Statistical Analysis Plan

Protocol CDX011-04

A Randomized Multicenter Pivotal Study of CDX-011 (CR011-vcMMAE)
in Patients with Metastatic, GPNMB Over-Expressing,
Triple Negative Breast Cancer
(The "METRIC" Study)

Version: Final v1.1

Date: August 31, 2017

Glossary of Abbreviations

Abbreviation or Term	Definition
ADC	Antibody-Drug Conjugate
ADI	Actual Dose Intensity
AE	Adverse Event
AR	Autoregressive
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BSA	Body Surface Area
CR	Complete Response
CRF	Case Report Form
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DOR	Duration of Response
DMC	Data Monitoring Committee
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOT	End of Treatment
ER	Estrogen Receptor
GPNMB	Glycoprotein NMB
HRQOL	Health-Related Quality of Life
IDMC	Independent Data Monitoring Committee
IHC	Immunohistochemistry
IRC	Independent Review Committee
ITT	Intention-to-Treat
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MNAR	Missing Not at Random
MRI	Magnetic Resonance Imaging
NCI	National Cancer Institute (of the United States)
ORR	Objective Response Rate
OS	Overall Survival
PD	Pharmacodynamics or Progressive Disease
PDI	Planned Dose Intensity
PFI	Progression-Free Interval
PFS	Progression-Free Survival
PK	Pharmacokinetics

Abbreviation or Term	Definition
PR	Partial Response or Progesterone Receptor
PS	Performance Status
PT	Preferred Term
RECIST	Response Evaluation Criteria for Solid Tumors
RS	Raw Score
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SOC	System Organ Class
TA	Total Antibody or Tumor Assessment
TEAE	Treatment-Emergent Adverse Event
TNBC	Triple Negative Breast Cancer
WHO	World Health Organization

Table of Contents

List of Tables

1 INTRODUCTION

This statistical analysis plan (SAP) describes the planned data analysis specifications for Protocol CDX011-04, "A Randomized Multicenter Pivotal Study of CDX-011 (CR011-vcMMAE) in Patients with Metastatic, GPNMB Over-Expressing, Triple Negative Breast Cancer (The "METRIC" Study)" for the completion of the Clinical Study Report (CSR).

Any changes made to the planned analyses after this document has been finalized will be noted in the clinical study report.

1.1 Overview

This is a pivotal, open-label, prospectively controlled, randomized study of CDX-011 in patients with metastatic GPNMB-expressing triple-negative breast cancer. Eligible patients will be randomized (2:1) to receive CDX-011 or capecitabine.

Study treatment, and associated study visits at 3 week intervals, will continue until disease progression or intolerance. Tumor assessments will be performed at six week intervals for six months and nine week intervals thereafter, until disease progression. Upon discontinuation of study treatment, patients will be permitted to receive appropriate alternate anti-cancer therapies and will be followed for progression-free survival (if progression is not yet documented) and overall survival.

The study analyses of progression free survival (PFS), objective response rate (ORR), and duration of response (DOR) will be based on tumor response assessments performed by an independent review committee according to RECIST 1.1 guidelines (Eisenhauer, Therasse et al. 2009). The investigator's assessment of tumor response (according to RECIST 1.1) will guide clinical management and eligibility for continuation of study treatment.

An independent data monitoring committee (IDMC) will meet approximately every 6 months or as frequently as deemed necessary by the IDMC, starting from the first randomized patient, to review adverse events, serious adverse events, laboratory values, vital signs, mortality, and other safety-related data. The details of the DMC procedure and statistical deliverables are specified in the DMC Charter.

1.2 Schedule of Assessments

The schedule of assessments is described in Table 1 below.

Table 1. Schedule of Assessments

Visit	Screening[2]	Treatment Visits[3]							Disease Assessment Visit[4]	Survival Assessment[5]
		Cycle 1/ Day 1	Cycle 1/ Day 7	Cycle 1/ Day 14	Cycle 2/ Day 1	Cycles 3, 5, 7, etc. ("odd cycles") / Day 1	Cycles 4, 6, 8, etc. ("even cycles")/ Day 1	End of Treatment[6]		
Visit window[1]	*Day -28 to Day -1*		*+/-1 day*	*+/-1 day*	*+/-3 days*	*+/-5 days*	*+/-5 days*	*Within 28 days post-dosing*	*Every 6 (±1) weeks for 6 months and every 9 (±2) weeks thereafter, until progression.*	*Every 12 (±2) weeks until study closure*
Informed Consent and, if applicable, HIPAA	X									
Tumor tissue[8]	X[9]			X[10]						
Determination of TNBC status[11]										
Medical history[12]	X	X								
Physical examination[13]	X	X[7]			X	X	X	X		
Vital signs[14]	X	X[7]			X	X	X	X		
ECOG performance status	X	X[7]			X	X	X	X		
Electrocardiogram (ECG) [15]	X[15]	X[7,15]						X[15]		
Pregnancy test	X[16]	X[7,16]								
Hematology[17]	X	X[7]	X	X	X	X	X	X		
Blood chemistry[17]	X	X[7]			X	X	X	X		
Urinalysis[17]	X	X[7]				X		X		
Immunogenicity[18]		X[19]				X[19]		X[19]		
Pharmacokinetics[18]		X[20]			X[20]	X[20]	X[20]	X[20]		
PBMC collection[18]		X[21]			X[21]			X		
Molecular profiling[22]		X[22]								

Visit	Screening[2]	Treatment Visits[3]							Disease Assessment Visit[4]	Survival Assessment[5]
		Cycle 1/ Day 1	Cycle 1/ Day 7	Cycle 1/ Day 14	Cycle 2/ Day 1	Cycles 3, 5, 7, etc. ("odd cycles") / Day 1	Cycles 4, 6, 8, etc. ("even cycles")/ Day 1	End of Treatment[6]		
Visit window[1]	*Day -28 to Day -1*		*+/-1 day*	*+/-1 day*	*+/-3 days*	*+/-5 days*	*+/-5 days*	*Within 28 days post-dosing*	*Every 6 (±1) weeks for 6 months and every 9 (±2) weeks thereafter, until progression.*	*Every 12 (±2) weeks until study closure*
Questionnaire (EORTC QLQ-C30)[23]	X	X[7]				X		X		
Disease assessment[24]	X								X[24]	
Randomization[25]	X									
Administration of study treatment[26]		X			X	X	X			
Survival status										X
Concomitant medication review[27]	X	X			X	X	X	X	X[27]	X[27]
Adverse event monitoring[28]		X	X[29]	X[29]	X	X	X	X	X[30]	X[30]

1. A delay in study treatment or performance of study visits due to holidays, weekends, inclement weather or other unforeseen circumstances will be permitted and not considered a protocol violation. However, significant delays (i.e., greater than one week) should be discussed with the study medical monitor to reach consensus on subsequent scheduling.
2. No study procedures will be performed prior to receipt of signed Informed Consent. However, assessments performed according to standard of care prior to receipt of Informed Consent may be utilized to fulfill the screening requirement, if completed within the required window for screening.
3. Patients will receive CDX-011 or capecitabine (on a three-week cycle) until intolerance or progression of disease.
4. Disease assessments will be performed every six weeks (+/- 1 week) for six months (i.e., weeks 6, 12, 18, 24), every 9 weeks (+/- 2 weeks) thereafter (i.e., weeks 33, 42, 51, 60, 69, 78, etc.), until progression. Patients who discontinue study treatment without documented progression of disease as per RECIST 1.1 should continue to have Disease Assessment Visits until such criteria are met (regardless of intervening therapy). In the event that follow-up is discontinued without a final disease assessment documenting objective disease progression, the reason must be supplied; in particular, any admittance to hospice or other such similar end-of-life care must be reported. If a partial or complete response is noted, a follow-up radiographic assessment must be done no sooner than 28 days later to confirm response. If surgical intervention or localized radiation become indicated (either for palliation or down-staging of previously non-resectable tumor), these interventions should be avoided if clinically feasible until after the 12 week response assessment. Prior to any intervention (such as surgical resection, palliative radiation or alternate anti-

cancer therapy), every effort should be made to perform a tumor response assessment in order to document progression and/or confirm an objective response. Patients who undergo surgical resection or radiation in the absence of progression may continue to receive study treatment until remaining lesions meet criteria for progression of disease.

5. Subsequent to the End of Treatment Visit, all patients will be followed at 12 (±2) week intervals until study closure. These visits may be performed by telephone. However, for patients who discontinue study treatment in the absence of progression, survival assessments may be combined with Disease Assessment Visits, and conducted every 6 (±1) or 9 (±2) weeks until progression, and then at 12 (±2) week intervals until study closure.
6. The End of Treatment Visit should be performed within 28 days after last dose of study treatment and prior to initiation of alternate therapies.
7. Assessments do not need to be repeated if completed within the previous 24 hours as part of the screening assessment.
8. Assessment of GPNMB expression (by IHC) will be performed at a central laboratory. Additional analyses to be performed centrally may also include GPNMB expression by RT-PCR, examination of tumor markers using IHC or other molecular analyses, evaluation of tumor infiltrating leukocyte populations, biomarkers related to immune activation, and localization of CDX-011, CR011, or MMAE at the tumor site. Sample collection, processing and shipping instructions will be provided separately.
9. Tumor specimen(s) submitted for screening must have been obtained in the setting of advanced disease. Submission of additional samples from other collection dates, when available, is encouraged. Tissue may be submitted and tested at any time prior to or during the 28-day window for screening, provided that the patient has signed an appropriate consent (either a tumor tissue-specific consent or full study consent).
10. In the event of a repeat resection or biopsy during treatment or following progression, submission of tissue sample for central analysis is strongly encouraged.
11. Determination of triple-negative status will be done at a local laboratory. Tumor sample(s) used to determine eligibility must be obtained in the setting of advanced disease. Laboratory reports will be required to provide quantitative results of sufficient detail to verify eligibility for all patients enrolled (see study entry criteria).
12. Medical history includes demography, race, ethnicity, history of breast cancer, previous therapy, and pre-existing diseases. At Cycle 1, Day 1, medical history is updated with any adverse events occurring prior to administration of study drug.
13. Complete physical exam should be performed at screening; thereafter, symptom-directed exams are acceptable.
14. Vital signs to include height (at screening only), weight, respiration, pulse, temperature, and resting systolic and diastolic blood pressure. For patients receiving CDX-011, vital signs should be assessed pre-infusion, at 45 (±15) minutes during the infusion, and within one-half hour following completion of the infusion. (Note: weight is only assessed once per visit.) For patients receiving capecitabine, vital signs should be assessed once at the indicated visits.
15. A second original copy of the ECG tracing should be retained for possible submission to Celldex.
16. Serum or urine pregnancy test only for women of childbearing potential. Patients of non-childbearing potential include those who are ≥60 years, surgically sterilized, or postmenopausal with absence of menses for at least 1 year. However, women <60 with therapy-induced amenorrhea will require a pregnancy test unless additional evidence (oophorectomy or serial measurement of FSH and/or estradiol) are available to ensure postmenopausal status.

17. Laboratory assessments will include the following, when indicated. Hematology results must be reviewed prior to dosing.

Hematology:	**Clinical Chemistry:**	**Urinalysis**
Hemoglobin	Sodium	pH
Hematocrit	Potassium	Protein
Mean corpuscular volume (MCV)	Chloride	Glucose
Erythrocyte count (RBC)	Bicarbonate	Specific gravity
Leukocytes (WBC)	Glucose (nonfasting)	Blood
Platelets	Blood urea nitrogen (BUN)	
Differential:	Creatinine	
Neutrophils	Calcium	
Lymphocytes	Phosphate	*Microscopic examination must be performed at*
Monocytes	Alkaline phosphatase	*baseline and, if clinically indicated, at subsequent*
Eosinophils	Alanine transaminase (ALT/SGPT)	*visits (if urinary infection is suspected then a*
	Aspartate transaminase (AST/SGOT)	*negative urine culture is required prior to*
Differential should be reported consistently	Total protein	*enrollment).*
throughout the study as either an absolute	Albumin	
count (preferred) or as a percentage.	Lactate Dehydrogenase (LDH)	
	Total Bilirubin	

18. Analyses will be performed by centralized laboratories. Sample collection, processing and shipping instructions will be provided separately.
19. Samples drawn only for patients receiving CDX-011. On CDX-011 dosing days, samples are collected prior to dosing.
20. Samples for pharmacokinetic analyses will be collected for CDX-011-treated patients only. Analysis may also include circulating GPNMB or other soluble mediators. At each CDX-011 dosing day, samples will be collected prior to dosing and at end of infusion (at or within 15 minutes of completion of infusion).
21. PBMC collection will occur for up to 150 patients enrolled at a subset of study centers in the United States (regardless of treatment arm). On CDX-011 dosing days, samples are collected prior to dosing. Analysis will include examination of GPNMB expression on myeloid suppressor cells. Celldex will notify the sites when the requisite number of patients providing PBMC samples has been reached or that further data is not needed.
22. Blood sample will be stored for retrospective BRCA1/2 mutation status and other potentially relevant molecular markers.
23. EORTC QLQ-C30 Questionnaire (Protocol Appendix 6) completed only for patients who are fluent in a language in which the questionnaire is validated (information will be provided by Celldex). Each questionnaire must be completed by the patient prior to any other assessments or procedures for that visit. It should be completed without assistance if possible. If the patient requires assistance to complete the questionnaire, the assistance should be provided by clinical or study staff as opposed to family members or significant others, and the nature of required assistance should be documented. If a scheduled questionnaire is missed, the patient should complete the questionnaire at the next study visit
24. Imaging-based evaluation per RECIST 1.1 should be performed in accordance with the Site Manual provided by Celldex (or designee). Contrast-enhanced CT of the chest, abdomen, and pelvis, as well as all other suspected disease sites is required. MRI exams of the brain, abdomen, and pelvis can be performed in lieu of a CT; however MRI exams of the chest are not recommended. In the event that a chest MRI is performed, a non-contrast chest CT is strongly recommended to evaluate the lung parenchyma. Brain and/or bone scans are required for any patients with a history of metastases to bone and/or brain or where symptomatology raises the suspicion for bone and/or brain metastases. Lesions identified on bone scans should be confirmed by a CT or MRI at baseline, and, if identified as target lesions due to soft tissue component, they should continue to be followed by the same methodology (i.e., CT or MRI scan). However, bone lesions followed as non-target disease may be subsequently followed by bone scans only. Lesions that cannot be imaged but are assessable by clinical exam may be assessed using color photography including a ruler (preferred method) or measured with calipers. Normally, all target and non-target disease sites should be evaluated at each assessment. However, for patients with non-target bone disease, bone scans need only be repeated every twelve weeks. The same method of assessment and the same technique should be used to

characterize each identified and reported lesion at baseline and during follow-up. Target lesions selected for tumor measurements should be those where surgical resection or radiation are not indicated or anticipated.

25. Randomization may be performed up to 5 days prior to initiation of study treatment, but should occur only after confirming all eligibility criteria have been met.
26. Unless otherwise specified, all study assessments should be performed prior to administration of study treatment, and may be performed up to 24 hours prior to treatment administration if assessments remain within the specified visit window.
27. All concomitant medication will be documented in the CRF if taken within 28 days prior to Study Day 1, and either (whichever occurs sooner): a) 28 days after last dose of study treatment or b) initiation of alternate anti-cancer therapy. In addition, all anti-cancer medications and concomitant medications required to treat CDX-011-related SAEs taken throughout the duration of study follow-up should also be recorded.
28. For patients who develop grade 3 rash possibly related to CDX-011 and who provide appropriate consent, punch biopsies and photographs of the rash site, as well as uninvolved skin, are strongly encouraged. Samples may be submitted for central analyses including quantification of GPNMB expression; in these cases, collection, processing and shipping instructions will be provided separately.
29. Adverse event monitoring on Cycle 1 Day 7 and Cycle 1 Day 14 can be performed in person or by telephone to determine if the patient is experiencing any adverse events.
30. Events occurring > 28 days after discontinuation of study treatment are only reportable if serious (SAE) and potentially treatment-related.

2 STUDY OBJECTIVES

Primary Objective

- To evaluate the anti-cancer activity of CDX-011 in metastatic, GPNMB over-expressing, triple-negative breast cancer as measured by the duration of progression-free survival (PFS).

Secondary Objectives

- To further assess the anti-cancer activity of CDX-011 in metastatic, GPNMB over-expressing, triple-negative breast cancer, as assessed by the objective response rate (ORR), duration of response (DOR), and overall survival (OS).
- To further characterize the safety of CDX-011 in metastatic, GPNMB over-expressing, triple-negative breast cancer.
- To obtain pharmacokinetic parameters and to explore the relationship between patient-specific measures of exposure and safety and activity parameters.

Exploratory Objective

- To assess whether treatment with CDX-011 is associated with improvement in quality of life and/or cancer-related pain as reflected by reduced analgesic use.

3 ANALYSIS ENDPOINTS

3.1 Primary Endpoint

- Progression-free survival (PFS)

3.2 Secondary Endpoints

The following secondary efficacy endpoints will be evaluated:

- Objective response rate (ORR)
- Duration of response (DOR)
- Overall survival (OS)

3.3 Exploratory Endpoints

- Quality of life as measured by EORTC-QLQ-C30 and/or cancer-related pain as reflected by reduced analgesic use

3.4 Safety Endpoints

The following safety-related endpoints will be evaluated:

- Incidence of adverse events
- Deaths on study
- Discontinuations of study treatment due to adverse events
- Hematology, chemistry, and other laboratory parameters
- Vital sign parameters
- Electrocardiogram (ECG) parameters

4 RANDOMIZATION AND BLINDING

4.1 Randomization

Eligible patients will be centrally randomized in a 2:1 ratio to receive CDX-011 or capecitabine. For subjects enrolled in the study under protocol amendments 3 and 4, the randomization will be stratified by the following factors:

- 0-1 prior line of chemotherapy for advanced disease vs. 2 prior lines of chemotherapy for advanced disease
- Progression-free interval ≤ 6 months after receipt of taxane therapy vs. progression-free interval > 6 months after receipt of taxane therapy
- Received anthracycline therapy previously vs. no prior anthracycline therapy

There are no subjects enrolled under the original protocol. In protocol amendments 1 and 2, different stratification factors are considered. Specifically,

- Strata Under Amendment 1:
  - No lines of prior chemotherapy for advanced disease vs. 1 line prior chemotherapy for advanced disease – 0 vs. 1
  - "Resistant" to anthracycline therapy (i.e., progression-free interval of ≤ 6 months after completing treatment) vs. "Exposed" to anthracycline therapy (i.e., progression-free interval of > 6 months after completing treatment).

- Strata Under Amendment 2 and 2.1:
  - 0-1 prior line of chemotherapy for advanced disease vs. 2 prior lines of chemotherapy for advanced disease
  - "Resistant" to anthracycline therapy (i.e., progression-free interval of ≤ 6 months after completing treatment) vs. "Exposed" to anthracycline therapy (i.e., progression-free interval of > 6 months after completing treatment).

In the stratified analysis for this study, the stratification factors in Amendment 3 will be used. Given the changes in protocol amendments in stratification factors, the primary analysis of efficacy will be based on the actual stratification factors that are corrected after randomization. In addition, the secondary analysis using the *stratification factors as randomized* will also be performed, with the following mapping approach:

- All patients randomized under Amendment 1 will be considered as 0-1 prior line of chemotherapy for advanced disease.
- All patients randomized under Amendment 1 and 2.1 will be considered as received anthracycline therapy previously.
- For all patients randomized under Amendment 1 and 2.1, their Progression-free interval after receipt of taxane therapy will be determined based on the actual data reported in the clinical database.
- All patients randomized under/after Amendment 3 use the final stratification factors in randomization.

4.2 Blinding

This is an open-label study; therefore, neither the patients nor the investigators will be blinded to study treatment. Radiological and clinical disease assessments will be reviewed centrally in a blinded manner by the independent review committee (IRC). Members of the IRC will be blinded to study treatment. For data integrity consideration, a blinding process was established and specified in the document of *CDX011-04 Blinding and Unblinding Procedures*. In brief, two biostatistical groups are established: a blinded team (Celldex biostatistics team responsible for execution of the Statistical Analysis Plan, day-to-day interactions with the study team, and final study analyses) and an unblinded team (PPD biostatistics team responsible for supporting the Data Monitoring Committee (DMC) meetings). The blinded team will not have access to randomization and treatment data, and the unblinded team will have full access to the database.

5 SAMPLE SIZE CALCULATION

The primary efficacy endpoint is progression free survival (PFS). For patients with metastatic advanced GPNMB-expressing breast cancer in the treatment setting under investigation, median PFS of 4 months is anticipated following randomization to capecitabine. It is hypothesized that CDX-011 will increase median PFS in such patients by 2.25 months (i.e., 4.0 vs 6.25 months). Under the assumption of proportional hazards, such an increase corresponds to a hypothesized hazard ratio of 0.64. If this hypothesized hazard ratio is true, 203 PFS events (total of two arms) determined by IRC will provide 85% power with 2-sided type I error of 0.05. Under the assumption of exponential distribution for each arm and uniform enrollment over 2 years, and 10% drop out rate (PFS events cannot be observed), 300 patients (200 in the CDX-011 arm and 100 in the capecitabine arm) are needed, and it is anticipated that 203 PFS events will be observed in approximately 26 months from the date the first patient is randomized. The sample size calculation is performed using SAS v9.4.

6 ANALYSIS POPULATIONS

6.1 Efficacy Analysis

6.1.1 Intent-to-Treat (ITT) Population

The Intent-to-Treat (ITT) population will include all patients who are randomized. Patients will be analyzed according to the treatment arm to which they are randomized (CDX-011 or capecitabine).

6.1.2 Per-Protocol Population

The Per-Protocol population excludes patients that may substantially affect the results of the primary analysis, due to important deviations from the protocol. The final determination on protocol violations, and thereby the composition of the Per-Protocol population, will be documented in a separate memo prior to locking the clinical database and final analysis. Patients in the Per-Protocol population will be included in the treatment arm to which they actually received.

6.2 Safety Population

The safety population will include all patients who receive at least one dose of study treatment (either CDX-011 or capecitabine). Patients will be analyzed according to the treatment which they have actually received (CDX-011 or capecitabine). A baseline measurement and at least one laboratory, vital sign, or other safety-related measurement

obtained after at least one dose of study treatment may be required for inclusion in the analysis of a specific safety parameter.

7 GLOSSARY/ANALYTIC DEFINITIONS

7.1 Study Treatment

The term "study treatment" represents CDX-011 or capecitabine.

7.2 Study Day 1

Study Day 1 corresponds to the date of the first dose of study treatment.

7.3 Study Day

Study Day represents the elapsed number of days from Study Day 1, inclusive.

Study Day n = (Date of assessment – Date of Study Day 1) + 1 day,

if Date of assessment ≥ Date of Study Day 1. Or,

Study Day n = (Date of assessment – Date of Study Day 1),

if Date of assessment < Date of Study Day 1. Unless otherwise specified, the timing of all study-related events, assessments, and interventions will be calculated relative to Study Day 1. Study Day −1 will be the day before Study Day 1, and in general, negative days will be measured backwards starting from Study Day −1.

7.4 Cycle 1 Day 1

The first 21-day treatment cycle will begin with the first dose of study treatment, and will be denoted as "Cycle 1 Day 1". In general, the start date of each cycle (i.e., Day 1 of Cycle k, where k=1 to maximum number of cycles started) equals the date of the first dose of study treatment administered in the corresponding cycle. If study treatment is permanently discontinued, then the cycle number and cycle day for study-related assessments done after the last cycle of dosing will be replaced in the tables, figures, and listings with the visit name (e.g., End of Treatment Visit, Disease Assessment Visit WK24). The visit date and study day also will be used to identify the timing of such assessments.

7.5 Cycle *k* Day *j*

Day *j* represents the elapsed number of days since the first dose of study treatment in Cycle *k*.

$$\text{Day } j = (\text{Date of assessment} - \text{Date of Day 1 in Cycle } k) + 1$$

Unless otherwise specified, the timing of study-related visits and assessments will be calculated relative to Day 1 in each cycle.

7.6 Baseline

Unless specified otherwise, the baseline values will be those values measured closest to but not after the first dose of study treatment on Day 1 of Cycle 1.

8 PATIENT INFORMATION

8.1 Patient Disposition

The number of patients who are included in ITT, Per-Protocol, and Safety will be summarized by treatment arm. The reason for exclusion from one or more analysis populations will be provided.

The number of patients who are in the treatment phase and post-treatment follow-up phase of the study will be summarized by treatment arm. Patients discontinuing from study treatment and the primary reason for discontinuation will be summarized by treatment arm. Patients who withdrew from study and with the primary reason for discontinuation will be summarized in a similar manner.

8.2 Demographics and Baseline Characteristics

The following demographic and baseline characteristics considered pertinent to the outcomes for patients with metastatic breast cancer will be summarized for ITT population:

- Age, age group (≤40, 41 – 64, ≥65), sex, race, ethnicity, region, height, weight, body surface area (BSA), body mass index (BMI), ECOG performance status
- Breast cancer history, including time since initial diagnosis, time since diagnosis of advanced and/or metastatic breast cancer, number of prior relapses, stage at initial diagnosis, any prior CNS involvement, histopathology, estrogen receptor (ER) status at initial diagnosis and for advanced/metastatic disease, progesterone

receptor (PR) status at initial diagnosis and for advanced/metastatic disease, HER2 receptor status at initial diagnosis and for advanced/metastatic disease, ER/PR combined status for advanced/metastatic disease, ER/PR/HER2 combined status advanced/metastatic disease, GPNMB expression level in tissue used for study eligibility, current sites of disease including visceral involvement, number of organs involved (1, 2, >=3)

- Breast cancer treatment history, including prior surgery and use of radiation, number of prior lines of anticancer therapies, and setting received, and for specific, therapies of interest; number of prior lines of treatment for advanced disease; received prior anthracycline (Y/N), and progression-free interval after receipt of taxane therapy (≤ 6 months, >6 month)

The medical history term will be assigned standardized terms using the Medical Dictionary for Regulatory Activities (MedDRA). It will be summarized based on the number and percentage of patients by MedDRA dictionary terms and will also be presented in data listings.

9 PRIOR AND CONCOMITANT MEDICATIONS

Prior and concomitant medications will be coded to the World Health Organization (WHO) drug generic term using the WHO Drug Dictionary. Prior medications are defined as medications that started before the first dose of study drug. Concomitant medications are defined as medications that (1) started before the first dose of study drug and continued after first dose, or (2) started on or after the first dose of study drug and within 28 days after last dose. Prior and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) classification therapeutic subgroup and WHO generic term. All medication data will be listed individually and summarized by treatment arm according to anatomical therapeutic class and generic name.

10 EFFICACY ANALYSES

10.1 Multiplicity Consideration

This study has no planned interim analysis. For the primary analysis of PFS at final analysis, the test will be performed at two-sided 5% significance level. For multiplicity adjustment consideration for secondary endpoints, when the primary PFS analysis is positive, to strongly control the type I error at 5% level, Hochberg procedure (Hochberg 1988) will be used for ORR and OS. Specifically, Table 2 below describes the statistical testing scenarios:

Table 2: Hochberg testing procedure of the secondary endpoints ORR and OS

P values	Reject
$P_{(2)} \leq 0.05$	$H_{(1)}$ and $H_{(2)}$
$P_{(1)} \leq 0.025$ and $P_{(2)} > 0.05$	$H_{(1)}$
$P_{(1)} > 0.025$ and $P_{(2)} > 0.05$	None

Note: $P_{(1)}$ and $P_{(2)}$ are ordered *P* values of two tests for $H_{(1)}$ and $H_{(2)}$ respectively such that $P_{(1)} \leq P_{(2)}$, where $H_{(1)}$ and $H_{(2)}$ are two tests for ORR and OS.

According to this procedure, ORR and OS will both be considered positive if each analysis yields a *P* value of ≤ 0.05. Alternatively, any analysis resulting in a *P* value > 0.05 will not be considered positive. If only one analysis results in a *P* value > 0.05, the other must yield a *P* value ≤ 0.025 to be considered positive.

DOR will be analyzed as descriptive only and the statistical testing is not planned.

The HRQOL for this study is considered exploratory. The primary measure of HRQOL is the global health status score based on the EORTC QLQ-C30 questionnaire. No adjustment will be made for HRQOL analyses, and the p-values will be evaluated in an exploratory manner using a nominal significance level 0.05 (2-sided).

10.2 Subgroup Analyses

The following subgroup analyses will be performed for efficacy endpoints PFS, ORR, and OS:

- Age (≤40, 41 – 64, ≥65)
- Race (White vs other)
- Prior line of chemotherapy for advanced disease (0-1 vs >=2)
- Levels of GPNMB expression (25% - <50% or >=50%)
- Progression-free interval (PFI) after receipt of taxane therapy (≤ 6 months vs > 6 months)
- Prior anthracycline therapy (Yes vs No)
- Estrogen Receptor (ER)/Progesterone Receptor (PR) level (both <1% vs either >=1%)

- Visceral disease (Yes vs No)
- CNS involvement (Yes vs No)
- Number of disease site (<=3 vs >3)
- ECOG PS (0 vs 1)
- Region (North America vs Other)

The subgroups of PFI, prior anthracycline, and prior line of chemotherapy for advanced disease are determined based on actual clinical data. The subgroup analyses will be performed with the corresponding stratification factors if appropriate. If the number of subjects is too small in a particular stratum to produce reliable estimate, unstratified analysis may be used.

10.3 Progression-free Survival (PFS)

PFS is defined as the time from the date of randomization to documented disease progression, or death due to any cause, whichever is earlier.

10.3.1 Primary PFS Analysis

The primary analysis of PFS will be based on the Independent Review Committee (IRC) assessment according to RECIST 1.1. Unless specified otherwise, all recorded PFS events will be included in the primary analysis regardless of stopping randomized therapy. Table 3 below describes the detailed censoring rules for the primary PFS analysis. The determination of PFS is also detailed in the flow chart in the Appendix (Section 17).

Table 3: Date of Progression of Censoring for the Primary PFS analysis (IRC)

Situation	Date of Progression or Censoring	Outcome
Disease progression between scheduled disease assessments	Date of progression	Progressed
Death between scheduled disease assessment	Date of death	Progressed
Death before the first planned disease assessment at 6 weeks (plus 1 week window) without any disease assessment	Date of death	Progressed
No baseline disease assessments	Date of randomization	Censored
No post-baseline disease assessments without death before the first planned disease assessment	Date of randomization	Censored
Discontinuation of study treatment without progression (i.e., due to toxicity, withdrew consent, or lost to follow-up) and no further radiographic assessment[a]	Date of last radiological assessment without documented progression disease	Censored
Alternate anticancer treatment started without documentation of disease progression beforehand[b]	Date of last radiological assessment prior to start of further anticancer treatment	Censored
Death or disease progression after missing two or more consecutively scheduled disease assessments[c]	Date of last radiological assessment without documented progression disease prior to death/progression	Censored
Patients who have not progressed and are alive	Date of last radiological assessment without documented progression disease	Censored

a. For a patient who discontinues study treatment without PD and continues to have radiographic assessments, post-treatment radiographic assessments will be considered for PFS analyses according to the rules in this table.

b. Alternate anticancer treatment interventions that are diagnostic (e.g., biopsy) or palliative (e.g., local radiation for painful nodule) will not routinely require censoring unless they are determined to interfere with the assessment of the patients' disease status.

c. Disease assessments are planned every 6 (±1) weeks for the first 6 months (24 weeks); and every 9 (±2) weeks, thereafter, until disease progression. The interval of two missed scheduled assessments is calculated according to the scheduled assessment frequency and the corresponding window (1 or 2 weeks depending on assessments scheduled before or after 6 months).

The median PFS by treatment group will be estimated using Kaplan-Meier method and the Kaplan-Meier curves by treatment group will be presented.

The primary inferential comparison of this endpoint will use the stratified log-rank test stratified by the actual stratification factors corrected in the clinical database after randomization. A secondary analysis based on *the stratification factors as randomized* will also be performed (Section 4.1). Estimation of the hazard ratio for treatment and its corresponding 95% confidence interval will be determined using a stratified Cox proportional hazards model with the same stratification factors.

Subgroup analyses for PFS (Section 10.2) will be performed and a forest plot will be presented.

10.3.2 PFS Analysis (Investigator)

The analysis approach is the same as the PFS analysis based on IRC assessment.

Additionally, differences between the IRC and investigators assessments of disease progression will be summarized by treatment arm. Such differences will include the early discrepancy rate and late discrepancy rate (Zhang et al 2013), which are based on the frequency that the investigator's evaluation declares progression earlier or later than the IRC, respectively.

10.3.3 Sensitivity PFS Analysis

Sensitivity analysis will be performed to assess potential sources of bias due to 1) censoring for alternate anti-cancer therapy; 2) symptomatic disease progression.

Alternate anti-cancer therapy prior to documentation of disease progression

It is anticipated some patients may receive subsequent anti-cancer therapy prior to documentation of progressive disease. Censoring alternative anticancer treatment prior to documented progression may be problematic since it ignores the issue of informative censoring and contravenes the basic principle of an intention-to-treat analysis. Patients who stop taking randomized study treatment prior to documented progression frequently do so due to either toxicity of the drug or due to deterioration in the status of their disease. In such cases, the investigator may judge that immediate intervention, commonly in terms of the introduction of a new cancer treatment, is in the best interest of the patient without necessarily waiting for confirmatory radiographic evidence of progressive disease (Carroll 2007).

To address the above concerns, the following sensitivity analysis will be performed: Death or disease progression that's documented after the initiation of an alternate anti-cancer therapy will be considered as a PFS event.

Symptomatic progression / hospice care events

A sensitivity analysis will be performed with symptomatic deterioration or admission to hospice care considered a PFS event with investigator's disease assessments. Patients with symptomatic deterioration will be considered to have disease progression the day after the date of the last disease assessment showing no evidence of disease progression. For patients who enter hospice care without prior symptomatic deterioration or disease progression per RECIST 1.1, the date of admission to the hospice care will be considered as PFS event date.

10.4 Objective Response Rate (ORR)

ORR will be estimated based on the proportion of patients in each treatment arm with measurable disease at baseline whose best overall response during the course of study treatment is complete response (CR) or partial response (PR). Tumor response will be assessed using RECIST 1.1. The primary analysis of ORR will be based upon evaluations by the IRC regardless of tumor response (CR or PR) confirmation. The secondary analysis of ORR will be based on the confirmed (no sooner than 28 days later) tumor response (CR or PR). Supplemental analysis will be performed based on investigator's review.

Exact 95% confidence intervals of ORR will be calculated by treatment arm using Clopper-Pearson method. The inferential comparison of the observed ORRs will be made using the Cochran-Mantel-Haenszel Chi-square test, stratified by the randomization stratification factors.

Subgroup analyses will be performed for the list of subgroups specified in Section 10.2. Waterfall plots will be used to depict graphically for individual patients their maximum percentage shrinkage from baseline in the sum of the longest diameters in target lesions.

Best overall response will be classified as 'not evaluable / unknown' for patients with early progression or early death. Early progression will include patients who discontinue study treatment due to worsening disease but without objective evidence of disease progression, and their best overall response cannot be determined because tumor assessments were either incomplete or were never performed or not repeated. Similarly, early death will include patients who die without objective evidence of disease

progression beforehand, and their best overall response cannot be determined because tumor assessments were either incomplete or were never performed or not repeated.

10.5 Duration of Response (DOR)

DOR will be calculated for patients who achieve CR or PR regardless of confirmation. For such patients, DOR is defined as the number of months from the start date of PR or CR (whichever response is recorded first), to the first date that progressive disease (PD) or death is documented.

The same censoring rules of the primary PFS analysis will be applied to DOR.

Median DOR will be estimated according to the Kaplan-Meier method. Inferential comparison between treatment arms for DOR is not planned.

As noted previously, the primary analysis of DOR will be based on the time point and best overall responses and PFS events determined by blinded independent central review. A supplemental analysis of DOR will be performed based on investigator assessments using the same data cutoff and censoring rules specified in PFS for the primary analysis.

10.6 Overall Survival (OS)

OS is defined as the time from randomization to the date of death due to any cause. Patients who are alive or lost to follow-up as of the data analysis cutoff date will be censored at last known alive date before cutoff.

The OS will be summarized descriptively using the Kaplan-Meier method. Median follow-up for OS will be estimated according to the Kaplan-Meier estimate of potential follow-up (Schemper 1996). The KM plot will be produced by treatment arm.

The primary inferential comparison of this endpoint will use the stratified log-rank statistic to test the null hypothesis that the distribution of OS times is equal for the 2 treatment arms.

Subgroup analyses will be performed for the subgroup factors listed in Section 10.2. For each subgroup, the hazard ratio for treatment arm will be estimated and the KM plot will be produced. Forest plots will be used to graphically display the hazard ratios and 95% confidence intervals across subgroups. Kaplan-Meier estimates will be provided for each treatment arm within each subgroup.

Survival rate at one-, two-, and three-year after randomization will be estimated for each treatment arm using the Kaplan-Meier product-limit method. Greenwood's formula will be used to calculate the standard error of the Kaplan-Meier estimate and corresponding 95% confidence interval.

To supplement the analysis of OS, a summary of the subsequent anti-cancer therapies and interventions initiated after the discontinuation of study treatment will be provided. Such therapies will be collected during the scheduled post-treatment follow-up assessments.

10.7 Health-Related Quality of Life (HRQOL)

HRQOL will be assessed in a subset of patients using the EORTC quality of life core questionnaire-30 (QLQ-C30, version 3). The EORTC QLQ-C30 consists of 30 individual items, each representing single- or multi-item scales. Of the 30 questions, 24 aggregate into 9 multi-item scales representing the following HRQOL dimensions: 5 functioning scales (physical, role, emotional, cognitive, and social), 3 symptoms scales (fatigue, nausea and vomiting, and pain), and 1 global measure of health status (Table 4). The remaining 6 single-item scales assess the following symptoms: dyspnea, insomnia, appetite loss, constipation, diarrhea, and the perceived financial impact of the disease treatment.

The individual items on both EORTC questionnaires will be scaled and scored according to the EORTC QLQ-C30 Scoring Manual (Fayers 2001). The raw scores will be transformed to a linear scale ranging from 0 to 100, where higher scale scores indicate better HRQOL for the global health status and functioning scales, and worse HRQOL for the symptom scales. Differences from baseline of at least 10 points will be classified as the minimum clinically meaningful change in an HRQOL measure (Osoba 1998; Maringwa 2011). Changes of more than 20 points will be classified as large effects. For example, an increase of 10 points on a functional scale will represent a moderate improvement, whereas a decrease of 10 points will represent moderate worsening. Changes of less than 10 points will be regarded as no change, or as clinically irrelevant.

The functional scales will be scored as follows:

$$Score = \left\{1 - \frac{(RS - 1)}{range}\right\} \times 100$$

where *range* is the difference between the possible maximum and minimum response to individual items. Where the majority of items take values from 1 to 4, giving *range* = 3.

Raw score (RS) is the mean of the component items:

$$RS = (I_1 + I_2 + I_3 + \ldots + I_n)/n$$

Similarly, the symptom scales/items and global health status will be scored as follows:

$$Score = \{(RS - 1)/range\} \times 100$$

For multi-item scales, if at least half the items in the scale are complete, then the scale score will be calculated based on the items for which data are available. Otherwise, the scale score will be set to missing. Under this approach, none of the single-item measures will be imputed.

Table 4: Scoring the EORTC QLQ-C30 (version 3.0)

Scales/Items	Number of Items	Item Range	Item Numbers
Global health status			
Global health status	2	6	29, 30
Functional scales			
Physical functioning	5	3	1 to 5
Role functioning	2	3	6, 7
Emotional functioning	4	3	21 to 24
Cognitive functioning	2	3	20, 25
Social functioning	2	3	26, 27
Symptom scales/items			
Fatigue	3	3	10, 12, 18
Nausea and vomiting	2	3	14, 15
Pain	2	3	9, 19
Dyspnea	1	3	8
Insomnia	1	3	11
Appetite loss	1	3	13
Constipation	1	3	16
Diarrhea	1	3	17
Financial difficulties	1	3	28

Baseline HRQOL assessments will be administered to patients prior to randomization. Post-baseline HRQOL assessments are planned for Day 1 of Cycles 3, 5, 7, etc. (i.e., "odd cycles") and End of Treatment (EOT) visit.

Summarization of the HRQOL data will be based on the actual study visits (including unscheduled visits and EOT visits) mapped at Weeks 6, 12, 18, etc.

If multiple HRQOL assessments are associated with the same time point, the assessment performed closest to the planned study day will be used for purposes of summarizing the data. If 2 measurements are equidistant from the planned study day, the measurement occurring prior to the planned study day will be used for summarization.

The primary HRQOL analysis will be based on the subset of patients in the ITT population who complete the HRQOL questionnaire at baseline.

Missing assessments present a challenge in analyzing and interpreting HRQOL data, especially when the data are missing non-randomly. For example, patients who experience increased toxicity or disease progression, hospice entry, or death are more likely to miss one or more assessments. Similarly, in a longitudinal study data may be missing at later assessments because of death or withdrawal due to disease progression. This leaves purportedly healthier patients, with better HRQOL scores on study longer and displaying a misleading trajectory of higher scores over time. If the reasons for missing data are ignored and the analyses are based only on the observed data, the estimates may be biased in that they do not accurately reflect the entire population of study patients (Troxel 1998)

Mixed effects models for repeated measures data based on maximum likelihood theory provide valid results when data are missing at random (MAR). Data are MAR if they depend on the observed data only. If the probability of missing depends on the unobserved HRQOL scores, then the missing data are considered non-ignorable, or missing not at random (MNAR). This type of missing data can occur in association with death, disease progression, or toxicity. Distinguishing between MAR and MNAR is not trivial and relies on assumptions that are themselves untestable because the required data are missing. Applications of mixed models have been developed that can assist in the analysis of non-ignorable missing data. One such application is the pattern-mixture models. In this method, patients are stratified by the missing data pattern, and mixed models are created within each stratum. Parameter estimates for each stratum are then combined into a weighted average for the entire study population.

The pattern of missing HRQOL data in this study will be examined with patients classified into groups based on the timing of study dropout for purposes of future HRQOL assessment. The chi-square test and two-sample t-test, as appropriate, will be used to compare baseline characteristics between dropout groups. Because it is not possible to definitively determine whether missing data are ignorable, sensitivity analyses

(comparing the mixed effects model and pattern-mixture model, described below) will be performed to ensure that treatment group effects are consistent across different analytical methods with different missing data assumptions and to evaluate the range of possible treatment group differences.

The initial HRQOL analysis will evaluate the EORTC QLQ-C30 scores (global health status and each individual function/symptom scale) longitudinally using a linear mixed effects model for repeated measures. This model, which assumes missing data are MAR, will describe the rate of change in HRQOL scores over time for each treatment arm, taking into account the between-patient variability by incorporating each patient's individual starting point and individual rate of change. The MIXED procedure in SAS will be used with patients as the random effect; treatment arm will be defined as the fixed effect. The model also will include linear time effects (i.e., week) and time by treatment interaction terms. A covariance structure that follows an autoregressive (AR) process dependent only on the prior HRQOL assessment will be used (i.e., AR-1). If there is model convergence issue, then compound symmetry covariance structure will be used.

Because missing HRQOL data are anticipated to be related to worsening prognosis, a sensitivity analysis will be performed using a pattern-mixture model to assist in the analysis of potentially non-ignorable missing data. For this analysis, patients will be stratified into groups based on their treatment arm and timing of study dropout for HRQOL assessment (anticipated to be largely due to disease progression and death). Patients will be classified as "dropouts" if their last HRQOL assessment was prior to 6 months and as "completers" if their last assessment was 6 months or later. Mixed effects models will be created within each stratum, which will provide for separate estimates of treatment effect for "dropouts" and "completers". Parameter estimates for each stratum will then be combined into a weighted average for the entire treatment arm. As noted previously, selected demographics and disease characteristics also will be summarized for "dropouts" and "completers" separately to evaluate if the missing HRQOL data are MAR.

A check of the distributional assumptions for the longitudinal models will be performed, along with diagnostics to examine residuals and the presence of any outlying and influential observations. The results from the data checking and model diagnostics will be presented in a separate report.

The HRQOL objective is considered exploratory. For all HRQOL scales, the p-values will be evaluated in an exploratory manner using a nominal significance level of 0.05 (2-sided).

10.8 Cancer-Related Pain

The improvement in cancer-related pain will be summarized using the proportion of days using analgesic to the total number of days during treatment in the study. The WHO Drug anatomical therapeutic class and generic name that represent analgesic drugs will be used to identify the use of analgesic therapy after the study treatment is administered.

10.9 Exploratory analysis

Exploratory analysis may be performed for PFS and OS adjusted for baseline prognostic factors such as total prior regimens, GPNMB expression levels, stromal GPNMB expression levels, and other factors using multivariate Cox proportional hazards models.

In addition, exploratory analysis of rash in cycle 1 will be performed to assess its correlation with clinical outcomes.

11 STUDY TREATMENT

CDX-011 will be administered as an intravenous infusion over 90 (±10) minutes on Day 1 of each 21 day cycle. Subjects will continue to be dosed once every three weeks until disease progression or unmanageable treatment-related toxicities.

The duration of exposure to study treatment, total number of cycles initiated, cumulative dose administered, and relative dose intensity will be summarized. The calculation for the above statistics is specified below:

- Duration of Exposure (days) = Date of first dose of last cycle – Date of first dose + 21 days
- For CDX-011, the actual dose at each cycle (mg/kg) = Total Vol infused (mL) / [Vol of Infusion bag (mL)] * Volume of CDX-011 added to bag (mL) * 5 (mg/mL) / Weight at the cycle (kg).
- For capecitabine, the actual dose at each cycle (mg/m^2) = Total doses at each cycle (mg) / BSA (m^2).
- Cumulative dose is the sum of all actual doses for all cycles.
- Actual dose intensity (ADI) = Cumulative Dose / Duration of Treatment (weeks).
- Relative Dose Intensity = Actual Dose Intensity / Planned Dose Intensity (PDI).

Dose modifications and reasons will be summarized by treatment. For CDX-011, dose modifications include dose delayed since last dose, dose reduced since the last dose, infusion interrupted, infusion slowed, infusion prematurely discontinued, and other. For capecitabine, dose modifications include dose delayed/interrupted since the last dose, dose level reduced since the last dose, prematurely discontinued, and other.

12 SAFETY ANALYSES

12.1 General Considerations

Safety will be assessed by clinical review of all relevant parameters including adverse events (AEs), serious adverse events (SAEs), laboratory values (hematology, serum chemistry, and urinalysis), vital signs, physical examinations, 12-lead electrocardiograms (ECGs), and ECOG performance status. The safety analyses will be performed using the safety population defined in Section 6.2.

12.2 Adverse Events

Summary tables and listings will be provided for all reported treatment emergent adverse events. Treatment emergent adverse events (TEAEs) are defined as adverse events that start on or after the first administration of study treatment and through 28 days after the last dose of study treatment. If it cannot be determined whether an adverse event is treatment-emergent (based on AE start date or, if AE start date is missing or partial AE start date), conservatively, the adverse event will be considered treatment-emergent event.

For partially missing AE start dates (missing day and/or month), the following imputation rules will be applied:

If both day and month of the AE start date are missing, the day and month will be imputed with the day and month of the first study dose date if the year is equal to the year of the first dose. Otherwise, the month and day is imputed as the first day of the year (01 Jan).

If only day is missing, and if the year and month are equal to the first dose date, the AE start date will be imputed as the first dose date. Otherwise, the day will be imputed as "01". As a general consideration, imputed AE start date cannot be after AE end date if available.

The reported adverse event term will be assigned standardized terms using the Medical Dictionary for Regulatory Activities (MedDRA). The version of MedDRA that is actually used for this study will be noted in the clinical study report.

Treatment emergent adverse events will be summarized based on the number and percentage of patients experiencing events by MedDRA system organ class and preferred term. The causal relationship between the occurrence of an adverse event and study treatment will be judged by the investigator as "related" or "unrelated". A worst-case scenario approach will be taken to handle missing relatedness data, i.e. the summary table of treatment-related AEs will include events with the relationship to study treatment reported as "Related" or "Missing". In the event a patient experiences multiple episodes of the same adverse event, then the event with the highest severity grade and/or strongest causal relationship to study treatment will be used for purposes of incidence tabulations.

Tabular summaries of the following treatment emergent adverse events will be provided:

- Overview of TEAEs
- TEAEs by SOC and PT
- Treatment-related TEAEs by SOC and PT
- TEAE by PT by decreasing frequency
- TEAEs by SOC and PT and worst toxicity grade
- Treatment-related TEAEs by SOC and PT and worst toxicity grade
- TEAEs with action of study treatment delayed/interrupted
- TEAEs with action of treatment reduced
- TEAEs with action of study treatment discontinued
- Serious TEAEs by SOC and PT
- Treatment-related serious TEAE by SOC and PT
- TEAEs with fatal outcome by SOC and PT
- Treatment-related TEAEs with fatal outcome by SOC and PT

In addition to the tabular summaries, adverse events will be reported in individual patient listings.

12.3 Laboratory Values

Hematology and serum chemistry tests will be evaluated at Screening, Day 1 of each treatment cycle, and End of Treatment visit. The hematology and serum chemistry values (actual value and change from baseline value) will be summarized in a descriptive manner by calculating the mean, standard deviation, median, and range by treatment arm as follows:

- Baseline value
- Minimum post-baseline value
- Maximum post-baseline value
- Average post-baseline value
- Last post-baseline value

Whenever available, laboratory values will be assigned toxicity grades using the NCI CTCAE version 4.0. These criteria may include qualifying definitions (e.g., clinical adverse event and/or requirement for concomitant medication) in addition to the specific laboratory value in the definition of the toxicity grades for some laboratory tests. For such tests, the qualifying definitions will not be used for the assignment of toxicity grades.

Shifts in laboratory values to outside the laboratory standard normal range will be evaluated for selected laboratory tests by assessing, relative to the baseline value, the maximum increase and/or decrease observed at each time point, throughout the course of treatment/observation. The number and proportion of patients with directional shifts above or below the standard normal range will be summarized for selected laboratory tests.

Similar analyses will be performed for shifts in NCI CTCAE v4.0 toxicity grades relative to the baseline toxicity grade by visit and worst grade during treatment.

Figures of cumulative incidence distributions may be provided that depict for each treatment arm, the cumulative probability for the development of treatment emergent hematological toxicity (e.g., neutropenia [ANC], thrombocytopenia [platelets], anemia [hemoglobin]) over time measured from the first dose of study treatment. For each hematological toxicity of interest, separate cumulative incidence distributions will be provided for the following CTCAE toxicity grade(s): grades 1-2 combined, grade 3, grade 4, and grades 3-4 combined.

The onset time of each hematological toxicity at a certain toxicity grade will be based on the earliest occurrence, if more than one occurrence is reported. The use of blood transfusions (platelets, red blood cells) and/or growth factor support will be summarized to supplement these analyses.

12.4 Deaths

All deaths that occur on study will be reported in patient listing format. The listing will include the primary cause of death and the number of days since the date of the last dose of study treatment. The primary cause of death will be summarized by treatment group.

12.5 Vital Signs

Vital signs (diastolic and systolic blood pressure, heart rate, respiration rate, and temperature) will be evaluated at Screening, Day 1 (pre-dose for all patients) of each treatment cycle, and End of Treatment visit. The reported values and change from baseline values for each vital sign will be summarized in a descriptive manner by calculating the mean, standard deviation, median, and range by treatment arm as follows:

- Baseline value
- Minimum post-baseline value
- Maximum post-baseline value
- Average post-baseline value
- Last post-baseline value

For patients receiving CDX-011, in addition to the vital sign measurements taken before the start of the infusion on Day 1 of each cycle, measurements also will be obtained at 1) 45 (±15) minutes after the start of the infusion and 2) within 30 minutes after completion of the infusion. For these patients, vital signs will be further assessed by calculating 1) the difference between the pre-dose value and that obtained 45 (±15) minutes after the start of the infusion (time period 1), and 2) the difference between the pre-dose value and that obtained within 30 minutes after completion of the infusion (time period 2). For each patient, the maximum increase and maximum decrease in each time period that is determined across all visits will be summarized in a descriptive manner.

12.6 ECOG Performance Status

ECOG performance status will be evaluated at Screening, Day 1 of each treatment cycle, and End of Treatment visit. ECOG performance status scores will be summarized in a

descriptive manner by treatment arm using the conventions described for laboratory values and vital signs. In addition, post-baseline scores will be summarized descriptively using shifts from baseline.

12.7 Electrocardiograms

ECG assessments are planned at Screening, Day 1 of Cycle 1, and End of Treatment visit. ECG results will be summarized using shifts from baseline for the number and percentage of patients with abnormal (clinically significant and not clinically significant, separately) and normal findings, as reported by the local investigator. Additionally, descriptive statistics will be tabulated for QT interval values (corrected using Frederica's method) along with changes from baseline to End of Treatment. Categorical analyses of QTc interval data will be provided using shifts from baseline for the number and percentage of patients meeting or exceeding the following threshold values: >450 ms, >480 ms, >500 ms. Additionally, the number and percentage of patients with QTc interval increases from baseline >30 ms and >60 ms will be reported. These thresholds are based on the October 2012 FDA Guidance for Industry, '*E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs*'.

12.8 Physical Examinations

Physical examinations are to be performed at Screening, Day 1 of each treatment cycle, and End of Treatment visit. Clinically significant abnormalities that are identified during physical examinations will be reported as adverse events. The disposition of each physical examination (whether performed or not) and the date the physical examination was performed will be provided in individual patient listings only. Summary tabulation of the physical examination data reported for this study is not planned.

13 PHARMACOKINETICS AND PHARMACODYNAMICS

13.1 Pharmacokinetics

Concentration of the antibody-drug conjugate (ADC), total antibody (TA) and free MMAE will be obtained and summarized. Population PK analyses will be performed. The influence of key intrinsic factors, such as weight and gender, on variability in PK will also be evaluated. The impact of circulating GPNMB levels on pharmacokinetic parameters may be examined. The pharmacokinetic parameters will be summarized. The relationships between patient-specific measures of exposure and safety/activity parameters will be explored.

13.2 Pharmacodynamics

Pharmacodynamics will be evaluated via assessment of tumor tissue obtained via voluntary biopsy or re-resection. Parameters evaluated may include localization of CDX-011, CR011 or MMAE at the tumor site and/or GPNMB expression levels in tumor tissue, as well evaluation of tumor infiltrating leukocytes. Soluble GPNMB levels in the circulation may also be examined.

PK analyses and PD analyses results will be provided in a separate report.

13.3 Immunogenicity

Patients will be monitored for the development of anti-CDX-011 and anti-CR011 antibodies, and whether these antibodies are neutralizing. Immunogenicity analysis may be performed and reported separately.

14 CHANGE FROM PROTOCOL OR APPROVED SAP

Version 1.1 has a clarification of ORR analysis regarding measurable disease in Section 10.4.

- Version 1.0: ORR will be estimated based on the proportion of patients in each treatment arm whose best overall response during the course of study treatment is complete response (CR) or partial response (PR).
- Version 1.1: ORR will be estimated based on the proportion of patients in each treatment arm with measurable disease at baseline whose best overall response during the course of study treatment is complete response (CR) or partial response (PR).

15 REFERENCES

Carroll KJ: Analysis of progression-free survival in oncology trials: Some common statistical issues. *Pharmaceutical Statistics* 6: 99–113, 2007.

Eisenhauer, E. A., P. Therasse, et al. (2009). "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)." Eur J Cancer **45**(2): 228-247.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomly A, on behalf of the EORTC Quality of Life Group. *The EORTC QLQ-C30 Scoring Manual (3rd Edition)*. European Organization for Research and Treatment of Cancer, Brussels, 2001.

Maringwa J, Quinten C, et al: Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. *Annals of Oncology* 22: 2107–2112, 2011.

Osoba D, Rodrigues G, et al: Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology* 6: 139–144, 1998.

Schemper M, Smith TL: A note on quantifying follow-up in studies of failure time. *Controlled Clinical Trials* 17: 343–346, 1996.

Troxel A, Fairclough D, et al: Statistical analysis of quality of life with missing data in cancer clinical trials. *Statistics in Medicine* 17: 653–666, 1998.

US Food & Drug Administration: Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. May 2007

US Food & Drug Administration: Guidance for Industry. E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs. October 2012

Hochberg Y: A sharp Bonferroni procedure for multiple tests of significance. *Biometrika (1988), 75, 4, pp.800-2*

Zhang J, Zhang L et al: Assessment of audit methodologies for bias evaluation of tumor progression in oncology clinical trials. *Clin Cancer Res*, 2013

16 SHELLS FOR TABLES, LISTINGS, AND FIGURES

The mock tables, listings and figures are provided in a separate document.

17 APPENDIX

Study CDX011-04 Progression Free Survival (PFS) Determination Flow Chart – Primary Analysis