Title: Randomized, Open-label, Phase 3 Trial of A+AVD Versus ABVD as Frontline Therapy in Patients With Advanced Classical Hodgkin Lymphoma

NCT Number: NCT01712490

SAP Approve Date: 10 March 2015

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

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

CLINICAL STUDY PROTOCOL C25003
Brentuximab vedotin
A Randomized, Open-label, Phase 3 Trial of A+AVD Versus ABVD as Frontline
Therapy in Patients With Advanced Classical Hodgkin Lymphoma

Protocol #: C25003

SAP Version: Date of Statistical Analysis Plan:
Draft   22 May 2012
Draft   10 March 2015

Approval Signatures

_____   Prepared by: PPD
Date   PPD   Clinical
                                  Biostatistics

_____   Reviewed by: PPD
Date   PPD   Clinical Biostatistics

_____   Approved by: PPD
Date   PPD   , Oncology Clinical
                                  Research

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND GLOSSARY OF TERMS

Abbreviation	Term
A	ADCETRIS™
ABVD	doxorubicin [Adriamycin], bleomycin, vinblastine, dacarbazine
AE	adverse event
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANC	absolute neutrophil count
AUC	area under the plasma concentration versus time curve
AVD	doxorubicin [Adriamycin], vinblastine, dacarbazine
A+AVD	doxorubicin [Adriamycin], vinblastine, dacarbazine, brentuximab vedotin [ADCETRIS™]
BPT	bleomycin pulmonary toxicity
BSA	body surface area
C_{max}	maximum plasma concentration
CO_2	carbon dioxide
CR	complete response
CR(u)	unconfirmed complete response
CT	computed tomography
DNA	deoxyribonucleic acid
DOCR	duration of complete response
DOR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EDC	electronic data capture

Abbreviation	Term
EOS	End of Study (visit)
EOT	End of Treatment (visit)
ICF	informed consent form
ICH	International Conference on Harmonisation
IDMC	independent data monitoring committee
IPFP	International Prognostic Factor Project
IPS	[Hasenclever] International Prognostic Score
IRB	institutional review board
IRF	independent review facility
ITT	intent-to-treat
IV	intravenous; intravenously
IVRS	interactive voice response system
MedDRA	Medical Dictionary for Regulatory Activities
Millennium	Millennium Pharmaceuticals, Inc., and its affiliates
mPFS	modified progression-free survival
MRI	magnetic resonance imaging
MRU	medical resource utilization
NCI	National Cancer Institute
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
OS	overall survival
PD	progressive disease (disease progression)
PFS	progression-free survival
PK	pharmacokinetic(s)
PR	partial response
PRO	patient-reported outcome
QALY	quality-adjusted life year
QOL	quality of life
QTc	rate-corrected QT interval (millisec) of electrocardiograph
SAE	serious adverse event
SD	stable disease
$t_{1/2}$	half-life
T_{max}	first time to maximum plasma concentration
ULN	upper limit of the normal range
V_z	volume of distribution in the terminal phase
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 will 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 open-label, randomized, 2-arm, multicenter, phase 3 study has the primary objective of comparing the modified progression-free survival (mPFS) obtained with A+AVD against that obtained with ABVD.

Approximately 1240 patients will be randomized 1:1 into 2 treatment arms:

- A+AVD: Doxorubicin 25 mg/m^2, vinblastine 6 mg/m^2, dacarbazine (DTIC) 375 mg/m^2, brentuximab vedotin 1.2 mg/kg

- ABVD: Doxorubicin 25 mg/m^2, bleomycin 10 units/m^2, vinblastine 6 mg/m^2, dacarbazine (DTIC) 375 mg/m^2

ABVD and A+AVD will be administered intravenously on Days 1 and 15 of each 28-day cycle. Brentuximab vedotin will be administered intravenously over 30 minutes at a dose of 1.2 mg/kg; the brentuximab vedotin infusion is to be started within approximately 1 hour after completion of AVD therapy. Patients may receive up to 6 cycles of therapy (ABVD or A+AVD).

Randomization will be stratified by region (Americas versus Europe versus Asia), and number of International Prognostic Factor Project(IPFP) risk factors (0-1 versus 2-3 versus 4-7).

1.2 Study Objectives

Primary Objective:

- To compare the modified progression-free survival (mPFS) obtained with brentuximab vedotin plus AVD (abbreviated A+AVD) versus that obtained with ABVD for the frontline treatment of advanced classical HL

Key secondary objective:

- To determine if A+AVD improves overall survival (OS) versus that obtained with ABVD

Other secondary objectives:

- To determine if A+AVD improves complete remission (CR) rate at the end of randomized regimen versus that obtained with ABVD

- To determine the safety profile of A+AVD relative to that of ABVD

- To determine the event-free survival (EFS) obtained with A+AVD and ABVD

- To determine the disease-free survival (DFS) obtained with A+AVD and ABVD

- To determine if A+AVD improves overall objective response rate (ORR, defined as CR + PR) versus that obtained with ABVD

- To determine the duration of response (DOR) and duration of complete response (DOCR) obtained in the A+AVD and ABVD arms

- To determine the rate of patients receiving irradiation for HL not in CR in the A+AVD and ABVD arms

- To determine the rate of patients in CR at the end of frontline therapy in the A+AVD and ABVD arms

- To determine the rate of Cycle 2 PET negativity in patients treated with A+AVD versus those treated with ABVD

- To determine if A+AVD improves health-related quality of life (HRQoL) versus ABVD

- To describe the PK of brentuximab vedotin, MMAE, and total antibody (TAb) in blood

- To determine the immunogenicity of brentuximab vedotin

Exploratory objectives:

- To investigate any differences in lung-specific patient reported outcomes (PRO) between the treatment arms

- To assess any impact of brentuximab vedotin dosing on serum concentrations of AVD

- To investigate any differences between the treatment arms in the rate of patients alive without HL at 3 and 5 years

- To assess changes in tumor biomarker expression before and after treatment

- CCI

- To assess other PROs

- To assess medical resource utilization

- To assess fertility

2. POPULATIONS FOR ANALYSIS

2.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population will include all patients 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 all efficacy analyses of efficacy endpoints unless specified otherwise.

2.2   Per-Protocol Population

The Per-Protocol (PP) population will include all randomized patients who do not have major protocol violation as determined by the project clinician. All decisions to exclude patients from the PP population will be made prior to 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. The PP population will be used as supportive analysis for the primary efficacy endpoint mPFS.

2.3   Response-Evaluable Population

The response-evaluable population is defined as the subset of the ITT population with diagnosis as confirmed by an independent pathology review facility, with measurable disease at baseline, who receive at least 1 dose of study drug, and have at least 1 postbaseline response assessment. The response-evaluable population will be used for the analyses of CR rate, overall response rate, and duration of response.

2.4   Safety Population

The safety population will include all enrolled patients who have received at least 1 dose of study medication. All patients in the safety population will be analyzed according to the actual treatment received.

All safety analyses will be performed using the safety population.

2.5   Pharmacokinetics Population

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

2.6 Pharmacodynamics Population

The pharmacodynamics population will include enrolled patients with sufficient dosing and sufficient pharmacodynamics data to reliably measure pharmacodynamics parameters. The pharmacodynamics population will be used for pharmacodynamics analyses.

3. HYPOTHESES AND DECISION RULES

3.1 Statistical Hypotheses

Primary hypothesis to be tested:

> The primary null hypothesis is that there is no difference in modified progression-free survival (mPFS) between the 2 treatments of A+AVD and ABVD. The alternative hypothesis is that A+AVD improves mPFS.

Key secondary hypothesis to be tested:

> The null hypothesis is that there is no difference in overall survival (OS) between the 2 treatments of A+AVD and ABVD. The alternative hypothesis is that A+AVD improves OS.

Hypotheses for the secondary endpoints will also be tested.

3.2 Statistical Decision Rules

3.2.1 Testing Significance for Primary Endpoint

Modified PFS will be tested at a 1-sided significance level of 0.025. We will conclude that the A+AVD treatment can improve mPFS compared with the ABVD treatment if the test is statistically significant.

3.2.2 Testing Process and Significance for Key Secondary Efficacy Endpoint

Key secondary endpoint will be tested at 1-sided, 0.025 level only when the test of the primary endpoint (mPFS) is statistically significant. If the test is statistically significant, we will conclude that the A+AVD treatment can make improvement in OS compared with the ABVD treatment.

4.   INTERIM ANALYSIS

4.1   Interim Analysis

There will be 2 formal interim analyses in this study.

The first formal interim analysis to be performed is a futility analysis. The CR rate at the end of frontline therapy will be analyzed when the first approximately 348 patients have completed the regimen to which they were randomized (ie, received the planned study drug regimen with no more than 2 missed doses of A+AVD or ABVD) or have discontinued treatment prior to completion. Recommendation by the independent data monitoring committee (IDMC) whether to terminate the study based on this interim analysis will be determined upon evaluation of the overall safety information and efficacy data, specifically if the CR rate per independent review facility (IRF) for the A+AVD arm is at least 5% lower than that of the ABVD arm and trends in mPFS and other efficacy endpoints suggest inferior efficacy in the A+AVD arm. This futility analysis does not require adjustment of the type I error.

The second formal interim analysis is for OS to be performed at the time of the final mPFS analysis. Overall type-I error for OS will be controlled using the O'Brien-Fleming method with a Lan-DeMets alpha spending function, with final OS analysis scheduled for when 112 deaths have occurred.

4.2   Independent Data Monitoring Committee (IDMC)

An IDMC will review safety and efficacy data at the interim analyses. The IDMC will provide a recommendation regarding study continuation based on the safety and efficacy parameters. In the event that the study is terminated early based on an IDMC recommendation, Millennium will notify the appropriate regulatory authorities. Detailed information regarding the composition of the IDMC and detailed IDMC procedures will be provided in the IDMC charter.

The first formal safety review will occur after the first 100 patients have completed 2 cycles (8 weeks) of treatment or discontinued prior to completing 2 cycles of treatment. Subsequently, IDMC safety reviews will be performed periodically per 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 percent (of non-missing) per category for categorical data, unless specified otherwise.

5.1 Sample Size Justification

The primary endpoint of the study is mPFS, and the study is powered on the following assumption: a 2-year mPFS of 81% for patients in the A+AVD treatment group versus 73% for patients in the ABVD treatment group (HR = 0.67, assuming an emergent plateau in the PFS event rate after 2 years). A total of 260 mPFS events will provide 90% power to detect a hazard ratio of 0.67 at a 1-sided significance level of 0.025 using a log-rank test. Approximately 1240 patients will be randomized to achieve (with 95% probability) 260 mPFS events in about 60 months assuming 36 months of accrual, a 5% annual dropout rate, and 24 months of mPFS follow-up after last patient in.

5.2 Randomization and Stratification

The randomization scheme will be generated by Millennium. Prior to dosing, a randomization number will be assigned to each patient. The randomization schedule also includes 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 A+AVD or ABVD. Patients will be stratified by region (Americas versus Europe versus Asia), and number of IPFP risk factors (0-1 versus 2-3 versus 4-7).

5.3 Unblinding

This is an open-label study; investigators and patients will know the individual treatment assignments. However, aggregate efficacy data will be blinded to the sponsor's study team, investigators, and patients throughout the study conduct. The IRF will be blinded to treatment assignments.

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. Data that are potentially spurious or erroneous will be examined under the auspices of standard data management operating procedures.

In general, missing data will be treated as missing and no data imputation will be applied, unless otherwise specified. For Quality of Life Data, missing elements may be substituted with the average of non-missing items per published methods of analysis.

Last observation carried forward method and multiple imputation method may be considered for some clinical outcomes as deemed appropriate.

5.4.1.1   Missing/Partial Dates in Screening Visit

The following rules apply to dates recorded in the screening visits, with the exception of prior therapies (Section 5.4.1.2).

- If only the day-component is missing, the first day of the month will be used if the year and the month are the same as those for the first treatment. Otherwise, the 15th will be used.

- If only a year is present, and it is the same as the year of the first treatment, the 15th of January will be used unless it is later than the first treatment, in which case the date of the first of January will be used.

- If only a year is present, and it is not the same as the year of the first treatment, the 15th of June will be used, unless other data indicates that the date is earlier, in which case the 15th of January will be used. .

5.4.1.2   Missing/Partial Dates in Adverse Events/Concomitant Therapies/Subsequent Therapies

Every effort will be made to avoid missing/partial dates in on-study data. If the resolution date of a resolved adverse event (AE) or the stop date of a concomitant therapy is missing, the following rules are to be used unless conflicting data exists: if month and year are present and the day of the month is missing, the last day of the month is imputed. If only a

year is present, the 31st of December is used. After imputation, the imputed dates will be compared against the date of death, if available. If the date is later than the date of death, the date of death will be used as the imputed date instead.

In cases where the onset date of an adverse event is completely or partially missing, the following imputation rules will be used:

1. When month and year are present and the day of the month is missing,

   - If the onset month and year are the same as the month and year of first treatment with study drug, the day of first treatment or the day-component of the resolution date are imputed, whichever is earliest.

   - If the onset month and year are not the same as the month and year of first treatment with study drug, the first day of the month is imputed.

2. When only a year is present, or no components of the onset date are present,

   - If the resolution date is available, the earlier of the resolution date (possibly imputed) and the date of first treatment will be used.

   - If the resolution date is missing, and the onset-year is the same as the year of first treatment with study drug, then the date of first treatment with study drug is used.

   - Otherwise if only a year is present, the 1st of January of that year is imputed.

3. If none of the previous rules can be applied, then the date of first treatment with study drug is imputed as the onset date.

The imputation rules for missing/partial start dates of concomitant therapies will be the same as the above with the exception as follows:

For prior therapy data, no imputation will be done for start dates.

The imputation rules for missing/partial start dates of subsequent therapies recorded as concomitant medications will be the same as the above with exceptions as follows.

1. When month and year are present and the day of the month is missing,

a. If the month and year of the start date are the same as the month and year of treatment termination, the day of treatment termination or the day-component of the stop date is imputed, whichever is earliest.

b. If the start month and year are not the same as the month and year of treatment termination, the first day of the month is imputed.

2. When only a year is present, or no components of the start date are present, the date will not be imputed.

5.4.1.3   Lab Values Below the Lower Limit of Quantification

If the numeric value of a laboratory test (excluding PK analyses) is not available because it is below the lower limit of quantification (LLOQ), a logical value of < LLOQ should be used whenever applicable. In cases where a numeric value is required, eg, calculating the mean, the LLOQ will be used.

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 prior to, the start of study drug administration. If a patient is not dosed, the date of randomization will be used.

In the case that laboratory values from both the central lab and local lab are available before randomization, the most recent value from the central lab will be used as baseline, even if the sample date of the central lab result is prior to that of any local lab value.

5.4.3   Windowing of Visits

All data will be categorized based on the scheduled visit at which it was collected. These visit designators are predefined values that appear as part of the visit tab in the 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.

5.4.5   Withdrawals, Dropouts, Lost to Follow-up

No patients will be replaced in this study.

5.5 Patient Disposition

A disposition of patients includes the number and percentage of patients for the following categories: patients in each of the study population, primary reason to discontinue from the treatment, patients discontinued from the study, and primary reason to discontinue from the study. All percentages will be based on the number of patients in the ITT population.

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, including gender, age, race, weight, height, BSA, primary diagnosis, and other parameters as appropriate. No inferential statistics will be carried out.

The formulation for BSA is:

BSA = sqrt(height(cm) x weight (kg) / 3600).

5.6.2 Medical History

Medical history will be listed for all patients.

5.6.3 Baseline Disease Status

Baseline disease status will be summarized by the treatment groups, including disease primary diagnosis, time since initial diagnosis (months), IPFP risk factors, Ann Arbor stage, and baseline Eastern Cooperative Oncology Group (ECOG) performance status.

Patient distribution by stratification factors will also be summarized.

5.7 Treatments and Medications

5.7.1 Concomitant Medications

Concomitant medications will be coded by preferred term using the World Health Organization (WHO) Drug Dictionary. The number and percentage of patients taking concomitant medications will be tabulated.

Concomitant procedures will not be coded, but will be presented in a data listing in the clinical study report (CSR).

5.7.2   Study Treatments

The study treatments for this study are A+AVD and ABVD. ABVD consists of doxorubicin 25 mg/m^2, bleomycin 10 units/m^2, vinblastine 6 mg/m^2, and dacarbazine (DTIC) 375 mg/m^2. A+AVD consists of doxorubicin 25 mg/m^2, vinblastine 6 mg/m^2, dacarbazine (DTIC) 375 mg/m^{2,} and brentuximab vedotin (ADCETRIS) 1.2 mg/kg. A+AVD and ABVD will be administered by IV infusion on Days 1 and 15 of each 28-day cycle, for a total of 6 cycles. Details of study treatments administration is described in the protocol.

5.7.2.1   Extent of Exposure

The exposure to each component of study treatments A+AVD and ABVD will be characterized by total amount of dose received (mg or units), total number of doses received, number of treatment cycles, numbers and percentages of patients by maximum number of treatment cycles completed(1, 2, … , 6).

A treatment cycle is defined as a cycle in which the patient received any amount of study drug.

Relative dose intensity (%) is defined as: 100 x (total dose received)/(total dose intended). Total dose intended is the summation of the intended doses in all treatment cycles. The intended dose in each cycle is determined by the dose level when the patient was randomized.

Dosing data will also be presented in a by-patient listing.

5.7.2.2   Treatment Modifications

Action on each component of study treatments A+AVD and ABVD will be summarized by Cycles 1 through 6.

Patients who switch to another frontline therapy before completing A+AVD and ABVD will be summarized by time of switch and by treatment to which they switch.

5.8 Efficacy Analyses

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

5.8.1 Primary Efficacy Endpoint

The primary endpoint is mPFS (modified progression-free survival) per IRF assessment using the Revised Response Criteria for Malignant Lymphoma.

5.8.1.1 Definition

mPFS is defined as:

The time from the date of randomization to the date of the first of (1) documentation of progressive disease (PD); (2) death due to any cause; (3) for patients who are confirmed non-complete responders per IRF, receipt of anticancer chemotherapy or radiotherapy for HL after completion of frontline therapy, as defined in Table 5-1; these patients' mPFS event date will be the date of the first PET scan post completion of frontline therapy demonstrating the absence of a CR, defined as a Deauville score of ≥ 3.

Table 5-1   Completion of Frontline Therapy

Treatment History	Completion of Frontline Therapy
Did not switch therapy	Upon receipt of planned study drug regimen with no more than 2 missed doses of A+AVD or ABVD[a]
Switched therapy before completion of A+AVD or ABVD	Upon conclusion of 1 alternative anticancer regimen[b] for HL subsequent to A+AVD or ABVD discontinuation

a   A "missed dose" refers to administration of the full study regimen of either A+AVD or ABVD. Patients may miss individual agents within the A+AVD or ABVD regimen (such as patients who discontinue bleomycin for pulmonary toxicity)without this counting as a missed dose.
b   Receipt of chemotherapy OR radiation.

5.8.1.2 Handling of Missing Assessment and Censoring for mPFS Primary Analysis

The determination of the date of progression by the IRF will be used for the primary analysis. Sensitivity analyses will also be performed to evaluate the robustness of results.

For the mPFS primary 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 evidence of disease progression regardless of violations and

discontinuation of study drug. For those patients who receive new anticancer chemotherapy or radiotherapy for HL after completion of frontline therapy without confirmed noncomplete-response, mPFS will be determined by PD or death. Detailed handling rules for missing assessments and censoring for the analysis of mPFS are presented in Table 5-2.

Table 5-2   Handling of Missing Assessments and Censoring for mPFS Primary Analysis

Situation	Date of Progression or Censoring	Outcome
No baseline and/or no postbaseline assessment, no subsequent anticancer therapy after frontline therapy, no death	Date of Randomization	Censored
Disease progression documented between scheduled visits	Date of next scheduled visit	Progressed
No documented mPFS event	Date of last adequate assessment[a]	Censored
Lost to follow-up, withdraw consent before any documented mPFS event	Date of last adequate assessment[a]	Censored
Treatment discontinuation for undocumented disease progression after the last adequate assessment	Date of last adequate assessment[a]	Censored
mPFS event after more than one missed visit	Date of last adequate assessment[a]	Censored

a   Adequate assessment is defined as there is sufficient data to evaluate a patient's disease status.

5.8.2   Primary Efficacy Analysis

Final analysis of mPFS will be performed when 260 mPFS events have been observed, which is estimated to occur by 24 months after the last patient is randomized.

Stratified log-rank testing will be used to compare mPFS between the 2 treatment arms as the primary analysis. The stratification factors include region and number of IPFP risk factors at baseline (Section 5.2). The hazard ratios along with the 95% confidence interval (CI; 2-sided) will be estimated using the stratified Cox model with treatment as the explanatory variable. The Kaplan-Meier (K-M) survival curves and survival probability at 2 and 3 years along with the 2-sided 95% CIs will also be provided for each treatment group.

In addition, a stratified Cox regression model will be used to further evaluate the treatment effects on mPFS after adjusting for some prognostic factors. Besides treatment and the stratification factors (Section 5.2), the following prognostic factors will be included in the model simultaneously: age, race (white, non-white), baseline ECOG score, baseline cancer

stage, baseline B symptoms, and PET results from Cycle 2. Additional exploratory analyses may be performed if deemed necessary.

For patients with mPFS events, the reasons leading to the determination of mPFS will be tabulated. For patients without mPFS events, the main reason for censoring will also be tabulated.

The proportional hazard assumptions will be examined and sensitivity analysis may be conducted if appropriate.

5.8.2.1   mPFS Sensitivity Analyses

Sensitivity analyses will be performed for mPFS to evaluate the robustness of treatment effects. To satisfy EMA requirements, a sensitivity analysis will be performed by treating the last two categories of Table 5-2 as events, the events time will follow the description provided in Table 5-3. In the second sensitivity analysis, mPFS based on the investigators' determinations of disease progression will be analyzed in the same manner as the primary analysis. In the third sensitivity analysis, if confirmed non-complete response constitutes the mPFS event, the mPFS event date will be the date of receipt of first dose of second-line therapy. Additional sensitivity analyses for mPFS will be performed based on the alterations of the handling of missing assessment and censoring in Table 5-3, on the basis of 1 alteration at a time, not on combined alterations unless otherwise specified.

Table 5-3   Handling of Missing Assessments and Censoring for mPFS Sensitivity Analysis

Situation	Date of Progression or Censoring	Outcome
Disease progression documented between scheduled visits	Date of documented disease progression	Progressed
Treatment discontinuation for undocumented disease progression after the last adequate assessment	Date of last adequate assessment	Progressed
New anticancer chemotherapy or radiotherapy after completion of frontline therapy without confirmed non-complete-response	Date of first dose of second-line therapy	Progressed
New anticancer chemotherapy or radiotherapy after completion of frontline therapy without confirmed non-complete-response	Date of first dose of second-line therapy	Censored
mPFS event after more than one missed visit	Date of mPFS event	Progressed
Lost to follow-up before any documented mPFS event	Next scheduled assessment	Progressed
Lost to follow-up before any documented mPFS event	A+AVD: Next scheduled assessment	progressed
	ABVD: Last adequate assessment	Censored
Lost to follow-up before any documented mPFS event	A+AVD: Last adequate assessment	Censored
	ABVD: Next scheduled assessment	Progressed

5.8.2.2   Other Analysis

In addition, the primary analysis of mPFS will be performed for the PP population, as well as for the subgroups specified below. The statistical model will be adjusted accordingly to fit the subgroup analyses.

Table 5-4   Subgroups for Subgroup Analysis

Subgroup	Definition of Group
Age	< 60 years; $\geq$ 60 years
Region	North American, Europe, Asia
Number of IPFP risk factors	0-1; 2-3; 4-7
Baseline cancer stage	stage III; Stage IV
Baseline B symptoms	Present; absent
Cycle 2 PET	Positive; negative
Cycle 2 PET Deauville Score	< 5; 5
Receipt of alternative frontline therapy	Yes; No
Baseline extra nodal sites	0, 1, >1

As HL is a curable disease, a mixture cure survival model will be performed:

$$Survival_i = P_i + (1 - P_i)S_i$$

where P_i denotes the fraction of patients cured in each arm, S_i is the mPFS survival function for those who are not cured. S_i is assumed to follow Weibull distribution with the same shape parameter for both treatment arms. If Weibull distribution assumption for S_i is not appropriate, other parametric distributions will be considered. Two additional exploratory analyses will also be performed. One for mPFS with definition of frontline therapy restricted to no-switch in therapy. The other for PFS, which is defined as the earlier of (1) documentation of PD or (2) death due to any cause. The statistical methods will be similar to those used for mPFS.

5.8.3   Key Secondary Efficacy Endpoint

OS is designated as a key secondary endpoint. The key secondary endpoint will be tested at a 1-sided 0.025 level when the test of mPFS is statistically significant.

OS is defined as the time from the date of randomization to the date of death. Patients without documented death at the time of analysis will be censored at the date last known to be alive.

There will be 2 formal analyses for OS, an OS interim analysis at the time of the final mPFS analysis, and the OS final analysis when 112 deaths have occurred. Overall type I error will be controlled using the O'Brien-Fleming method with a Lan-DeMets alpha spending function.

Stratified log-rank testing will be used to compare OS between the 2 treatment arms. The stratification factors include region and number of IPFP risk factors at baseline (Section 5.2). The hazard ratios along with the 95% CIs (2-sided) will be estimated using a stratified Cox regression model. The Kaplan-Meier method will be used to estimate the distribution of the OS endpoint for each treatment. The 25th, 50th (median), 75th percentiles of survival times (if estimable), and the 2-sided 95% CIs will be presented. Analysis of OS will be performed based on the ITT population.

In addition, a stratified Cox regression model will be used to further evaluate the treatment effects on OS after adjusting for some prognostic factors. Besides treatment and the stratification factors (Section 5.2), the following prognostic factors will be included in the model simultaneously: age, race (white, non-white), baseline ECOG score, baseline cancer stage, baseline B symptoms, and PET results from Cycle 2. Additional exploratory analyses may be performed if deemed necessary.

Subgroup analyses may be performed using subgroups defined for mPFS analyses. Additional analyses of OS may be performed if deemed necessary.

5.8.4 Other Secondary Efficacy Endpoints

5.8.4.1 Overview of Statistical methods

Other secondary efficacy endpoints mainly consist of 2 types of variables: time to event variables (eg, event-free survival) and binary outcome variables (eg, CR). For time to event variables, the statistical methods will be the same as that for time to death outlined in Section 5.8.3. For binary outcome variables, the statistical methods will be the same as that for CR outlined in Section 5.8.4.2. Other endpoints that don't belong to these 2 categories will be addressed separately.

5.8.4.2 Complete Remission Rate at the End of Randomized Regimen per IRF

CR rate per IRF is defined as the proportion of patients who achieve CR at the end of treatment with randomized regimen (ABVD or A+AVD) as determined by an IRF.

The number and percentage of patients who experienced a CR at the end of treatment with ABVD/A+AVD will be summarized by treatment groups. The response rates between the 2 treatment groups will be tested using Cochran-Mantel-Haenszel (CMH) chi-square test, stratified by the stratification factors (Section 5.2). The CMH chi-square p-value and the relative risk (and odds ratio), along with its 95% 2-sided CI will also be provided. In

addition, the absolute treatment difference in CR response rates will be provided along with the 95% 2-sided CI.

In addition, a logistic regression model will be used to further evaluate the treatment effects on CR rate after adjusting for some prognostic factors. Besides treatment and the stratification factors (Section 5.2), the following prognostic factors will be included in the model simultaneously: age, race (white, non-white), baseline ECOG score, baseline cancer stage, baseline B symptoms, and PET results from Cycle 2. The odds ratio and its associated 95% CIs will be presented. Subgroup analyses may be performed using subgroups defined for mPFS analyses. Additional exploratory analyses may be performed if deemed necessary.

In the primary analysis of CR using ITT population, nonevaluable patients will be treated as nonCR. Sensitivity analyses for CR per IRF will be performed using the response-evaluable population. CR rate per investigator will also be analyzed similarly using the ITT population.

5.8.4.3   Complete Remission Rate at the End of Frontline Therapy

CR rate at the end of frontline therapy per IRF is defined as the proportion of patients who achieve CR at the end of frontline therapy as determined by an IRF. CR at the end of frontline therapy is a binary outcome variable.

5.8.4.4   Event-Free Survival

EFS is defined as the time from randomization until any cause of treatment failure: disease progression, premature discontinuation of randomized treatment for any reason, or death due to any cause, whichever occurs first. EFS is a time-to-event variable.

5.8.4.5   Disease-Free Survival

DFS is defined as the time from CR to disease progression or to death from lymphoma or acute toxicity from treatment. Analyses of DFS will be performed based on the subset of the ITT population achieving a CR. DFS is a time-to-event variable.

5.8.4.6 Overall response rate

Overall response rate (ORR) is defined as the proportion of patients who achieve CR or PR at the end of treatment with randomized regimen (ABVD or A+AVD) as determined by an IRF.

ORR is a binary outcome variable.

5.8.4.7 Duration of Response (DOR) or Complete Response (DOCR)

DOR in subjects with confirmed response is the time between first documentation of response (PR or CR) and disease progression. DOCR in subjects with confirmed CR is the time between first documentation of CR and disease progression. DOR and DOCR per IRF will be analyzed based on the subset of ITT population who had response. DOR and DOCR are time-to event-variables.

5.8.4.8 PET Negativity Rate

PET negativity rate at Cycle 2 is defined as the proportion of patients with negative Cycle 2 PET results. PET negativity is a binary outcome variable.

5.8.5 Other Efficacy Endpoints

Alive without HL rate at 3 years and 5 years is defined as the proportion of patients who are alive without classical Hodgkin lymphoma at 3 years or 5 years after the patient's randomization date.

Alive without HL rates between the 2 treatment arms will be compared using a stratified Cochran-Mantel-Haenszel (CMH) test.

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 based on serum or plasma samples collected from patients who meet 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. Exploratory safety-PK, efficacy-PK, and if possible, PK-pharmacodynamic relationships will be determined.

Descriptive statistics (eg, number of patients, mean, standard deviation, median, minimum, and maximum) will be used to summarize concentrations of analyte for brentuximab vedotin-treated patients.

The pharmacokinetics of doxorubicin, vinblastine, and dacarbazine will be compared between the treatment arms. Descriptive statistics (eg, number of patients, mean, standard deviation, median, minimum, and maximum) will be used to summarize concentrations of analyte. Geometric mean ratios of the AUC will be calculated for each AVD component (doxorubicin, vinblastine, and dacarbazine).

5.9.2 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.

Relationships between ATA development and safety and efficacy will be explored.

5.9.3 Biomarker Analysis

Absolute and change from baseline value of circulating biomarkers will be summarized by time point using descriptive statistics, as applicable. Descriptive statistics (eg, number of patients, mean, standard deviation, median, minimum, and maximum) will also be provided to summarize disease markers (CCI), tissue levels of potential resistance markers (CCI

), and qualitative and semiquantitative measures of markers (CCI

), and change from baseline values of these markers, as applicable. The association between these and clinical response or safety endpoints might be explored.

5.10 Resource Utilization and Patient-Reported Outcome Analysis

Analyses of patient-reported outcomes (PROs) and health economics will be performed using the ITT population.

5.10.1   Patient-Reported Outcomes Analysis

Patient-reported outcome (PRO) assessments based on EORTC QLQ-C30, FACIT Dyspnea 10, and FACT/GOG-NTX Additional scale will be analyzed to determine if response to therapy is accompanied by changes in the quality of life.

The analysis will be based on scores from the global health status/QoL scale of QLQ-C30, shortness of breath scale of FACIT Dyspnea 10, and the sensory scale of FACT/GOG-NTX. The analysis of change of scores from baseline will use mixed-effects models with repeated measures at each time point specified in the Schedule of Events of the protocol and the 95% CIs of the estimates will be provided.

Descriptive summaries of all the scales and individual item scores observed will be generated at each scheduled assessment time point by treatment.

Missing data will be summarized with the proportion of missing responses for each item over time. Manuals for scoring and handling missing data published for QLQ-C30, FACIT Dyspnea 10, and FACT/GOG-NTX will be used to impute missing data initially.

A sensitivity analysis will be performed to evaluate the impact on the analysis results from missing data imputation. Conditional on the patterns of missing data, multiple imputation methods, including a pattern-mixture model, will be considered. Any deaths that occur before the end of treatment (EOT) are to be imputed by a value zero and will be considered missing otherwise.

5.10.2   Health Economics Analysis Using Medical Resource Utilization and Utility

EQ-5D scores will be summarized in descriptive statistics for treatment arms.

MRU data will be summarized in descriptive statistics of medical encounters (length of stay, inpatient, outpatient, and reason), number of missing days from work or other activities by patient and care-giver for treatment arms.

Further modeling will be performed separately at post hoc analyses.

5.11   Safety Analyses

Safety will be evaluated by the incidence of TEAEs, severity and type of AEs, and by changes from baseline in the patient's vital signs, ECOG performance status, ECG, and

clinical laboratory results. Exposure to study drug and reasons for discontinuation will be tabulated.

These analyses will be performed using the safety population.

5.11.1 Adverse Events

Treatment-emergent adverse events are defined as any AE that occurs after administration of the first dose of study drug and through 30 days after the last dose of frontline therapy.

The MedDRA dictionary will be used to code the investigator's adverse event terms. Adverse event tables will summarize subject incidence by treatment group actually received. The severity of adverse events will be assessed using NCI CTCAE version 4.03.

5.11.1.1 Adverse Events

The number and percentage of patients experiencing at least 1 treatment emergent adverse event will be tabulated by MedDRA primary system organ class, high level term, preferred term, and treatment group. For the number of patients with AEs, patients reporting the same event more than once will have that event counted only once within each body system, once within each high level term, and once within each preferred term. AEs will be coded using the MedDRA dictionary.

The number and percentage of patients experiencing at least 1 treatment emergent AE considered related to treatment will also be summarized. If the event is missing relationship data, the event will be included in the total column for the corresponding treatment group. Multiple occurrences of the same event are counted once per subject using the most related event.

AEs will also be summarized by intensity grade. Again, multiple occurrences of the same event are counted once per subject using the maximum intensity. If the intensity of an AE is missing, the event will be included in the total column for the corresponding treatment group.

The following tabulations will also be presented:

- Incidence of treatment-emergent adverse events by System Organ Class (SOC), high-level term, and preferred term

- Incidence of treatment-emergent adverse events reported by at least 10% of patients in either treatment group by MedDRA preferred term

- Incidence of treatment-emergent adverse events considered by the investigator to be related to study drug by MedDRA SOC, high-level term and preferred term

- Incidence of treatment-emergent adverse events that caused study drug dose modification by MedDRA SOC, high-level term and preferred term

For treatment-emergent adverse events with NCI CTCAE toxicity Grade 3 or higher, the following summaries are to be produced:

- Incidence of Grade ≥ 3 treatment-emergent adverse events, by SOC, high-level term and preferred term

- Incidence of Grade ≥ 3 treatment-emergent adverse events considered by the investigator to be related to study drug, by SOC, high-level term and preferred term

5.11.1.2 Serious Adverse Events

The number and percentage of patients experiencing at least 1 treatment-emergent serious AE (SAE) will be summarized by MedDRA primary system organ class, high-level term, and preferred term. The following summaries are to be produced:

- Incidence of treatment-emergent serious adverse events, by SOC, high-level term and preferred term

- Incidence of treatment-emergent serious adverse events considered by the investigator to be related to study drug, by SOC, high-level term and preferred term

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

5.11.1.3 Deaths

All deaths occurring on-study and during follow-up will be summarized. Cause of death will be summarized as well in this table. Frequencies of deaths due to study treatment-related adverse events will also be reported.

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 AE status).

5.11.1.4   Adverse Events Resulting in Discontinuation of Study Drug

Adverse events results in discontinuation of study drug will be tabulated by treatment. 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 non-numeric qualifier (eg, less than (<) a certain value, or greater than (>) a certain value), the given numeric value will be used in the summary statistics, ignoring the non-numeric qualifier. In case a central lab value is missing, the local lab, if available, will be used.

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, leukocytes with differential, neutrophils (absolute neutrophil count [ANC])

- Serum chemistry: Blood urea nitrogen, Creatinine, Total bilirubin, Urate, Lactate dehydrogenase, Gamma-glutamyl transpeptidase (GGT), Phosphate, Albumin, Alkaline phosphatase (ALP), AST, ALT, Glucose, Sodium, Potassium, Calcium, Chloride, Carbon dioxide, Magnesium.

- Other: Hemoglobin A1C

Laboratory data of hematology and clinical chemistry up to 30 days after last dose or the End of Treatment visit date, whichever is later, will be reported in SI units. Normal ranges from the central and local laboratories used in this study will be listed.

Summary statistics (mean, standard deviation, median, and range) will be calculated for the raw data and for their percentage changes from baseline at each time point of assessment and for the percentage changes from baseline to the last value. Individual values outside the

normal ranges will be identified (by "H" for high and "L" for low) in the data listings displaying the absolute values for each subject.

Shift tables from baseline to worst value (ie, worst grade) on study (from treatment start to 30 days after last dose or the End of Treatment visit date, whichever is later) will be provided for the following laboratory parameters: hemoglobin, ANC, WBC, platelets, AST, ALT, bilirubin and HbA1C.

5.11.3 Electrocardiograms

Investigators' assessments of ECG monitoring (normal, abnormal and clinically significant, or abnormal and not clinically significant), including unscheduled or retested measurements, will be presented in a listing.

5.11.4 Vital Signs

The actual values of vital sign parameters including temperature, pulse rate, and systolic and diastolic blood pressure when available, will be summarized over time.

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

Not applicable.

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm)
PPD	Biostatistics Approval	10-Mar-2015 18:13
PPD	Biostatistics Approval	10-Mar-2015 19:17
PPD	Clinical Approval	10-Mar-2015 20:30