Title: A Randomized, Open-Label, Phase 3 Trial of brentuximab vedotin (SGN-35) Versus Physician's Choice (Methotrexate or Bexarotene) in Patients With CD30-Positive Cutaneous TCell Lymphoma

NCT Number: NCT01578499

SAP Approve Date: May 25, 2016

Certain information within this Statistical Analysis Plan has been redacted (ie, specific content is masked irreversibly from view with a black/blue bar) to protect either personally identifiable (PPD) information or company confidential information (CCI).

This may include, but is not limited to, redaction of the following:

- Named persons or organizations associated with the study.

- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.

- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

STATISTICAL ANALYSIS PLAN

A Randomized, Open-Label, Phase 3 Trial of Brentuximab Vedotin (SGN-35) Versus Physician's Choice (Methotrexate or Bexarotene) in Patients With CD30-Positive Cutaneous T-Cell Lymphoma

Protocol #: C25001

SAP Version:		Date of Statistical Analysis Plan:
Draft	4.0	26 Feb 2014
Final	1.0	25 May 2016

The Final version of this SAP is unchanged in content from the Draft 26Feb2014 version. The re-signature was performed when the cover page was changed from "Draft" to "Final" (this was the only change made to the SAP).

PPD

TABLE OF CONTENTS

List of Abbreviations and Definitions of Terms

Abbreviation	Term
ADC	antibody-drug conjugate
AE	adverse event
AIC	Akaike information criterion
ALCL	anaplastic large cell lymphoma
ALK	anaplastic lymphoma kinase
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANC	absolute neutrophil count
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
ATA	antitherapeutic antibodies
ATC	anatomical therapeutic chemical
BIC	Bayesian information criterion
BSA	body surface area
BUN	blood urea nitrogen
BV	brentuximab vedotin
CD30+	CD30-positive
CDF	cumulative distribution function
CI	confidence interval
C_{max}	maximum plasma concentration
C_{min}	minimum plasma concentration
CO_2	carbon dioxide
CR	complete response
CT	computed tomography
CTCL	cutaneous T-cell lymphoma
C_{trough}	trough concentration
CV	coefficient of variation
DOR	duration of response
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EFS	event-free survival
EOT	end of treatment
GGT	gamma-glutamyl transpeptidase
GRS	global response score
HDL	high-density lipoprotein
HLT	High Level Term
IDMC	independent data monitoring committee
IRF	independent review facility

Abbreviation	Term
ITT	intent-to-treat
IV	intravenous; intravenously
IVRS	interactive voice response system
LDH	lactic dehydrogenase
LDL	low density lipoprotein
MedDRA	Medical Dictionary for Regulatory Activities
MF	mycosis fungoides
MID	minimum important difference
MMAE	monomethylauristatin E
MRU	medical resource utilization
mSWAT	modified severity weighted assessment tool
MTX	methotrexate
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	objective response rate
ORR4	objective response lasting at least 4 months
pcALCL	primary cutaneous anaplastic large cell lymphoma
PD	progressive disease (disease progression)
PFS	progression-free survival
PK	pharmacokinetic(s)
PP	per protocol
PR	partial response
PRO	patient-reported outcome
PT	Preferred Term
SAE	serious adverse event
SAP	statistical analysis plan
sCD30	soluble CD30
SOC	System Organ Class
SOE	Schedule of Events
T3	triiodothyronine
T4	thyroxine
TEAE	treatment-emergent adverse event
TSH	thyroid-stimulating hormone
WHO	World Health Organization

1. INTRODUCTION

In general, the purpose of the Statistical Analysis Plan (SAP) is to provide a framework that addresses the protocol objectives in a statistically rigorous fashion, with minimized bias or analytical deficiencies. Specifically, this plan has the following purpose:

To prospectively (a priori) outline the types of analyses and data presentations that addresses the study objectives outlined in the protocol, and to explain in detail how the data will be handled and analyzed, adhering to commonly accepted standards and practices of biostatistical analysis in the pharmaceutical industry.

1.1 Study Design

This is an open-label, randomized, multicenter study. All patients must have histologically-confirmed CD30-positive (CD30+) cutaneous T-cell lymphoma (CTCL) by central laboratory assessment and pathology review. Patients are to be stratified by diagnosis: primary cutaneous anaplastic large cell lymphoma (pcALCL) or mycosis fungoides (MF), and randomized to treatment with brentuximab vedotin monotherapy (Arm 1) up to 16 cycles total (48 weeks), or methotrexate (MTX) or bexarotene per physician's choice (Arm 2) up to a total of 48 weeks.

Objective response, lasting at least 4 months (ORR4), per independent review facility (IRF) will be assessed by global response score (GRS), which consists of skin evaluation (mSWAT assessment), nodal and visceral radiographic assessment, and detection of circulating Sézary cells (MF only). Complete response (CR) or partial response (PR) per GRS will be confirmed by sustained response in skin per mSWAT assessment at the subsequent cycle. All randomized patients will be followed for survival every 12 weeks for a minimum of 24 months after the EOT visit, and then every 6 months until patient withdrawal, death, or study closure.

Patients with a documented PR or CR at the end of Cycle 3 and at any disease assessment thereafter may continue treatment with study drug for up to 16 cycles with brentuximab vedotin or 48 weeks with bexarotene or methotrexate; patients with stable disease may continue therapy at the discretion of the investigator. Patients with progressive disease (PD) at any time during the study will be discontinued from study treatment. Patients with increasing mSWAT before assessment at the end of Cycle 3 will be allowed to continue therapy until the Cycle 3 assessment if, in the investigator's opinion, the findings are due to tumor flare rather than disease progression.

1.2 Study Objectives

1.2.1 Primary Objectives

The primary objective is:

- To determine ORR, lasting at least 4 months, with brentuximab vedotin in patients with CD30+ MF or pcALCL compared to that achieved with therapy in the control arm

1.2.2 Secondary Objectives

The key secondary objectives are:

- To determine complete response (CR) rate with brentuximab vedotin compared to that achieved with therapy in the control arm

- To determine progression-free survival (PFS) with brentuximab vedotin compared to that achieved with therapy in the control arm

- To determine burden of symptoms during treatment with brentuximab vedotin compared to that achieved with therapy in the control arm

Other secondary objectives include:

- To assess duration of response (DOR) with brentuximab vedotin

- To assess duration of skin response with brentuximab vedotin

- To determine event-free survival (EFS) with brentuximab vedotin compared to that achieved with therapy in the control arm

- To describe the pharmacokinetics (PK) of brentuximab vedotin and monomethylauristatin E (MMAE) in blood

- To determine the immunogenicity of brentuximab vedotin

- To assess patient-reported quality of life outcomes

- To assess the safety of brentuximab vedotin

1.2.3 Exploratory Objectives

The exploratory objectives include:

CCI

2. POPULATIONS FOR ANALYSIS

2.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population will include all patients identified as CD30+ by the Ventana CD30 (Ber-H2) assay and randomized to treatment. All patients in the ITT population will be analyzed according to the treatment they were randomized to receive and not according to what they actually received, if different.

The ITT population will be used for the primary efficacy analysis and analyses on all efficacy endpoints unless specified otherwise.

2.2 All-Enrolled Population

The All-Enrolled population will include all patients randomized to treatment. All patients in the All-Enrolled population will be analyzed according to the treatment they were randomized to receive and not according to what they actually received, if different.

The all-enrolled population will be used as a supplement to the analysis of the ITT population for the primary efficacy endpoint.

2.3 Per-Protocol Population

The Per-Protocol (PP) population will include a subset of ITT patients who do not have major protocol violations as determined by the medical monitor. All decisions to exclude patients from the PP population will be made before the study database lock.

The PP population will be used as a supplement to the analysis of the ITT population for the primary efficacy endpoint.

All patients in the PP population will be analyzed according to the actual treatment received.

2.4 Response-Evaluable Population

The Response-Evaluable population is defined as a subset of the ITT population with measurable disease at baseline and with at least 1 post baseline response assessment. The Response-Evaluable population will be used for the sensitivity analyses of ORR4, CR rate, ORR, time to response, and duration of response, as needed.

2.5 Safety Population

The Safety population will include patients who receive at least 1 dose of study drug. All patients in the Safety population will be analyzed according to the actual treatment received. Patients receiving any dose of brentuximab vedotin will be grouped into the study drug group for the purposes of safety reporting.

All safety analyses will be performed using the Safety population.

2.6 Pharmacokinetics Population

The PK population will include patients with sufficient dosing and PK data to reliably estimate PK parameters as determined by a clinical pharmacologist.

PK analyses will be performed using the PK population.

2.7 Pharmacodynamics Population

The Pharmacodynamics population will include patients with sufficient dosing and sufficient pharmacodynamics data to reliably measure pharmacodynamics parameters.

Pharmacodynamic analyses will be performed using the Pharmacodynamics population.

3.   HYPOTHESES AND DECISION RULES

3.1   Statistical Hypotheses

The primary null hypothesis to be tested is that the proportion of patients achieving ORR4 per IRF is not different between the brentuximab vedotin (BV) treatment group and the physician's choice treatment group (MTX or bexarotene); i.e.,

$$H_0: \ ORR4_{BV} = ORR4_{Control}$$

The alternative hypothesis is that ORR4 per IRF in the brentuximab vedotin treatment group is higher than that of the physician's choice group, i.e.,

$$H_A: \ ORR4_{BV} > ORR4_{Control}$$

The key secondary hypotheses to be tested include:

- CR rate per IRF is not different between the brentuximab vedotin treatment group and the physician's choice treatment group.

- PFS is not different between the brentuximab vedotin treatment group and the physician's choice treatment group.

- Changes in symptom domain per the Skindex-29 assessment tool are not different between the brentuximab vedotin treatment group and the physician's choice treatment group.

3.2   Statistical Decision Rules

The primary null hypothesis will be tested at a type I error rate of 0.05 (2 sided). We will conclude that brentuximab vedotin treatment can significantly improve the ORR4 compared to the physician's choice (MTX or bexarotene) in patients with CD30+ CTCL if the ORR4 for the brentuximab vedotin treatment group is statistically significantly higher than that of the physician's choice treatment group in the primary efficacy analysis (refer to Section 5.8.1.1).

A fixed sequential testing procedure will be used to test between the primary endpoint and the key secondary endpoints; that is, the key secondary endpoints will be tested only if the primary endpoint is statistically significant.

To test the key secondary endpoints, a weighted Holm procedure[1] will be used. The weighted Holm testing procedure is applied at the final analysis at the 0.05 level (2-sided) with weights: 0.7, 0.2, and 0.1 for CR, PFS, and the symptom domain of the Skindex-29, respectively. Once the unadjusted 2-sided p-values $p_i, i = 1,2,3$ are obtained for each endpoint, q-values $q_i, i = 1,2,3$ are calculated by dividing the p-values with their respective weights $q_i = \dfrac{p_i}{c_i}$, where c_i is the weight for endpoint i, $i = 1,2,3$. Then the ordered q-values, $q_{[1]} \leq q_{[2]} \leq q_{[3]}$, are compared with 0.05 divided by the sum of weights corresponding to that q-value and any larger q-values in a step-down fashion starting with the smallest q-value. Specifically, we identify the smallest i such that $q_{[i]} > \dfrac{0.05}{\sum\limits_{j=i}^{3} c_{[j]}}$, and claim significance for endpoints corresponding to $q_{[1]}, ..., q_{[i-1]}$. If $i = 1$, then no endpoint is significant, and if i does not exist, then all endpoints are significant.

Assuming that the probabilities of success for the endpoints are 0.85, 0.55, and 0.50 with a 0.05 (2-sided) individual alpha level and a multivariate normal distribution with an equal correlation of 0.5 among all endpoints, the power to detect at least 1 significant endpoint is 84%, and the powers to show significance for CR, PFS, and the symptom domain of the Skindex-29 using the weighted Holm method are 82%, 48%, and 40%, respectively.

The weighted Holm testing procedure controls the type I error unconditionally. The impact on power has been evaluated through simulation for a range of correlation and/or success probability assumptions. Holding the success probabilities as assumed, the power to detect at least 1 significant endpoint decreases from 89% to 81% when correlation between endpoints increases from 0.1 to 0.9. Holding the correlation as assumed, the power to detect at least 1 significant endpoint increases from 79% to 88% when probabilities of success for all 3 endpoints increases from 0.8, 0.5, and 0.45 to 0.9, 0.6, and 0.55, respectively.

Primary and key secondary endpoints are tested at the 0.05 level, and the key secondary endpoints are gated by the primary endpoint. Due to the sequential procedure property, the overall type I error rate for testing the primary and key secondary endpoints is strongly controlled at the 0.05 level.

4. INTERIM ANALYSIS

4.1 Interim Analysis

There will be no interim efficacy analysis.

4.2 Independent Data Monitoring Committee

An independent data monitoring committee (IDMC) will be established to review safety data periodically per the IDMC charter. The first formal safety review will occur after approximately 30 patients have been randomized in total and received at least 1 cycle of study treatment, or 6 months after the first patient is randomized (whichever comes first).

Detailed information regarding the composition of the IDMC and IDMC procedures will be provided in the IDMC charter.

5. STATISTICAL METHODOLOGY

In general, summary tabulations will be presented that display the number of observations, mean, standard deviation, median, minimum, and maximum for continuous variables, and the number and percentage (of nonmissing) per category for categorical data, unless specified otherwise.

5.1 Sample Size Justification

Approximately 124 patients (approximately 62 patients per treatment arm) will be randomized to the study (124 patients are targeted for the ITT population). The sample size was calculated to provide 90% power to detect a 30% improvement in ORR4 in the brentuximab vedotin treatment group, assuming ORR4 for the MTX or bexarotene alone treatment group is 40%. This calculation is based on a 2-sided χ^2 test with a significance level of $\alpha = 0.05$, and a 10% dropout rate using nQuery Advisor 7.0. A minimum of 30 patients (15 per treatment arm) with pcALCL will be included in the study.

5.2 Randomization and Stratification

The randomization scheme will be generated by Millennium. Before dosing, a randomization number will be assigned to each patient. The randomization schedule will also include the study-specific identifiers (company name, protocol name, and protocol number) and the date and time the schedule was generated.

Patients will be randomized in an overall ratio of 1:1 to Arm 1 (brentuximab vedotin) or Arm 2 (physician's choice of MTX or bexarotene) using an interactive voice response system (IVRS), stratified by baseline disease diagnosis (MF or pcALCL).

5.3   Unblinding

This is an open-label study; investigators and patients will know their individual treatment assignment. However, aggregate efficacy results by treatment group will be blinded to the study team, investigators, patients, and IRF throughout the study conduct.

5.4   Data Handling

5.4.1   Methods for Handling Missing Data

All available efficacy and safety data will be included in data listings and tabulations. The relevance of missing sample data will be assessed. Data that are potentially spurious or erroneous will be examined according to standard data management operating procedures.

In general, missing data will be treated as missing, and no data imputation will be applied unless otherwise specified.

5.4.2   Definition of Baseline Values

Unless otherwise specified, the baseline value is defined as the value collected at the time closest to, but before, the start of study drug administration.

5.4.3   Windowing of Visits

All data will be categorized based on the scheduled visit at which they were collected. These visit designators are predefined values that appear as part of the visit tab in the electronic case report form (eCRF).

5.4.4   Justification of Pooling

All data from all sites will be pooled. Study center or treatment-by-center interaction will not be included in any statistical analysis due to the rarity of the disease and potentially limited number of patients at each center.

5.4.5   Withdrawals, Dropouts, Loss to Follow-up

No dropouts will be replaced in this study.

5.5 Patient Disposition

The disposition of patients includes the number and percentage of patients for the following categories: patients treated (safety population), patients in the ITT population, patients in the All-Enrolled population, patients in the PP population, patients in the Response-Evaluable population, patients in the PK and Pharmacodynamics populations, patients completed, and patients discontinued from the study. All percentages will be based on the number of patients randomized.

The primary reason for study and treatment discontinuation will also be summarized in this table.

A listing will present data concerning patient disposition.

5.6 Demographics and Baseline Disease Characteristics

5.6.1 Demographics

Demographic and baseline characteristics will be summarized by treatment groups. Baseline demographics and baseline characteristics to be evaluated will include age, gender, race, height, weight, body surface area (BSA), primary diagnosis (stratification factor), Eastern Cooperative Oncology Group (ECOG) performance status (0, 1, 2), and other parameters as appropriate.

No inferential statistics will be generated.

5.6.2 Medical History

General medical history and prior medications will be listed for all patients.

5.7 Treatments and Medications

5.7.1 Concomitant Medications

Concomitant medications will be coded by generic term using the 2002 version of the World Health Organization (WHO) Drug Dictionary. The number and percentage of normal subjects taking concomitant medications from screening through the end of the on-study period will be tabulated by Anatomical Therapeutic Chemical (ATC) classification pharmacological subgroup and WHO drug generic term.

Concomitant procedures will not be coded but will be presented in a data listing.

5.7.2 Study Treatments

Patients will be randomized to receive study treatment, either brentuximab vedotin or bexarotene or MTX.

Brentuximab vedotin will be administered by outpatient intravenous (IV) infusion given over approximately 30 minutes on Day 1 of each 21-day cycle. The dose of brentuximab vedotin is 1.8 mg/kg. Study treatment will be administered through a dedicated IV line and cannot be mixed with other medications. Details on the dosing of study drug can be found in the protocol.

MTX will be given once weekly as a single dose of 5 to 50 mg orally. Dosages may be gradually adjusted to achieve optimal clinical response but must not exceed 50 mg/week. Once optimal clinical response is achieved, the MTX dosage should be decreased to the lowest effective dose. For patients deriving benefit, MTX dosing should be continued up to 48 weeks.

The recommended initial dose of bexarotene capsules is 300 mg/m^2/day. Bexarotene capsules should be taken as a single oral daily dose with a meal. Bexarotene capsules should be continued as long as the patient is deriving benefit, up to 48 weeks. All patients (with normal or normalized triglyceride levels) who are to be given bexarotene should receive 7 days of pretherapy with fenofibrate 145 to 200 mg daily starting from Cycle 1 Day 1 (i.e., the initial dose of bexarotene starts on Day 8 of Cycle 1).

5.7.2.1 Extent of Exposure

The exposure to study drug will be characterized by total amount of dose taken in mg, total number of dose taken, number of treated cycles, duration of treatment, numbers and percentages of patients in treated cycle categories (1, 2…,16), and dose intensity (mg/week). Dose intensity (mg/week) will be calculated as total dose administered (mg)/(3 * number of treated cycles) where a treated cycle is defined as a 21-day period during which the patient received any amount of brentuximab vedotin (scheduled for a single dose in a 21-day cycle), MTX (single weekly dose), or bexarotene (single oral daily dose). The duration of treatment is defined as time from the first study dose to 21 days after the last study dose ([last dose date + 21] – first dose date) of brentuximab vedotin, or time from the first study dose to 7 days after the last study dose ([last dose date + 7] – first dose date) of MTX, or time from the first study dose to 1 days after the last study dose ([last dose date + 1] – first dose date) of bexarotene.

Relative dose intensity (%) will be calculated as (Total Dose Administered/Total Dose Expected) * 100 and summarized by treatment group.

All extent-of-exposure data will be summarized as continuous variables for brentuximab vedotin, methotrexate, and bexarotene in the safety population. Study drug administration and exposure information will also be presented in a by-patient data listing.

5.7.2.2   Treatment Modifications

Action on study drug (including dose reduction) will be summarized by treatment group, by cycle, and overall.

5.8   Efficacy Analyses

All efficacy evaluations will be conducted using the ITT population unless specified otherwise.

5.8.1   Primary Efficacy Endpoint

The primary endpoint of this study is ORR4 per IRF. ORR4 per IRF is defined as the proportion of patients achieving an objective response that lasts at least 4 months (i.e., duration from first response to last response is $\geq$ 4 months) on study as determined by an IRF. Patients whose first response occurs after the start of subsequent anti-cancer therapy but otherwise meet the primary endpoint criteria will be excluded. For objective response, a patient must achieve a CR or PR. The objective response will be considered maintained for patients with a previous CR who experience recurrent disease (i.e. Relapse) unless the criteria for progressive disease are met. Determination of objective response will be based on a GRS, which consists of skin evaluation (mSWAT assessment) by investigator, nodal and visceral radiographic assessment by IRF, and detection of circulating Sézary cells (MF only) by IRF. Skin evaluation is done at screening, before dosing on Day 1 of Cycles 1, 2, and 3, at the end of every cycle beginning at Cycle 3, EOT, and at post treatment follow-up visits. Radiographic and blood assessments are less frequent due to ethical and logistical considerations.

For patients without nodal or visceral involvement (as determined by baseline CT scans), CT scans will be performed at screening and:

- During the cycle following the first skin response, ideally corresponding to the confirmation of the skin response, and 6 cycles (a minimum or 4 months) after that confirmed skin response.

- If there is a question of new or progressive disease in the LN/viscera

For patients with baseline nodal/visceral disease, CT scans will be performed at screening and:

- At the end of Cycles 3, 6, 9, 12, and 15, and per the follow-up schedule until disease progression

- If there is a question of new or progressive disease in the LN/viscera

- At EOT if the previous CT scan was performed more than 8 weeks before EOT

A blood sample for Sézary cell enumeration in patients with MF will be collected at screening; at the end of Cycles 3, 6, 9, 12, and 15, at EOT, and per the follow-up schedule until progressive disease or study closure.

GRS will be determined by IRF at the end of each cycle beginning at Cycle 3, EOT, and at post treatment follow-up visits until disease progression, death or study closure. The GRS assessment at each cycle will incorporate the skin response at that cycle as well as the most recent nodal, visceral, and blood responses prior to or at the time of the GRS assessment time point. CR or PR per GRS will be confirmed by sustained response in skin per mSWAT assessment at the subsequent cycle.

For patients with skin only disease at baseline, GRS is essentially determined by skin response because disease progression in node/viscera/blood is atypical in the presence of a continuing skin response. In addition, the primary endpoint is further supported by CT confirmation of continued nodal/visceral noninvolvement at the cycle after the first skin response and 6 cycles after the first confirmed skin response. For patients with baseline nodal/viscera/blood involvement, GRS assessment at each cycle will incorporate the skin response at that cycle as well as nodal, visceral, and blood responses based upon the most recent respective compartment assessments prior to the GRS assessment time point. Blood results should be within 3 cycles. The CT scans should also be within 3 cycles for patients with nodal/viscera involvement.

5.8.1.1 Primary Efficacy Analysis

The primary endpoint, ORR4 per IRF, will be analyzed using a Cochran-Mantel-Haenszel test stratified by baseline disease diagnosis (pcALCL or MF) based on the ITT population.

The 95% confidence interval (CI) of the difference in ORR4 between the 2 treatments will be provided. No imputation will be conducted for missing data. Patients who do not have any post baseline response assessment as specified in the protocol, or patients who have no response before dropout, will be counted as nonresponders.

Sensitivity analyses will be performed for ORR per IRF in the study as well as ORR4 per IRF defined by the response criteria used for Romidepsin approval and published by Whittaker 2010.[2] Analysis for ORR4 per IRF defined by the response criteria used for Romidepsin approval will also be performed for MF patients only. Sensitivity analyses will be performed for ORR and ORR4 based on investigator's GRS assessment as well.

Another sensitivity analysis will be conducted for ORR4 per IRF in the ITT population to assess the robustness of the primary analysis's use of most recently available CT scans and blood assessments for GRS. For patients with baseline nodal/visceral/blood disease, the sensitivity analysis will only use GRS assessed at time points with concurrent nodal/viscera/blood component assessments as specified in the protocol SOE while for patients with skin-only disease, the analysis will use the same GRS assessment frequency as the primary analysis.

5.8.1.2 Other Analysis

The same analysis on ORR4 per IRF using a Cochran-Mantel-Haenszel test stratified by baseline disease diagnosis (pcALCL and MF) will be performed for the All-Enrolled population, Response-Evaluable population, as well as the PP population. The 95% CI of the difference in ORR4 between the 2 treatments will be provided. No imputation will be conducted for missing data. Patients who do not have any post baseline response assessment as specified in the protocol, or patients who have no response before dropout, will be counted as nonresponders.

Similarly, analyses will also be performed for the primary endpoint on, but not limited to, the following subgroups (with at least 10 patients in each subgroup per treatment arm), as applicable:

Subgroup	Group Definition
Baseline disease diagnosis	MF, pcALCL
ECOG performance status	0, 1, 2
Gender	Female, male
Age	< 65 years, ≥ 65 years
Region	North America, Europe, Asia, and rest of world
Race	White, nonwhite
Physician's choice (bexarotene)	Brentuximab vedotin vs bexarotene
Physician's choice (MTX)	Brentuximab vedotin vs MTX

Concordance between the IRF and investigator's GRS assessment for ORR4 will be assessed by percent positive, negative, and overall agreement.

CCI

5.8.2   Key Secondary Efficacy Endpoints

CR rate, PFS, and changes in symptom domain per Skindex-29 assessment tool are designated as key secondary endpoints.

CR rate is defined as the proportion of patients who achieved a CR as their best response on study as determined by an IRF. A sensitivity analysis will be performed for CR per the investigator's GRS assessment. PFS is defined as the time from randomization until PD per IRF or death due to any cause, whichever occurs first. The symptom domain of Skindex-29 includes 7 items: 1) My skin hurts; 2) My skin condition burns or stings; 3) My skin itches; 4) Water bothers my skin conditions (bathing, washing hands); 5) My skin is irritated; 6) My skin is sensitive; and 7) My skin condition bleeds. The following scoring system is used for all items: never (0), rarely (1), sometimes (2), often (3), and always (4). The total score for all 7 items will be converted to a scale of 0 (no impact on quality of life) to 100 (maximum impact on quality of life) by linear transformation. The questionnaire is administered on Day 1 of Cycles 1, 2, 4, 6, 8, 10, 12, 14, and 16.

5.8.2.1   Analysis on Key Secondary Efficacy Endpoint

5.8.2.1.1   Handling of Missing Assessment and Censoring for Analysis of
Progression-free Survival

Disease assessment data should be collected according to the intended schedule of assessment, and the date of PD/response should be assigned based on the time of the first documentation of PD/response regardless of violations or discontinuation of study drug.

For the PFS endpoint, patients who are lost to follow-up, withdraw consent, or those who discontinue treatment due to undocumented PD after the last adequate disease assessment will be censored at the last disease assessment. If death or PD occurs after a missed visit, then the patient is treated as progressed at the date of death or PD. Patients without baseline and/or no sufficient post baseline data for disease assessment and with no death recorded will be censored at the date of randomization. If PD is documented between scheduled visits, then the date of the documented PD is the date of progression. If the patient starts new antineoplastic therapy before PD, then the patient is treated as progressed at the date of assessment at which PD was documented.

The following sensitivity analyses will be performed for PFS to evaluate the robustness of treatment effects. They will be performed on the basis of 1 alteration at a time, not on combined alterations, unless specified otherwise.

- If the PD is documented between scheduled visits, then the date of the next scheduled visit is the date of progression.

- If treatment is discontinued for undocumented PD, then the patient is treated as progressed at the date of last disease assessment.

- Patients who start new antineoplastic therapy before PD will be censored at the date of the last disease assessment.

- Patients who start new antineoplastic therapy before PD will be treated as progressed at the date of the last disease assessment.

- If death or PD occurs after more than 1 missed visit, then the patient will also be censored at the last disease assessment before the missed visits.

Another sensitivity analysis will be conducted for PFS per IRF in the ITT population to assess the robustness of the primary analysis's use of most recently available CT scans and

blood assessments for GRS. For patients with baseline nodal/visceral/blood disease, the sensitivity analysis will only use GRS assessed at time points with concurrent nodal/viscera/blood component assessments as specified in the protocol SOE while for patients with skin-only disease, the analysis will use the same GRS assessment frequency as the primary analysis.

5.8.2.1.2   Key Secondary Efficacy Analysis

Because a fixed sequential testing procedure is used between the primary endpoint and the key secondary endpoints, the key secondary endpoints will be tested only if the primary endpoint is statistically significant. The weighted Holm procedure[1] will be used for testing the key secondary endpoints. See Section 3.2 for details of the statistical decision rules.

Stratified log-rank test statistics will be used to compare PFS between the 2 treatment arms. The hazard ratios along with the 95% CIs will be estimated using a stratified Cox regression model. The Kaplan-Meier method will be used to estimate the distribution of the time-to-event endpoints for each treatment. Median times (if estimable), along with the 2-sided 95% CIs, will be presented.

Assuming the median PFS for the brentuximab vedotin treatment group and the physician's choice treatment group (MTX or bexarotene) are 9 and 6 months, respectively, the number of PFS events is projected to be approximately 101 at the time of the planned final analysis. This is based on an approximately 2-year enrollment period, a dropout rate of 10%, and an α level of 0.01 (2-sided), as specified in Section 3.2 per the weighted Holm procedure.

Comparison of the CR rates between the 2 treatment groups will be conducted using the stratified Cochran-Mantel-Haenszel test. The 95% CI of the difference in the response rates between the 2 treatments will also be provided. No imputation will be conducted for missing data. Patients who do not have any post baseline response assessment as specified in the protocol will be counted as nonresponders.

For skin symptoms (Skindex-29), the primary analysis will be comparing the mean symptom reduction between brentuximab vedotin and the control arms. 'Symptom reduction' is defined for each patient as the maximum reduction from baseline (sum of the 7 items related to skin symptoms normalized to a 1-100 scale). The analysis of covariance (ANCOVA) model controlling for baseline covariates (treatment group, baseline score, disease diagnosis, ECOG) will be employed for the analysis. As a sensitivity analysis, comparison of the 2 treatment groups will be conducted using the linear mixed model with repeated measures at

each time point specified in the protocol SOE. The estimation of variance–covariance matrix and statistics such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) will be included in evaluating the linear mixed model. The 95% CIs of the difference of the changes from baseline between the 2 treatments will also be provided.

The total score for skin symptoms (Skindex-29) will be imputed with the mean of the other items if there is no more than 1 missing item; otherwise, it will be considered invalid and excluded from the analysis.

In addition, the entire distribution of change in Skindex-29 symptom scores at the individual patient level will be presented for each arm, respectively. The patient's largest score of change from baseline will be determined first, and then a cumulative distribution function (CDF) curve will be used to display the cumulative percentage of patients experiencing that change. It is noteworthy that the higher scores in Skindex-29 indicate worse levels of symptoms; therefore, the patient's largest score of change from baseline needs to be presented in the format of "reduction in scores" (e.g., a -10 of change needs to be presented as +10 in reduction) on the X-axis. Categorical definitions of responders or minimum important difference (MID) will be determined by scrutinizing these CDF curves.

Primary analyses of key secondary endpoints will be performed based on the ITT population. Sensitivity analyses will be performed for CR using similar statistical analyses based on the Response-Evaluable population.

5.8.3   Other Secondary Efficacy Endpoints

Other secondary endpoints include EFS, DOR, and duration of skin response. EFS is defined as the time from randomization until any cause of treatment failure: PD, discontinuation of treatment for any reason, or death due to any cause, whichever occurs first. DOR in subjects with a confirmed response is the time between first documentation of response and PD. Duration of skin response is the time between the first skin response and PD in skin.

5.8.3.1   Other Secondary Efficacy Analysis

The ITT principle will be used to determine the event time or censoring time for the EFS analysis. Disease assessment data should be collected according to the intended schedule of assessment, and the date of progression should be assigned based on the time of the first documentation of PD regardless of violations or discontinuation of study drug.

For EFS, patients who are lost to follow-up will be censored at last disease assessment. Patients who withdraw consent or start new antineoplastic therapy will be treated as if experiencing an EFS event. Patients without baseline and/or sufficient post baseline data for disease assessment and no treatment discontinuation or death recorded will be censored at the date of randomization. Analyses of EFS will be performed based on the ITT population.

DOR will be analyzed for the subgroup of patients with confirmed response (CR or PR) in the ITT population. Patients who are lost to follow-up, withdraw consent, or discontinue treatment due to undocumented PD after the last adequate disease assessment will be censored at the last disease assessment. If the patient starts new antineoplastic therapy before PD, the patient is treated as progressed at the date of assessment at which PD was documented.

Duration of skin response will be analyzed for the subgroup of patients with skin response (CR or PR in skin) in the ITT population. Patients who are lost to follow-up, withdraw consent, or discontinue treatment due to undocumented PD in skin after the last adequate skin assessment will be censored at the last skin assessment. Patients without sufficient skin assessment data after the initial skin response and with no death recorded will be censored at the date of the initial skin response. If the patient starts new antineoplastic therapy before skin PD, then the patient is treated as progressed at the date of the skin assessment at which PD in skin was documented.

EFS will be analyzed similarly as the key secondary efficacy endpoint PFS as described in Section 5.8.1. DOR and duration of skin response will be summarized descriptively using the Kaplan-Meier method.

5.9 Pharmacokinetic, Pharmacodynamic, and Biomarker Analysis

5.9.1 Pharmacokinetic Analyses

The PK of the antibody drug-conjugate (brentuximab vedotin), total antibody, and unconjugated drug (MMAE) will be derived from serum concentration versus time data for all patients who met study inclusion criteria, received study drug, and provided evaluable PK data. Population PK parameters will be calculated with an appropriate method based on a validated PK analysis program.

CCI .

5.9.2 Pharmacodynamic Analyses

Absolute and change from baseline values of circulating pharmacodynamic biomarkers such as soluble CD30 (sCD30), thymus and activation regulated chemokine (TARC or CCL17), and CTACK (CCL27) will be summarized by time point using descriptive statistics (e.g., number of patients, mean, standard deviation, median, minimum, maximum), as applicable. Descriptive statistics will also be provided to summarize biopsy pharmacodynamic parameters, such as pHisH3, and change from baseline values, as applicable.

5.9.3 Immunogenicity Analysis

All patients who were administered at least 1 dose of brentuximab vedotin will be evaluated for antitherapeutic antibody (ATA) development. A list/table of ATA status will be provided. Antibody neutralizing status (neutralizing or not neutralizing) will also be listed for patients who have positive antibody status.

Immunogenicity information, including ATA and neutralizing ATA, will be summarized in descriptive statistics as applicable.

The relationship between antibody status and efficacy and/or safety will be explored.

5.9.4 Biomarker Analysis

CCI

5.10 Analyses of Patient-Reported Outcomes and Health Economics

PRO endpoints include subscales of Skindex-29 and global and subscale scores of the EQ-5D and FACT-G questionnaires. The health economic endpoint includes medical resource utilization (MRU) enumerations. Furthermore, the symptom domain of Skindex-29 is one of the key secondary endpoints. Its analysis is described separately in Section 5.8.2.1.2. The analyses of the remaining PRO and MRU endpoints and the additional analysis of Skindex-29 symptom scores are described below.

Analyses of PROs and health economics will be performed using the ITT population.

5.10.1 Patient-Reported Outcomes Analysis

All PRO scores will first be scaled into numeric scores following published or prespecified scoring guidelines for each PRO instrument employed in this trial. The maximum allowed number of missing items to adequately measure a domain is instrument specific and found in the relevant guidelines. Missing pattern of PRO data (noncompliance, missing, dropout) will be examined and summarized. Additional imputation methods, such as last observation carried forward, or multiple imputation may be considered if appropriate.

For Skindex-29 (global, symptoms, emotions, and functioning), EQ-5D (global, EQ-5D descriptive system and the EQ visual analogue scale), and FACT-G (global, physical well-being, social/family well-being, emotional well-being, and functional well-being), scores will be summarized in descriptive statistics for the 2 treatment groups over time.

5.10.2 Health Economics Analysis Using Medical Resource Utilization

MRU data are collected through the medical encounter page of the eCRF.

MRU data will be summarized in descriptive statistics of hospitalization (length of stay, inpatient, outpatient, and reason), number of missing days from work or other activities by subject, and care-giver for the 2 treatment groups.

5.11 Safety Analyses

Safety evaluations will be based on the incidence, severity, type of adverse events (AEs), clinically significant changes, or abnormalities in the subject's physical examination, vital signs, ECOG performance status, and clinical laboratory results.

These analyses will be performed using the safety population.

5.11.1 Adverse Events

5.11.1.1 Adverse Events

Treatment-emergent AEs (TEAEs) will be tabulated by primary System Organ Class (SOC), High Level Term (HLT), and Preferred Term (PT). A TEAE is defined as any AE that occurs after administration of the first dose of study drug and up through 30 days after the last dose of study medication. AEs will be tabulated according to the most recent version (Version 8.1 or higher) of the Medical Dictionary for Regulatory Activities (MedDRA) by SOC, HLT, and PT and will include the following categories:

- TEAEs

- Study drug-related TEAEs

- Grade 3 or higher TEAEs

- Grade 3 or higher drug-related TEAEs

- TEAEs resulting in study drug discontinuation

- Serious AEs (SAEs), including study drug-related SAEs

- AEs of peripheral neuropathy identified by the broad search MedDRA SMQ "Peripheral neuropathy"

Additional analyses of peripheral neuropathy may also be presented.

The most commonly reported TEAEs (i.e., those events reported by ≥ 10% of all patients) will be tabulated by HLT and PT.

5.11.1.2   Serious Adverse Events

The number and percentage of subjects experiencing at least 1 treatment-emergent SAE will be summarized by MedDRA primary SOC, HLT, and PT.

In addition, a by-subject listing of the SAEs will be presented (the subject listing will contain all SAEs regardless of treatment-emergent status).

5.11.1.3   Deaths

A by-subject listing of the deaths will be presented. All deaths occurring on-study and during follow-up will be displayed (regardless of treatment-emergent status).

5.11.1.4   Adverse Events Resulting in Discontinuation of Study Drug

A by-subject listing of AEs resulting in discontinuation of study drug will be presented.

5.11.2   Laboratory Data

For the purposes of summarization in both the tables and listings, all laboratory values will be converted to standardized units. If a lab value is reported using a nonnumeric qualifier

(e.g., less than (<) a certain value, or greater than (>) a certain value), the given numeric value will be used in the summary statistics, ignoring the nonnumeric qualifier.

If a subject has repeated laboratory values for a given time point, the value from the last evaluation will be used.

The parameters to be analyzed are as follows:

- Hematology: hemoglobin, hematocrit, platelet count, neutrophils (absolute neutrophil count [ANC]), lymphocytes, monocytes, eosinophils, basophils, and leukocyte counts

- Clinical chemistries: blood urea nitrogen (BUN), creatinine, bilirubin (total), urate, lactate dehydrogenase (LDH), phosphate, gamma-glutamyl-transferase (GGT), albumin, alkaline phosphatase (ALP), aspartate aminotransferase (AST), alanine aminotransferase (ALT), glucose, sodium, potassium, calcium, chloride, CO_2, magnesium, amylase, and lipase

- Urinalysis: turbidity and color, pH, specific gravity, protein, ketones, bilirubin, blood, nitrite, urobilinogen, glucose, leukocytes

- Lipid panel: total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides

- Thyroid panel: thyroid-stimulating hormone (TSH), triiodothyronine (T3), and thyroxine (T4)

- Immune reconstitution assessments: lymphocyte counts, phenotypic T, B, and natural killer (NK) cell subsets ($CD4^+$ T cell counts, $CD8^+$ T cell counts, CD8/CD4 ratio, $CD19^+$ B cell counts), and NK cells ($CD3^-$ CD56 $CD16^+$ vs$CD3^-$ CD56 CD16$^-$)

- Other: hemoglobin A1C

The actual values in clinical laboratory parameters (hematology and serum chemistry, including lipid panel and thyroid panel) will be summarized at baseline and over time by treatment group; summary statistics for change from baseline in laboratory parameters will also be presented.

In addition, hematology and serum chemistry parameters will be summarized in shift tables comparing the baseline visit to the worst post baseline value (National Cancer Institute [NCI] Common Terminology Criteria for Adverse Events [CTCAE] grade, version 4.03, effective 14 June 2010[3]) for the following parameters.

- Hematology: hemoglobin, , platelet count, neutrophils (absolute neutrophil count [ANC]), lymphocyte counts, and leukocyte counts

- Clinical chemistries: bilirubin (total), alkaline phosphatase (ALP), aspartate aminotransferase (AST), alanine aminotransferase (ALT), amylase, and lipase

Urinalysis results will be presented in a listing.

5.11.3   Electrocardiograms

Electrocardiogram results will be presented in a listing.

5.11.4   Vital Signs

The actual values of vital sign parameters, including blood pressure (determined with the patient in a seated position after the patient has been sitting quietly for 5 minutes), heart rate, and oral temperature, when available, will be summarized over time.

5.11.5   Eastern Cooperative Oncology Group Performance Status

Eastern Cooperative Oncology Group performance status and change from baseline will be summarized. Shifts from baseline to the worst postbaseline score will be tabulated by treatment group. Patients who have ECOG performance scores that worsen postbaseline compared to baseline will be listed with other pertinent patient information.

5.11.6   Other Safety Assessments

Pregnancy testing results will be presented in a by-patient listing.

Additional safety analyses may be performed to most clearly enumerate rates of toxicities and to further define the safety profile of study drugs.

6.   CHANGES TO PLANNED ANALYSES FROM PROTOCOL

Not applicable

7. PROGRAMMING CONSIDERATIONS

7.1 Statistical Software

SAS version 9.1 (or higher) will be used for all analyses.

7.2 Rules and Definitions

Subject populations are defined in Section 2.

Baseline values are defined in Section 5.4.2.

Treatment-emergent AEs are defined in Section 5.11.1.1.

8. REFERENCES

1.   Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 1979;6:65-70.

2.   Whittaker SJ, Demierre MF, Kim EJ, Rook AH, Lerner A, Duvic M, et al. Final results from a multicenter, international, pivotal study of romidepsin in refractory cutaneous T-cell lymphoma. Journal of Clinical Oncology 2010;28(29):4485-91.

3.   U.S. Department of Health and Human Services, National Institutes of Health National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE). Version 4.03. 14 June 2010.