis. Biometrics, 1991; 47:871-881.
- MedDRA Dictionary (https://www.meddra.org/)
- NCI CTCAE Grade https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm

- Sampson HA. Second Symposium on the Definition and Management of Anaphylaxis: Summary Report—Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network Symposium. Annals of Emergency Medicine. 2006;47(4):373-80.

- FDA 2009. Guidance for industry. Drug-induced livery injury: premarketing clinical evaluation. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090.pdf

14. Appendices

14.1. General Schedule of Assessments

14.1.1. Schedule of Assessments for Balixafortide + Eribulin Treatment Arm

Table 2 - Schedule of Assessments for Balixafortide + Eribulin Treatment Arm: Screening, Cycle 1, and Cycle 2

Visits (Visit Window days)	Screening (-21 to -1)	Active treatment phase (21 Days per cycle)												
		Cycle 1						Cycle 2						
Assessment		**D1**	**D2**	**D3**	**D8**	**D9**	**D10**	**D1**	**D2**	**D3**	**D8**	**D9**	**D10**	**D21**
Informed consent	X													
IWRS/IRT	X	X	X	X	X	X	X	X	X	X	X	X	X	
Inclusion/exclusion criteria	X													
Demographics (gender, date of birth, race, ethnicity) and baseline data	X													
Medical history	X													
Patient reported health outcomes	X	X						X						
Physical examination[1]	X[2]	X[3]						X[3]						
Vital signs (body temperature, resting blood pressure, respiratory rate, pulse rate)[1]	X	X						X						
Weight, height[4]	X	X						X						
ECOG performance status	X	X						X						
Adverse events	X[5]	X----------X												
Prior/Concomitant medications	X	X----------X												
Clinical Laboratory tests[1]	X[6]	X[7]			X[7]			X[7]			X[7]			

Visits (Visit Window days)	Screening (-21 to -1)	Active treatment phase (21 Days per cycle)												
		Cycle 1						Cycle 2						
Assessment		**D1**	**D2**	**D3**	**D8**	**D9**	**D10**	**D1**	**D2**	**D3**	**D8**	**D9**	**D10**	**D21**
Pregnancy test	X[8]	X[9]						X[9]						
Contraceptive advice	X													
CT/MRI scan[10, 11]	X[12]													X[10]
Bone scan	X[12, 13]													
12-lead ECG[1]	X	X	X			X			X[14]			X[14]		
Document date and results of last tumor biopsy/fresh tumor biopsy	X													
Blood for PK			X		X	X			X		X	X		
Blood for immunogenicity testing		X			X			X			X			
Blood for exploratory biomarkers in all patients[15]		X	X	X	X	X								
Confirmation of eligibility and randomization		X												
Balixafortide treatment[16]		X	X	X	X	X	X	X	X	X	X	X	X	
Eribulin treatment			X			X			X			X		
For patients at selected sites														
Optional tumor biopsy[17]	X											X[18]		
Blood for other exploratory biomarkers		X	X		X							X[19]		

MRI: magnetic resonance imaging; RECIST: Response Evaluation Criteria In Solid Tumors; SAE: serious adverse event. **Footnotes:**

1. Additional assessments to those scheduled for this study should be performed as indicated at the Investigator's discretion (e.g. due to concerns about patient safety, patient's clinical status); the findings of an unscheduled examination should be recorded in the patient's source data and eCRF.
2. Standard, complete physical examination.
3. Targeted physical examination.

4. Height will be measured at Screening only.
5. SAEs will be recorded from the date that the informed consent form is signed.
6. Screening blood tests should be taken within 7 days prior to randomization.
7. Blood test results must be available within 24 hours prior to study medication administration.
8. Serum pregnancy test for women of childbearing potential only, no more than 7 days prior to Randomization. More frequent assessments can be performed if medically indicated as determined by the Investigator, and these evaluations will be recorded in the patient's source data and eCRF.
9. Urine pregnancy test for women of childbearing potential only.
10. CT/MRI scan will be performed at Screening, then from the date of randomization every 6 weeks (±7 days) during the first year, and then every 12 weeks (±7 days), thereafter, until PD is documented by RECIST v1.1. It is critical that tumor assessments are performed according to the Study Schedule in a timely and complete manner regardless of any dosing delay or interruption. Imaging assessments will be scheduled using the randomization date as the reference date for all timepoints and NOT scheduled according to the date of the previous timepoint. A delay in the imaging assessment to accommodate a treatment delay is not permitted. Following initial documentation of an objective response, a confirmatory CT/MRI scan will be performed at least 4 weeks later.
11. At the discretion of the Investigator, additional radiographic tumor assessments may be done at any time if there is clinical suspicion of disease progression; the results of the unscheduled scan should be recorded in the patient's source data and eCRF.
12. Record date and results of last scan on the patient's source data and eCRF. If no scan has been performed in the 30 days prior to Randomization a scan should be performed during Screening.
13. From the date of randomization, additional bone scans will be performed at the discretion of the Investigator.
14. After Cycle 2, the frequency of ECGs can be reduced to every 2 cycles if deemed clinically appropriate (i.e. absence of ECG findings in earlier assessments).
15. Blood will be collected from all patients to assess plasma levels of interferon gamma and additional cytokine markers.
16. Patients will start study medication within 3 days from randomization.
17. Patients at selected sites can optionally consent to fresh tumor biopsy at Screening and subsequent timepoints. If a biopsy has already been performed within 30 days prior to Randomization and tumor tissue is still available, then the fresh tumor biopsy is not required during Screening.
18. Alternatively, can be performed on Day 10 if preferred.
19. A blood sample will be taken only in patients at selected sites who are scheduled to have an optional, fresh tumor biopsy; the blood sample can be taken before or after the biopsy.

Table 3 - Schedule of Assessments for Balixafortide + Eribulin Treatment Arm: Cycle 3, Subsequent Cycles, and Post-treatment Phase

Visits (Visit Window days)	Active treatment phase (21 Days per cycle)															Post-treatment phase		
	Cycle 3								Subsequent Cycles							EoT (+7)	30-day safety FU (+7)	Long-term FU (±7)[1]
Assessment	D1	D2	D3	D4	D5	D8	D9	D10	D1	D2	D3	D8	D9	D10	D21			
IWRS/IRT	X	X	X			X	X	X	X	X	X	X	X	X		X		
Patient reported health outcomes	X								X							X		
Targeted physical examination[2]	X								X							X		
Vital signs (body temperature, resting blood pressure, respiratory rate, pulse rate)[2]	X								X							X		
Weight	X								X							X		
ECOG performance status	X								X							X		
Adverse events	X----------														----------X			
Prior/Concomitant medications	X----------														----------X		X[3]	
Clinical Laboratory tests[2]	X[4]					X[4]			X[4]			X[4]				X		
Pregnancy test	X[5]								X[5]							X[5]		
CT/MRI scan[6, 7]															X[6]	X[8]		X[1]
Bone scan[9]																		
12-lead ECG[2]		X[10]					X[10]			X[10]			X[10]			X		
Blood for PK		X		X[11]	X[11]					X[12]	X[12]							
Blood for immunogenicity testing	X								X[13]							X		

Visits (Visit Window days)	Active treatment phase (21 Days per cycle)															Post-treatment phase		
	Cycle 3								Subsequent Cycles							EoT (+7)	30-day safety FU (+7)	Long-term FU (±7)[1]
Assessment	D1	D2	D3	D4	D5	D8	D9	D10	D1	D2	D3	D8	D9	D10	D21			
Blood for exploratory biomarkers in all patients[14]	X	X	X	X[15]	X[15]	X	X											
Balixafortide treatment	X	X	X			X	X	X	X	X	X	X	X	X				
Eribulin treatment		X					X			X			X					
Document reason for treatment discontinuation																X		
Survival status, PD monitoring, all anticancer medicines received																		X
For patients at selected sites																		
Optional tumor biopsy																X		
Blood for other exploratory biomarkers	X	X														X		

AE: adverse event; MRI: magnetic resonance imaging; RECIST: Response Evaluation Criteria In Solid Tumors. **Footnotes:**

1. Regardless of treatment arm, patients discontinued from study medication for any reason, in the absence of PD, will undergo repeat imaging and tumor response assessments (including CT/MRI scans) every 8 weeks ±7 days (every 12 weeks ±7 days if the patient has been on the study for ≥1 year) until PD is documented as per RECIST v1.1, or death occurs, or the patient is lost to follow-up, or the patient withdraws consent (whichever occurs first). Once PD is recorded, the patient will be followed every 6 months (or more frequently) for Survival Follow-up only, until the patient dies, withdraws consent, or is lost to follow-up (whichever occurs first). Survival follow-up can be conducted by telephone.

2. Additional assessments to those scheduled for this study should be performed as indicated at the Investigator's discretion (e.g. due to concerns about patient safety, patient's clinical status); the findings of an unscheduled examination should be recorded in the patient's source data and eCRF.
3. Concomitant therapy/medications taken since previous visit to be recorded for patients with unresolved AEs.
4. Blood test results must be available within 24 hours prior to study medication administration.
5. Urine pregnancy test for women of childbearing potential only.
6. CT/MRI scan will be performed at Screening, then from the date of randomization every 6 weeks (±7 days) during the first year, and then every 12 weeks (±7 days), thereafter, until PD is documented by RECIST v1.1. It is critical that tumor assessments are performed according to the Study Schedule in a timely and complete manner regardless of any dosing delay or interruption. Imaging assessments will be scheduled using the randomization date as the reference date for all timepoints and NOT scheduled according to the date of the previous timepoint. A delay in the imaging assessment to accommodate a treatment delay is not permitted. Following initial documentation of an objective response, a confirmatory CT/MRI scan will be performed at least 4 weeks later.
7. At the discretion of the Investigator, additional radiographic tumor assessments may be done at any time if there is clinical suspicion of disease progression; the results of the unscheduled scan should be recorded in the patient's source data and eCRF.
8. CT/MRI scans will be performed in patients who are discontinued from study medication in the absence of PD; however, patients who have already demonstrated objective PD (according to RECIST v 1.1) do not need to have scans repeated at the EoT visit.
9. From the date of randomization, additional bone scans will be performed at the discretion of the Investigator.
10. After Cycle 2, the frequency of ECGs can be reduced to every 2 cycles if deemed clinically appropriate (i.e. absence of ECG findings in earlier assessments).
11. For patients at selected sites only.
12. Cycle 5 and Cycle 7 only.
13. After Cycle 3, blood samples are taken on Day 1 of every alternate cycle (e.g. Cycle 5, Cycle 7, Cycle 9 etc).
14. Blood will be collected from all patients to assess plasma levels of interferon gamma and additional cytokine markers.
15. Only in patients at selected sites scheduled to have PK sampling.

14.1.2. Schedule of Assessments for Eribulin Monotherapy Treatment Arm

	Screening (-21 to -1)	Active treatment phase (21 Days per cycle)										Post-treatment phase		
		Cycle 1		Cycle 2			Cycle 3		Subsequent Cycles			EOT (+7)	30-day safety FU (+7)	Long-term FU (±7)[1]
Assessment		D2	D9	D2	D9	D21	D2	D9	D2	D9	D21			
Informed consent	X													
IWRS/IRT	X	X	X	X	X		X	X	X	X		X		
Inclusion/exclusion criteria	X													
Demographics (gender, date of birth, race, ethnicity) and baseline data	X													
Medical history	X													
Patient reported health outcomes	X	X		X			X		X			X		
Physical examination[2]	X[3]	X[4]		X[4]			X[4]		X[4]			X[4]		
Vital signs (body temperature, resting blood pressure, respiratory rate, pulse rate)[2]	X	X		X			X		X			X		
Weight, height[5]	X	X		X			X		X			X		
ECOG performance status	X	X		X			X		X			X		
Adverse events	X[6]	X----------												X
Prior/Concomitant medications	X	X----------											X[7]	
Clinical laboratory tests[2]	X[8]	X[9]	X[9]	X[9]	X[9]		X[9]	X[9]	X[9]	X[9]		X		
Pregnancy test	X[10]	X[11]		X[11]			X[11]		X[11]			X[11]		
Contraceptive advice	X													
CT/MRI scan[12,13]	X[14]					X[12]					X[12]	X[15]		X[1]
Bone scan	X[14, 16]													
12-lead ECG[2]	X	X	X	X [17]	X[17]		X[17]	X[17]	X[17]	X[17]		X		
Document date and results of last tumor biopsy/fresh tumor biopsy	X													

	Screening (-21 to -1)	Active treatment phase (21 Days per cycle)										Post-treatment phase		
		Cycle 1		Cycle 2			Cycle 3		Subsequent Cycles			EOT (+7)	30-day safety FU (+7)	Long-term FU (±7)[1]
Assessment		D2	D9	D2	D9	D21	D2	D9	D2	D9	D21			
Blood for exploratory biomarkers in all patients[18]		X	X				X	X						
Confirmation of eligibility and randomization		X												
Eribulin treatment[19]		X	X	X	X		X	X	X	X				
Document reason for treatment discontinuation												X		
Survival status, PD monitoring, all anticancer medicines received														X
For patients at selected sites														
Optional tumor biopsy[20]	X				X							X		
Blood for other exploratory biomarkers		X			X[21]		X					X		

AE: adverse event; RI: magnetic resonance imaging; RECIST: Response Evaluation Criteria In Solid Tumors; SAE: serious adverse event.

Footnotes:

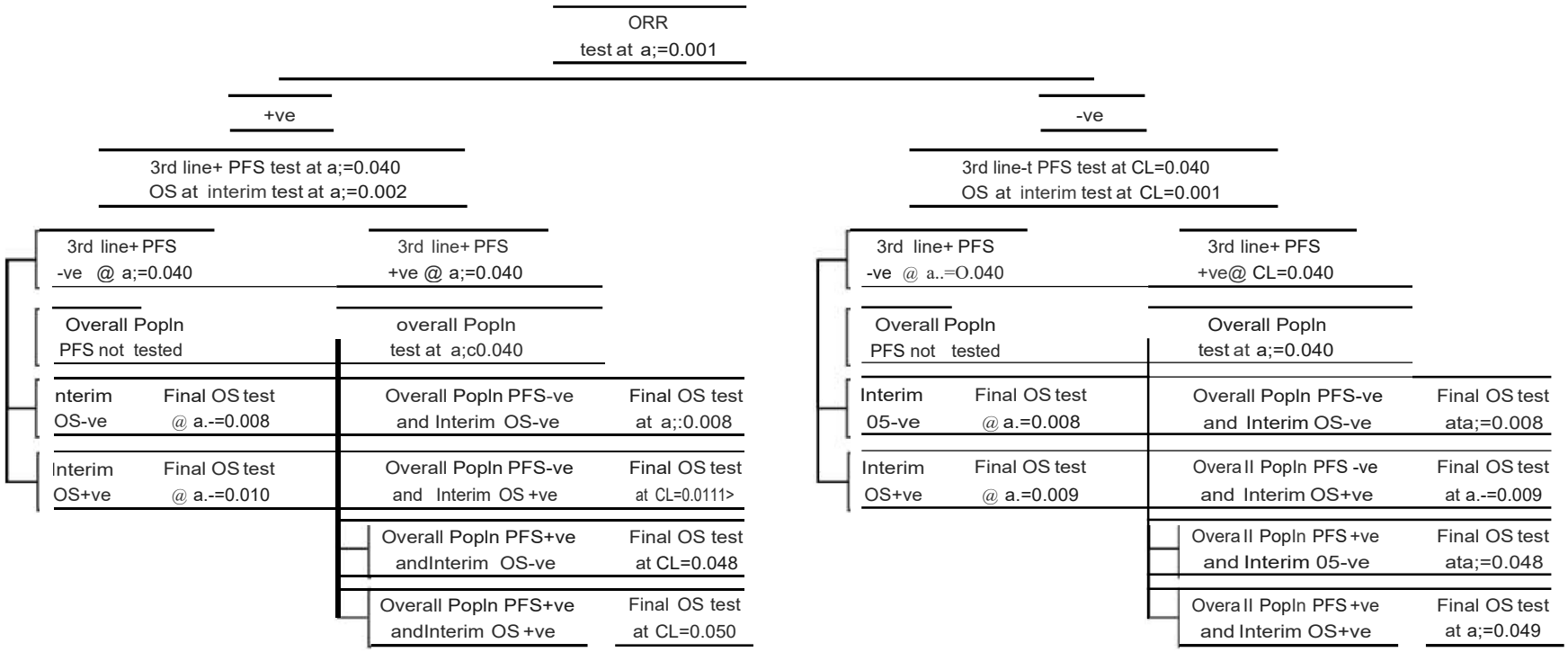
1. Regardless of treatment arm, patients discontinued from study medication for any reason, in the absence of PD, will undergo repeat imaging and tumor response assessments (including CT/MRI scans) every 8 weeks ±7 days (every 12 weeks ±7 days if the patient has been on the study for ≥1 year) until PD is documented as per RECIST v1.1, or death occurs, or the patient is lost to follow-up, or the patient withdraws consent (whichever occurs first). Once PD is recorded, the patient will be followed every 6 months (or more frequently) for Survival Follow-up only, until the patient dies, withdraws consent, or is lost to follow-up (whichever occurs first). Survival follow-up can be conducted by telephone.
2. Additional assessments to those scheduled for this study should be performed as indicated at the Investigator's discretion (e.g. due to concerns about patient safety, patient's clinical status); the findings of an unscheduled examination should be recorded in the patient's source data and eCRF.
3. Standard, complete physical examination.

4. Targeted physical examination.
5. Height will be measured at Screening only.
6. SAEs will be recorded from the date that the informed consent form is signed.
7. Concomitant therapy/medications taken since previous visit to be recorded for patients with unresolved AEs.
8. Screening blood tests should be taken within 7 days prior to randomization.
9. Blood test results must be available within 24 hours prior to study medication administration.
10. Serum pregnancy test for women of childbearing potential only, no more than 7 days prior to Randomization. More frequent assessments can be performed if medically indicated as determined by the Investigator, and these evaluations will be recorded in the patient's source data and eCRF.
11. Urine pregnancy test for women of childbearing potential only.
12. CT/MRI scan will be performed at Screening, then from the date of randomization every 6 weeks (±7 days) during the first year, and then every 12 weeks (±7 days), thereafter, until PD is documented by RECIST v1.1. It is critical that tumor assessments are performed according to the Study Schedule in a timely and complete manner regardless of any dosing delay or interruption. Imaging assessments will be scheduled using the randomization date as the reference date for all timepoints and NOT scheduled according to the date of the previous timepoint. A delay in the imaging assessment to accommodate a treatment delay is not permitted. Following initial documentation of an objective response, a confirmatory CT/MRI scan will be performed at least 4 weeks later.
13. At the discretion of the Investigator, additional radiographic tumor assessments may be done at any time if there is clinical suspicion of disease progression; the results of the unscheduled scan should be recorded in the patient's source data and eCRF.
14. Record date and results of last scan on the patient's source data and eCRF. If no scan has been performed in the 30 days prior to Randomization a scan should be performed during Screening.
15. CT/MRI scans will be performed in patients who are discontinued from study medication in the absence of PD; however, patients who have already demonstrated objective PD (according to RECIST v 1.1) do not need to have scans repeated at the EoT visit.
16. From the date of randomization, additional bone scans will be performed at the discretion of the Investigator.
17. After Cycle 2, the frequency of ECGs can be reduced to every 2 cycles if deemed clinically appropriate (i.e. absence of ECG findings in earlier assessments).
18. Blood will be collected from all patients to assess plasma levels of interferon gamma and additional cytokine markers.
19. Patients will start study medication within 3 days from randomization.
20. Patients at selected sites can optionally consent to fresh tumor biopsy at Screening and subsequent timepoints. If a biopsy has already been performed within 30 days prior to Randomization and tumor tissue is still available, then the fresh tumor biopsy is not required during Screening.

21. A blood sample will be taken only in patients at selected sites who are scheduled to have an optional, fresh tumor biopsy; the blood sample can be taken before or after the biopsy.

14.2. Type I Error Control for Primary and Key Secondary Efficacy Analyses

Figure 1 - Type I Error Control for Primary and Key Secondary Efficacy Analyses in the 3rd line+ patient population



14.3. Scoring for QLQ-C30 version 3.0

Table 4 - Scoring for QLQ-C30 version 3.0

	Scale	Number of items	Item range*	**Version 3.0** Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

14.4. Subscale Scoring for the FACT-B

Table 5 - Subscale Scoring for the FACT-B

Subscale	Item	Reverse?	Item response	Item Score	
PHYSICAL WELL BEING (PWB)	GP1	Yes	x	$4 - x$	
	GP2	Yes	x	$4 - x$	
	GP3	Yes	x	$4 - x$	
	GP4	Yes	x	$4 - x$	
	GP5	Yes	x	$4 - x$	
	GP6	Yes	x	$4 - x$	
	GP7	Yes	x	$4 - x$	
			Total	_____	
			×7	_____	
			Divide by number items answered	_____	= PWB subscale score
SOCIAL/FAMILY WELL BEING (SWB)	GS1	No	x	x	
	GS2	No	x	x	
	GS3	No	x	x	
	GS4	No	x	x	
	GS5	No	x	x	
	GS6	No	x	x	
	GS7	No	x	x	
			Total	_____	
			×7	_____	
			Divide by number items answered	_____	= SWB subscale score
EMOTIONAL WELL BEING (EWB)	GE1	Yes	x	$4 - x$	
	GE2	No	x	x	
	GE3	Yes	x	$4 - x$	
	GE4	Yes	x	$4 - x$	
	GE5	Yes	x	$4 - x$	
	GE6	Yes	x	$4 - x$	
			Total	_____	
			×6	_____	
			Divide by number items answered	_____	= EWB subscale score

Subscale Scoring for the FACT-B (cont.)

Subscale	Item	Reverse?	Item response	Item Score	
BREAST CANCER SUBSCALE (BCS)	B1	Yes	x	$4-x$	
	B2	Yes	x	$4-x$	
	B3	Yes	x	$4-x$	
	B4	No	x	x	
	B5	Yes	x	$4-x$	
	B6	Yes	x	$4-x$	
	B7	Yes	x	$4-x$	
	B8	Yes	x	$4-x$	
	B9	No	x	x	
	P2	Yes	x	$4-x$	
			Total	_____	
			×10	_____	
			Divide by number items answered	_____	=BC subscale score
FUNCTIONAL WELL-BEING (FWB)	GF1	No	x	x	
	GF2	No	x	x	
	GF3	No	x	x	
	GF4	No	x	x	
	GF5	No	x	x	
	GF6	No	x	x	
	GF7	No	x	x	
			Total	_____	
			×7	_____	
			Divide by number items answered	_____	= FWB subscale score

14.5. Algorithm for the best response determination according to RECIST v1.1 guidelines, as assessed by the local Investigator's review

a) Patients need to have two consecutive assessments of PR or CR to be a responder. PR or CR must be confirmed by 2 consecutive tumor evaluations spaced at least 4 weeks (28 days) apart.

b) For this study, up to three (3) intervening NEs will be allowed to confirm a CR or PR. Subsequent documentation of a CR may provide confirmation of a previously identified CR with up to three (3) intervening NEs (e.g., CR NE CR, CR NE NE CR, or CR NE NE NE CR). Similarly, subsequent documentation of a PR may provide confirmation of a previously identified PR with up to three (3) intervening NEs (e.g., PR NE PR, PR NE NE PR, or PR NE NE NE PR). However, only one (1) intervening SD will be allowed between PRs for confirmation (i.e., PR SD PR). The following table, based on Appendix from the RECIST 1.1 guidelines, summarizes the algorithm describing how Confirmed Response is determined from the overall tumor assessments as assessed by the local Investigator's

review. The order used to determine the First Time Point response is CR>PR>SD>PD, ignoring visits with missing tumor assessments.

c) The Confirmed Response is determined based on all overall tumor assessments.

d) Determine the first time point response as the first assessment of the non-confirmed best response (CR > PR > SD > PD) for a patient, then take the first occurrence of the non-confirmed best response and look at the subsequent assessments within the considered period (Results based upon scans performed after start of new anti-cancer therapy will be excluded for analysis derivations) using the following rules.

Case	**Overall response entered in eCRF by investigator**		**Confirmed Response (Best Response)***
	First Time Point Response (TPR)**	**Subsequent Time Point Response (TPR)**	
1	CR	CR	CR (if assessments at least 4 weeks (28 days) apart). (note: sequences of CR NE CR, CR NE NE CR, or CR NE NE NE CR would be considered as confirmed CR)
2	CR	PR	SD, PD or PR If CR truly met at first timepoint, any subsequent assessment of PR should make the disease PD at that point. That is, neither a PR nor SD may follow CR Therefore, SD, if CR assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization (i.e. date of assessment ~~d~~ate of randomiz ation $\geq$ 35 days), otherwise PD. However, confirmed response may be PR if subsequent scans suggest small lesions were still present at first assessment (in which case first assessment of CR should be changed to PR) * (if assessments at least 4 weeks (28 days) apart).
3	CR	SD	SD or PD SD, if CR or SD assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization, otherwise PD
4	CR	PD	SD or PD SD, if CR assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization, otherwise PD

Case	Overall response entered in eCRF by investigator		Confirmed Response (Best Response)*
	First Time Point Response (TPR)**	Subsequent Time Point Response (TPR)	
5	CR	NE**	SD or NE SD, if CR assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization otherwise NE
6	PR	CR	PR (if assessments at least 4 weeks (28 days) apart).
7	PR	PR	PR (if assessments at least 4 weeks (28 days) apart).
8	PR	SD (3) **	SD
9	PR	PD	SD or PD SD, if PR assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization, otherwise PD
10	PR	NE**	SD or NE SD, if PR assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization, otherwise NE
11	SD	SD	SD
12	SD	PD	SD or PD SD, if SD assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization, otherwise PD
13	SD	NE	SD or NE SD, if SD assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization, otherwise NE
14	CR, PR, SD	-	SD or NE SD, if assessment >=6 weeks (+/-7 days), i.e. 35 days after date of randomization and does not qualify for CR or PR, otherwise NE.
15	PD	No further	PD. Ignore all assessments after initial overall response of PD.

Case	Overall response entered in eCRF by investigator		Confirmed Response (Best Response)*
	First Time Point Response (TPR)**	Subsequent Time Point Response (TPR)	
16	NE	NE	NE Where all assessments are Not evaluable
17	NE	-	NE

* A Best Response of SD can only be made after the subject is on-study for a minimum of six (6) weeks (42 days) +/- seven (7) days, for a minimum time on-study of thirty-five (35) days. If the subject is on-study less than six (6) weeks (42 days) +/- seven (7) days or thirty-five (35) days, any tumor assessment indicating Stable Disease before this time period will have a Best Response of NE unless PD is identified.

** For this study, up to three (3) intervening NEs will be allowed between CRs or PRs but only one (1) intervening SD is allowed between PRs. Subsequent documentation of CR may provide confirmation of a previously identified CR for subjects where the intervening integrated responses, up to three (3), were NE. Subsequent documentation of PR may provide confirmation of a previously identified PR for subjects where the intervening integrated responses, up to three (3), were NE. If the post-NE TPR confirms the CR (or PR) then the Confirmed Response will be CR (or PR). For example, CR NE CR = confirmed CR, CR NE NE CR = confirmed CR, or CR NE NE NE CR = confirmed CR ; PR NE PR = confirmed PR, PR NE NE PR = confirmed PR, or PR NE NE NE PR = confirmed PR ; CR SD CR ≠ confirmed CR ; PR SD PR = confirmed PR. Note: in the following scenario, PR SD NE PR, the second PR is not a confirmed PR.

(1) TPR is SD if the increase from the first to the second assessment does not qualify for PD.

The confirmed response (best response) is defined when all Overall response assessments have been reviewed.