STATISTICAL ANALYSIS PLAN

Calithera Biosciences, Inc.

CX-839-007

Protocol Title:	A Multicenter Phase 2 Study of the Glutaminase Inhibitor CB-839 in Combination with Paclitaxel in Patients with Advanced Triple Negative Breast Cancer (TNBC) Including Patients of African Ancestry and Non-African Ancestry
Protocol Version and Date:	Original / 06 Jan 2017
Sponsor:	Calithera Biosciences, Inc. 343 Oyster Point Blvd, Suite 200 South San Francisco, CA 94080
Prepared By:	[redacted] Biostatistician Contracted to Calithera 343 Oyster Point Blvd, Suite 200 South San Francisco, CA 94080
Document Version and Date:	Draft Version 1.0; 8 Oct 2019

1 STATISTICAL ANALYSIS PLAN APPROVAL

Sponsor:	Calithera Biosciences, Inc.
Clinical Protocol Number:	CX-839-007
Protocol Title:	A Multicenter Phase 2 Study of the Glutaminase Inhibitor CB-839 in Combination with Paclitaxel in Patients with Advanced Triple Negative Breast Cancer (TNBC) Including Patients of African Ancestry and Non-African Ancestry
Document File Name:	Calithera_CX-839-007_SAP_v1.0_08OCT2019
Document Version and Effective Date:	Version 1.0; 08 Oct 2019

Approved By:

[redacted]

Director of Biometrics
Calithera Biosciences, Inc.

Date

[redacted]

Biostatistician II
[redacted] and Rare Disease

Date

[redacted]

Manager, Biostatistics
Precision for Medicine, Oncology and Rare Disease

Date

2 TABLE OF CONTENTS

Contents

TABLE OF TABLES

3 LIST OF ABBREVIATIONS

Abbreviation	Definition
AA	African Ancestry
AE	Adverse event
BID	Twice daily
CBR	Clinical Benefit Rate
CI	Confidence interval
CR	Complete response
CTCAE	Common Terminology Criteria for Adverse Events
DOR	Duration of Response
ECG	Electrocardiogram
EOT	End of Treatment
FDA	Food and Drug Administration
hr	Hour or hours
IV	Intravenous, intravenously
MedDRA	Medical Dictionary for Drug Regulatory Activities
mL	Milliliter
NCI	National Cancer Institute
ORR	Overall response rate
OS	Overall Survival
PD	Progressive Disease
PFS	Progression Free Survival
PK	Pharmacokinetic(s)
PR	Partial response
QTcF	Corrected QT interval, Fridericia's formula
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Stable disease
TEAE	Treatment-emergent adverse event
TNBC	Triple negative breast cancer
TTR	Time to recurrence

4 INTRODUCTION

The statistical analysis plan (SAP) provides details of the planned analyses and statistical methods for the study CX-839-007 (A Multicenter Phase 2 Study of the Glutaminase Inhibitor CB-839 in Combination with Paclitaxel in Patients with Advanced Triple Negative Breast Cancer (TNBC) Including Patients of African Ancestry and Non-African Ancestry). The background and rationale for the study can be found in the study protocol.

5 STUDY DESIGN

Protocol CX-839-007 is a Phase 2 open-label study of the combination of CB-839 with paclitaxel. Multiple single-arm cohorts will be enrolled in which 800 mg twice daily (BID) CB-839 will be administered in combination with the full approved dose of paclitaxel. Figure 1 illustrates the study design.

Patients eligible after completing all screening evaluations will be enrolled into one of four cohorts:

Cohort 1 – African ancestry with 3rd line+ Metastatic

a. Patients must self-identify as African ancestry (AA; includes African American).
b. At least 2 prior lines of systemic therapy for advanced/metastatic disease including a taxane.
   i. Prior taxane (paclitaxel, docetaxel, or nab-paclitaxel) for advanced/metastatic disease is required but must not have been received in the immediate prior line of therapy.
   ii. Systemic neoadjuvant and/or adjuvant therapy is considered a line of therapy for advanced/metastatic disease if the time to recurrence from completion of treatment was ≤ 12 months.

Cohort 2 – African ancestry 1st line Metastatic

a. Patients must self-identify as African ancestry (includes African American).
b. No prior systemic therapy for advanced or metastatic disease.
   i. Systemic neoadjuvant or adjuvant therapy, including taxane, is allowed if time to recurrence was > 12 mo.

Cohort 3 – Non-African ancestry 3rd line+ Metastatic

a. Patients do not self-identify as African ancestry.
b. Otherwise have the same criteria as Cohort 1.

Cohort 4 – Non-African ancestry 1st line Metastatic

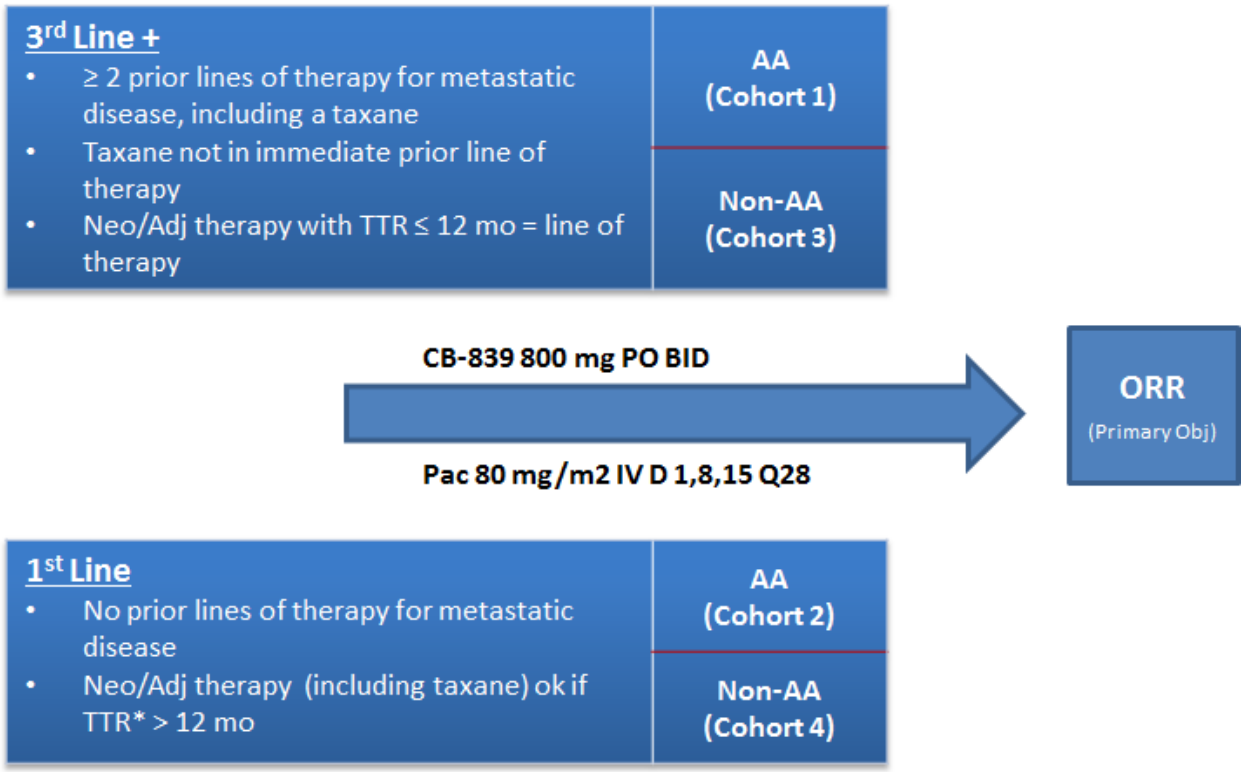
a. Patients do not self-identify as African ancestry.
b. Otherwise have the same criteria as Cohort 2.

Treatment administrations are per the instructions and schedule of events outlined in the protocol.

CB-839 will be administered orally 800 mg BID with food, on days 1 through 28 of each 28-day cycle. Paclitaxel (80 mg/m$^{2)}$ will be administered as an IV infusion over 1- 3 hr on Days 1, 8, and 15 of each 28-day cycle. Patients will receive the treatment in 28-day cycles until disease progression or unacceptable toxicity (whichever occurs first). Patients will be followed within 28

days of last dose date or early discontinuation of treatment for safety follow up. Patients who discontinue study treatment for reasons other than disease progression or death will remain in follow up including protocol-defined imaging. Long-term follow up for survival will continue until death or withdrawal of consent to follow up.

Figure 1: Study Design Schema



* TTR = Time to recurrence

5.1 Protocol Synopsis

Please refer to protocol synopsis in the original protocol dated 06 Jan 2017 (13). The Schedule of Assessments is in Appendix A: Schedule of Study Assessments.

5.2 Study Endpoints

5.2.1 Efficacy Endpoints

5.2.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the overall response rate (ORR) of patients treated with CB-839 plus paclitaxel (Pac-CB) for metastatic TNBC. Overall Response Rate is defined as the percentage of patients with complete response (CR) or partial response (PR) per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1. To be assigned a status of CR or PR, changes in tumor measurements must be confirmed by repeat assessment performed no less than 4 weeks after the criteria for response were first met.

5.2.1.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are as follows:

- Progression free survival (PFS) of patients treated with CB-839 plus paclitaxel (Pac-CB) for metastatic TNBC. PFS is defined as time from first dose date to the earlier of either progression of disease per RECIST v1.1 or death from any cause.
- Overall survival (OS) of patients treated with CB-839 plus paclitaxel (Pac-CB) for metastatic TNBC. Overall Survival is assessed by time from first dose date to death due to any cause.
- Duration of response (DOR) of patients treated with CB-839 plus paclitaxel (Pac-CB) for metastatic TNBC. DOR is defined as the time between the first documentation of a confirmed PR or a CR per RECIST v1.1 to the first documentation of PD or death.
- Clinical benefit rate (CBR) of patients treated with CB-839 plus paclitaxel (Pac-CB) for metastatic TNBC. Clinical Benefit Rate is defined as the percentage of patients with best response of CR, PR, or SD per RECIST v1.1 criteria. SD is defined as stable disease lasting ≥ 16 weeks for 3rd line + patients and ≥ 24 weeks for 1st line patients.

5.2.2 Safety Endpoints

The safety endpoints are:

- Type, incidence, seriousness, nature, severity, and drug-relatedness of adverse events graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0 (NCI CTCAE v4.0).

- Changes in vital signs, physical findings, and clinical laboratory test results following study treatment administration.

5.3 Determination of Sample Size

Approximately 112 response evaluable patients are expected in total to be enrolled and treated across four cohorts. If a patient does not meet criteria for inclusion into the Efficacy Analysis Set, the patient will be considered to be not evaluable for efficacy and will be replaced.

Cohorts 1 and 3: We assume that the ORR from single agent paclitaxel will be < 10% in this 3rd line+ patient population. With a sample size of 28 patients this design has a type I error rate of 0.14 if the true response rate is 10% and a power of 0.86 if the true response rate is 25%.

Cohorts 2 and 4: Simon's two-stage design (Simon, 1989) will be used. The null hypothesis that the true response rate is < 25% will be tested against the alternative hypothesis of ≥ 45% response rate. In the first stage, 15 patients will be accrued to each cohort. If there are 3 or fewer responses in these 15 patients, accrual to this cohort will be stopped. Otherwise, 13 additional patients will be accrued for a total of 28 patients. The null hypothesis will be rejected if 10 or more responses are observed in 28 patients (ORR ≥ 36%). This design has a type I error rate of 0.13 when the true response rate is 25% and power of 0.87 when the true response rate is 45%.

6 STUDY CONDUCT

6.1 Safety Data Monitoring

Safety data from the study is monitored on an ongoing basis via routine PrimeVigilance activities. In addition to real time medical review of emergent SAEs, a cross functional sponsor safety review team performs regular periodic aggregate data reviews across all CB-839 studies.

7 STATISTICAL METHODS

7.1 Analysis Sets

7.1.1 Efficacy Evaluable Population for ORR and CBR – Response Evaluable Analysis Set

All patients who have measurable disease at baseline, receive at least one dose of study drug (CB-839 or paclitaxel), and complete at least one post-baseline tumor assessment, will be considered evaluable for efficacy. In addition, patients who discontinue treatment for study-drug related toxicity or for disease-related death also are included in the efficacy evaluable population for ORR and CBR.

All response efficacy analyses will be performed by cohort.

7.1.2 Efficacy Evaluable Population for PFS – PFS Evaluable Analysis Set

All patients who receive at least one dose of any study-specific treatment (CB-839 or paclitaxel) will be included in the PFS evaluable analysis set. The PFS analysis will be performed by cohort.

7.1.3 Safety Analysis Set

All patients who receive at least 1 dose of any study-specific treatment (CB-839 or paclitaxel) will be included in the analysis of safety. All safety analyses will be performed by cohort.

7.2 Analysis of Study Conduct

Subject disposition, including enrollment, analysis populations, major protocol deviations (including major deviations of inclusion and/or exclusion criteria), reason for discontinuation from the study, and on-study status will be summarized by cohort and overall for the safety analysis set.

7.3 Analysis of Treatment Group Comparability

Demographic and baseline characteristics, including age, sex, race, ethnicity, and disease characteristics will be summarized by cohort and overall for the safety analysis set.

Baseline values are defined as the last available data obtained prior to the patient receiving the first dose of any study treatments on Cycle 1 Day 1 visits unless otherwise noted.

7.4 Efficacy Analysis

Efficacy analyses will be performed by cohort. Response to treatment will be evaluated using RECIST v1.1.

The four cohorts are based on patient's demographics and treatment history instead of randomization, therefore formal hypothesis testing among cohorts will not be performed.

7.4.1 Primary Analysis of Overall Response Rate

Overall Response Rate (ORR) will be assessed by cohort where ORR is defined as the percentage of patients with complete response (CR) or partial response (PR) per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1. To be assigned a status of CR or PR, changes in tumor measurements must be confirmed by repeat assessment performed no less than 4 weeks after the criteria for response were first met. Patients without assessed response will be treated as non-responders.

The analysis set for ORR will be the response evaluable analysis set. An estimate of ORR and its 95% confidence interval (CI) will be calculated with the Clopper-Pearson method. Best of Response (BOR) Rate is defined as the percentage of patients with their best of response being CR, PR, SD or PD, and will be estimated in the same approach as for ORR.

In Cohorts 1 and 3 respectively, the observed ORR will be compared to 10%. If stage I and II are both completed, in Cohorts 2 and 4 respectively, the observed ORR will be compared to 25%. One-sample binomial tests will be performed in the comparisons. The resultant p-values are descriptive in nature.

7.4.2 Secondary Analyses

Progression Free Survival: PFS is defined as time from the first dose date to the earlier of either progression of disease per RECIST 1.1 or death from any cause. If the disease progression assessment involves more than one date, the earliest date will be used as the event date.

The duration of PFS will be censored at the date of the last radiographic disease assessment if:

- Patient is alive and progression free at the time of analysis data cutoff.
- Disease progression or death occurs after missing data (including a non-evaluable status for overall response assessment) for two consecutive radiographic disease assessments.
- Patient receives non-protocol TNBC treatment prior to documentation of disease progression.

Patients missing baseline disease assessment will be censored at the first dose date (day 1). Patients who come off of study for reason other than PD or death should continue to be followed with radiographic assessment until PD by RECIST 1.1, death, withdrawal of consent, or initiation of another systemic anti-cancer treatment. Censoring rules are summarized in Table 2.

The analysis set for PFS will be the PFS evaluable analysis set. Kaplan-Meier methodology will be used to estimate median PFS for each cohort and construct survival curves for visual description. The Brookmeyer-Crowley methodology will be used to construct the 95% CI for the median PFS for each cohort (2).

Overall Survival (OS): OS is defined as the time from the first dose date to death due to any cause. For patients alive at time of analysis, OS will be censored at the time when the patient is last known to be alive. Patients who discontinue study for reason other than death should continue to be followed for survival status until death or withdrawal of consent. Because OS may require longer follow up than the primary PFS analysis, additional OS analyses may be performed after the time of PFS primary analysis. OS will be analyzed with the same methodologies and analysis set as PFS.

Duration of Response (DOR): For patients achieving a PR or a CR, the duration of response will be calculated as the time between the first documentation of a PR or a CR to the first documentation of PD or death, whichever occurs first, taking as reference for PD the smallest measurements recorded on study (including baseline, i.e. radiographic assessment at screening). For patients achieving first a PR then a CR, the PR date will be the starting date for response duration calculation. Patients who are alive and have not experienced disease progression at the time of the analysis will follow the censoring rules in Table 2 where only assessments after first response will be included. The determination of DOR is based on responders only. DOR will be analyzed with the same methodologies as PFS using response evaluable analysis set.

DOR will be reported for patients with confirmed OR.

Table 1: Primary Analysis: The Censor/Event Rules for Progression Free Survival and Duration of Response[a]

Situation	Date of Event or Censoring	Outcome
No baseline disease assessment	Date of first dose	Censored
No post-baseline assessments and no death	Date of first dose	Censored
No progression and no death	Date of last evaluable tumor assessment	Censored
Additional cancer therapy prior to documentation of disease progression or death	Date of last evaluable tumor assessment prior to first new cancer therapy	Censored
Documented RECIST progression per investigator or death within 2 scheduled scan intervals following previous evaluable radiological tumor assessment	First date of evaluation of overall response of PD or death is determined	Event
RECIST progression or death documented to occur after missing 2 scheduled disease assessments (including an overall response of non-evaluable) following previous evaluable radiological tumor assessment	Date of last evaluable tumor assessment with no progression prior to the first of these missed visits	Censored

PD = progressive disease; RECIST = Response Evaluation Criteria in Solid Tumors
[a]RECIST progression or death can occur either on study or during the survival follow up period after treatment discontinuation for symptomatic deterioration, adverse event, or other reason not related to disease and prior to the initiation of new cancer therapy.

7.3 Safety

Unless specified otherwise, safety analyses described in this section will be conducted for the Safety analysis set.

7.3.1 Treatment Exposure

Extent of exposure to both study treatments (CB-839 and paclitaxel), including treatment duration, total dose received (mg for CB-839 and mg/m^2 for paclitaxel), number of cycles, will be evaluated by summary statistics (N, mean, standard deviation, median, minimum and maximum). Percent of patients and cycles with dose delays and reductions will be calculated.

Exposures will be summarized by cohort for CB-839 and paclitaxel separately.

7.3.2 Adverse Events

Verbatim description of adverse events will be coded to System Organ Class (SOC) and Preferred Term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA) and graded by the investigator according to NCI CTCAE v4.03. Treatment-emergent AEs (TEAEs) are defined as AEs started or worsened on or after first dose and will be tabulated. The number and proportion of patients reporting a given TEAE will be tabulated according to the worst severity grade reported. Separate tables will be constructed for a) all reported TEAE's, b) serious TEAEs (SAEs), c) severe TEAEs (Grade $\geq$ 3), and d) TEAE's leading to permanent discontinuation or interruption of study treatments. The above tables will also be presented for TEAEs judged to be related to either study treatment.

Multiple occurrences of the same event will be counted once at the maximum grade. In order to accurately summarize the true TEAE rate, all TEAE summaries will not count grade 5 events for patients who died due to progressive disease.

All listings of adverse events will include both TEAEs and non-TEAEs.

Deaths reported during the study treatment period and the follow up period after treatment completion/discontinuation will be listed and summarized by cause.

7.3.3 Laboratory Data

Laboratory variables will be examined using mean change in value from baseline to scheduled time points. The baseline value of a variable is defined as the last value obtained on or before the date and time of the first dose of CB-839 or paclitaxel. Proportion of patients with laboratory measurement outside the normal range and by NCI CTCAE v4.03 grade will be summarized.

7.3.4 ECG and Vital Signs

Electrocardiogram (ECG), weight, and vital signs will be summarized by changes from baseline to scheduled time points using descriptive statistics. Baseline is defined the same way as for laboratory measurements.

7.4 Missing Data

Patients with missing response assessment will be treated as non-responders in analyzing ORR and CBR. Tables 2 will be followed to censor these patients in analyzing PFS and DOR. Missing laboratory, ECG and vital sign measurements will not be imputed and included in the analysis.

AEs with missing or partial start date that cannot be definitely determined to be earlier than first dose will be treated as TEAE. AEs with missing relationship to study treatments will be treated as related.

8 REFERENCES

1. Protocol CX-839-007, dated 16 Jan 2017

2. Brookmeyer, R. and Crowley, J. (1982) A confidence interval for the median survival time. Biometrics, 38, 29-41. doi:10.2307/2530286

3. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. Statistical principles for clinical trials (E9). International Conference on Harmonization; 1998.

4. Data Standards: Position on Use of SI Units for Lab Tests. U.S Food and Drug Administration; 25 October 2013. Available from: http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm37 2553.htm

5. U.S. Department of Health and Human Services. Common Terminology Criteria for Adverse Events (CTCAE), Version 4.03. 14 June 2010. Available from: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf.

APPENDIX A: SCHEDULE OF STUDY ASSESSMENTS

Visit	Screening	Cycle 1		Cycle 2+[17]	End of Treatment/ Follow up
	Day -28 to -1	Day 1 (-1 day)	Days 8 and 15 (± 2 days)	Day 1 (± 5 days)	EOT: Within 28 days post treatment discontinuation
Written Informed Consent	X[1]				
Inclusion/Exclusion Criteria	X				
Demographics and Medical History	X				
Physical Examination[2]	X	X	X	X	X
Height	X				
Weight	X	X	X	X	X
Vital Signs[3]	X	X	X	X	X
ECOG Performance Status	X	X		X	X
Duplicate 12-lead ECG with QTcF	X	X[4]			X
Urinalysis[5]	X				
Coagulation tests[5]	X				
Hematology	X	X[6]	X	X	X
Serum Chemistry levels	X	X[6]	X	X	X
Serum or Urine Pregnancy Test[7]	X				X
Pharmacokinetic (PK) Assay[8]			X	X[8]	
Whole blood & Saliva for biomarker analysis[9]	X				
Archival Tumor Collection[10]	X				
Pre-dose Tumor Biopsy[11]	X				
Optional Tumor Biopsies[12]				X	X
Radiographic Evaluation of Tumor Burden (diagnostic CT or MRI)	X[13]			X[14]	X
Neuropathy Assessment	X			X[15]	X
CB-839 Dosing		CB-839 will be administered twice daily (BID) with food every day of every cycle.			
Paclitaxel Dosing		Paclitaxel is administered on Days 1, 8, and 15 of every 28 day cycle.			
Adverse Events		X	X	X	X
Concomitant Medications	X	X	X	X	X
Follow up					X[16]

Explanation of Superscripts:

1. Informed consent must completed prior to any study-related screening procedures and may be completed before the 28-day screening window.
2. Complete physical exam is required at Screening and at End of Treatment. A symptom-directed physical exam can be done on all other visits. System exams are only required as clinically indicated.
3. Vital sign measurements include temperature, pulse, respiratory rate and resting systolic and diastolic blood pressure.
4. On C1D1, duplicate ECGs to be performed 2-4 hr post CB-839 dose
5. Assessments completed at screening. Investigators should monitor during the study if it is deemed necessary.
6. Does not need to be repeated if the Screening sample was obtained within 3 days prior to C1D1 unless a clinically significant change is suspected
7. Serum or urine pregnancy test is required of all patients of child-bearing potential. Screen pregnancy test must occur within 3 days prior to C1D1.
8. PK sample (3 mL of whole blood) will be collected predose and 4 hours (± 30 min) post-dose on C1D15, C2D1, and C3D1. Refer to laboratory manual for further instructions.
9. Biomarker analysis samples (3 mL of whole blood and saliva) will be taken at screening. Refer to laboratory manual for further instructions.
10. Archival tumor tissue must be submitted, if available, for correlative studies.
11. Fresh pre-dose tumor biopsies will be collected from all patients UNLESS an archival sample collected within 3 months prior to C1D1 is provided or if the tumor is inaccessible.
12. Optional fresh tumor biopsies may be collected from patients who consent to providing samples during the trial or at EOT (e.g., if patient is responding, progressing lesion, etc.).
13. Tumor assessments for Screening must be done within 21 days prior to C1D1. Whenever possible, imaging should be done at the same institution/facility and with the same modality which will be used to measure response during the patient's participation in the study.
14. Completed approximately every 8 weeks (2 cycles) during the first 13 cycles per RECIST 1.1. For patients with ≥ SD for at least 13 cycles who are on a steady dose for ≥ 2 cycles, radiographic evaluation of tumor burden may be reduced to every 3 cycles (i.e., at Cycles 16, 19, 22, etc.). Evaluations may occur more frequently as clinically indicated. Scans must be submitted for central review.
15. Neuropathy assessment will be performed at Screening, C3D1, C6D1, and at EOT. Refer to Study Reference Manual for detailed information.
16. Patients will be contacted every 3 months for the first 12 months after discontinuation and then every 6 months thereafter to confirm survival.
17. For patients with ≥ SD for at least 13 cycles who are on a steady dose for ≥ 2 cycles, study assessments may be reduced to every 3 cycles (i.e., at Cycles 16, 19, 22, etc.).