1.0 Title Page

Statistical Analysis Plan

Study M16-085

A Phase 2, Open-Label, Multicenter, Dose-Escalation and Expansion Study of Venetoclax and Pomalidomide Combination Therapy with Dexamethasone in Subjects with Relapsed or Refractory Multiple Myeloma

Date: 06 Feb 2018

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

List of Appendices

3.0 Introduction

This statistical analysis plan (SAP) serves as a guide for the analysis of the efficacy and safety endpoints pertaining to ABT-199 Study M16-085.

Unless noted otherwise, all analyses will be performed using SAS version 9.4 or higher (SAS Institute Inc., Cary, NC 27513) in the UNIX operating system.

This SAP will not be updated unless a proposed protocol amendment is expected to impact the data analysis.

4.0 Study Background

4.1 Objective

The objective of this study is to determine the safety profile, tolerability, preliminary efficacy, recommended Phase 2 dose, and pharmacokinetics (PK) of venetoclax and pomalidomide combination therapy with dexamethasone (ven/pom/dex) in subjects with relapsed or refractory (R/R) multiple myeloma (MM) who have received ≥ 3 prior lines of therapy, including a proteasome inhibitor (PI) and an immunomodulatory drug (IMiD), or who are double-refractory to a PI and an IMiD.

4.2 Study Design and Design Diagram

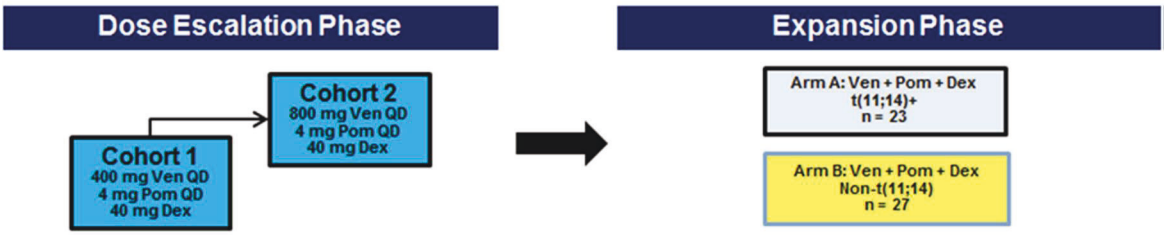
This is an open-label, multicenter study with two phases, diagrammed in Figure 1.

1. **Dose escalation** will consider 6 – 12 subjects confirmed to be t(11;14)+ by an analytically-validated FISH assay per central laboratory testing before enrollment. Dose escalation will be guided by a Bayesian optimal interval (BOIN) design, based on the cumulative number of subjects who experience a DLT at the current combination dose. The subjects will be enrolled in cohorts of at least 3 subjects. The first cohort will be administered a 400 mg dose of venetoclax orally, once daily (QD). If acceptable safety and tolerability are observed at the completion of Cycle 1 in at least 3 DLT-evaluable subjects, the dose of venetoclax will be

escalated to 800 mg QD (the target dose). All subjects will be administered the approved dose of pomalidomide (oral 4 mg QD) and dexamethasone (40 mg once weekly (qw)). To be considered DLT-evaluable, subjects must complete at least 80% of pomalidomide and venetoclax dosing (or experience a DLT) during Cycle 1. DLTs are defined in the operations manual. Additional information regarding the BOIN design is provided in Appendix A.

2. **Dose expansion** will consist of 2 independent arms. Arm A will consider t(11;14)+ subjects, and Arm B will consider non-t(11;14) subjects. All subjects who participate in this phase must be screened for t(11;14) translocation status before enrollment. Arm A will feature around 23 subjects and Arm B will feature around 27 subjects. All subjects will receive the ven/pom/dex combination until documented PD, unacceptable toxicity, withdrawal of consent, or they meet another criteria for discontinuation (see study protocol). Subjects may discontinue pomalidomide but continue receiving venetoclax QD for up to 2 years following the date of the last subject enrolled, provided they completed Cycle 1 of ven/pom/dex, continue to tolerate venetoclax, do not exhibit PD, and do not meet any criteria for discontinuation. AbbVie will work with the investigator to evaluate options for continued venetoclax therapy for subjects that continue to derive benefit after 2 years of treatment.

Figure 1. Study Design Diagram



4.3 Endpoints

4.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the overall response rate (ORR), which is defined as the proportion of subjects experiencing a stringent complete response (sCR), complete response (CR), very good partial response (VGPR), or partial response (PR).

4.3.2 Secondary Efficacy Endpoint

Key secondary efficacy endpoints include progression-free survival (PFS), duration of response (DOR), and time to progression (TTP). PFS for a given subject is defined as the number of days from the date of first dose until the date of the first documented PD or death due to any cause, whichever occurs first. All events of PD will be included, regardless of whether the event occurred while the subject was still taking study drug or had previously discontinued study drug. If the subject does not experience PD or death, the subject's data will be censored. See Table 1 for details.

Table 1. Event and Censoring Dates Used in PFS

Situation	Option for End-Date	Outcome
No baseline assessment	Date of first dose	Censor
PD or death at scheduled assessment date or before the next scheduled assessment	Date of PD or death (whichever occurs first)	Event
PD or death after exactly one missing assessment	Date of PD or death (whichever occurs first)	Event
PD or death after two or more missing assessments	Date of first dose (if no adequate assessment is available prior to PD/death) or date of the last adequate assessment prior to PD/death	Censor
No PD and no death	Date of first dose (if no adequate assessment is available) or date of last adequate assessment	Censor

Time to Disease Progression (TTP) for a given subject is defined as the number of days from the date of first dose to the date of first documented PD or death due to MM,

whichever occurs first. If the subject does not have an event of PD and the subject has not died due to MM, the subject's data will be censored. See Table 2 for details.

Table 2. Event and Censoring Date Used in TTP

Situation	Option for End-Date	Outcome
No baseline assessment	Date of first dose	Censor
PD or death due to multiple myeloma at scheduled assessment date or before the next scheduled assessment	Date of PD or death (whichever occurs first)	Event
PD or death due to multiple myeloma after exactly one missing assessment	Date of PD or death (whichever occurs first)	Event
PD or death due to multiple myeloma after two or more missing assessments	Date of first dose (if no adequate assessment is available prior to PD/death) or date of the last adequate assessment prior to PD/death	Censor
No PD and no death due to multiple myeloma	Date of first dose (if no adequate assessment is available) or date of last adequate assessment	Censor

Duration of Response (DOR) for a given subject is defined as the number of days from the date of that subject's first documented response (PR or better) to the date of first documented PD or death due to multiple myeloma, whichever occurs first. If the subject with a documented response does not have an event of PD and the subject has not died due to MM, the subject's data will be censored. See Table 3 for details. Subjects who never achieve a documented response will not be included in the analysis of DOR.

Table 3. Event and Censoring Date Used in DOR

Situation	Option for End-Date	Outcome
No baseline assessment	Date of first dose	Censor
PD or death due to multiple myeloma at scheduled assessment date or before the next scheduled assessment	Date of PD or death (whichever occurs first)	Event
PD or death due to multiple myeloma after exactly one missing assessment	Date of PD or death (whichever occurs first)	Event
PD or death due to multiple myeloma after two or more missing assessments	Date of the last adequate assessment prior to PD/death	Censor
No PD and no death due to multiple myeloma	Date of the last adequate assessment	Censor

4.3.3 Safety Endpoint

Safety evaluations include adverse event (AE) monitoring, physical examinations, vital sign measurements, electrocardiogram (ECG) variables, and clinical laboratory testing (hematology, chemistry, and urinalysis) as a measure of safety and tolerability for the entire study duration.

4.3.4 Pharmacokinetic Endpoints

PK samples will be collected and analyzed for venetoclax and pomalidomide concentrations. The maximum observed plasma concentration (C_{max}), time to C_{max} (T_{max}), and the area under the plasma concentration vs time curve (AUC) will be determined using noncompartmental methods.

4.3.5 Biomarker Research Endpoints

Biospecimens (blood, serum, plasma, bone marrow aspirate, and bone marrow core biopsy tissue) will be collected at specified time points throughout the study to evaluate known or novel disease-related or drug-related biomarkers. Types of biomarkers analyzed may include nucleic acids, proteins, lipids, or metabolites. Some biomarkers of

potential interest can be found in Section 3.2 of this SAP, and any noteworthy results may be included in the clinical study report. Details pertaining to biomarker research endpoints can be found in the operations manual.

4.4 Sample Size Justification

Approximately 60 subjects will be enrolled into this two-part study. Dose escalation will enroll around 10 subjects, and dose expansion will enroll around 50 subjects.

1. **Dose Escalation** will enroll at least 3 subjects per cohort according to a BOIN design, for a total of 6 – 12 subjects.
2. **Dose Expansion** will enroll around 23 subjects in Arm A, and 27 subjects in Arm B.
    a. Arm A: t(11;14)+: The historical ORR for pom/dex is 35%.[1,2] By adding venetoclax to this combination, an ORR of 70% is clinically meaningful for t(11;14)+ subjects. Allowing for a maximum one-sided type 1 error rate of 0.025, and assuming an ORR of 35% from the historical data, 23 subjects will provide 90% power if the true ORR for t(11;14)+ subjects treated with ven/pom/dex is 70%.
    b. Arm B: Non-t(11;14): The historical ORR for pom/dex is 35%. By adding venetoclax to this combination, an ORR of 60% is clinically meaningful for non-t(11;14) subjects. The sample size for Arm B will be approximately 27 subjects. Allowing for a maximum one-sided type 1 error rate of 0.1, and assuming an ORR of 35% from the historical data, 27 subjects will provide 90% power if the true ORR for non-t(11;14) subjects treated with ven/pom/dex is 60%.

4.5 Interim Analysis

During the dose escalation phase, review meetings will be conducted at the completion of Cycle 1 in at least 3 evaluable subjects per cohort.

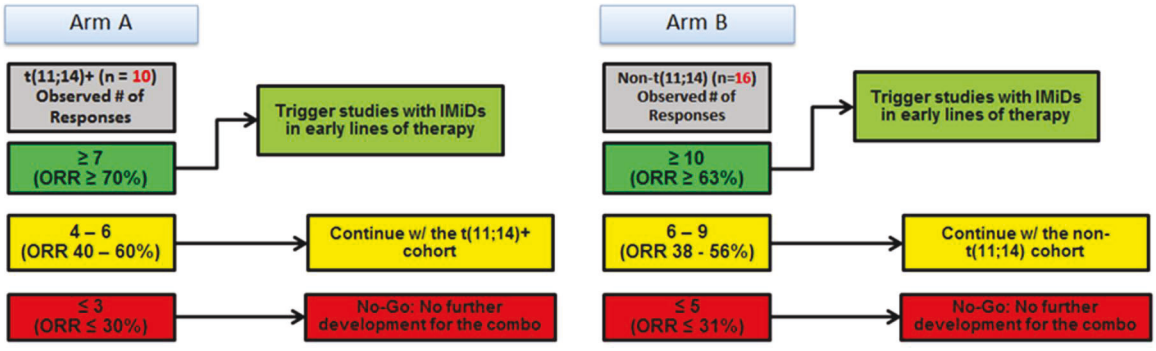
Since this is an open-label study, both arms of the dose-expansion phase will be continuously monitored. Interim analysis (IA) may be conducted when adequate data from approximately 10 and 16 subjects are available from Arms A and B, respectively. Figure 2 displays some guidelines for go/no-go decisions for the IA, broken down by arm. Observed ORRs in the green region may be used as evidence to accelerate studies of venetoclax in combination with IMiDs to earlier lines of therapy. Results in the yellow region suggest evaluating the depth and duration of response for more guidance.

Arm A: the probability of making the wrong "go" decision is 2.6%, given a true ORR of 35%, and the probability of making the wrong "no-go" decision is 1.1%, given a true ORR of 70%. The final analysis for Arm A will include the subjects from the escalation portion of the study, resulting in a grand total of around 29 t(11;14)+ subjects.

Arm B: the probability of making the wrong "go" decision is 2.3%, given a true ORR of 35%, and the probability of making the wrong "no-go" decision is 1.9%, given a true ORR of 60%.

The final go/no-go decision will be based on totality of efficacy and safety data.

Figure 2. Interim Analysis Guidelines



5.0 Analysis Populations and Baseline Characteristics

5.1 Analysis Populations

The full analysis set (FAS) includes all subjects who received at least 1 dose of study drug. The FAS will be used for all safety, efficacy, and baseline analyses.

Table 4. Baseline Characteristics

Variable	Levels
Age	< 65, ≥ 65
Sex	male, female
Race	white, black, Asian, other
Ethnicity	Hispanic, non-Hispanic
Type of myeloma	IgG, IgA, IgD, light chain
Prior lines of therapy	1 – 3, 3 – 6, > 6
Prior therapy with:	• PI + IMiD • PI (bortezomib, carfilzomib, ixazomib) • IMiD (thalidomide, lenalidomide, pomalidomide), • monoclonal antibody (daratumumab, elotuzumab) • stem cell transplant
BCL-2	low, high
ISS stage	I, II, III
ECOG	0, 1, 2, > 2
t(11;14)	positive, negative
t(4;14)	positive, negative
17p deletion	yes, no
13q deletion	yes, no
Hyperdiploid	yes, no

6.0 Efficacy Analyses

6.1 Primary Efficacy Analysis

For ORR, exact (binomial) one-sided tests of H0: ORR ≤ 35% vs H1: ORR > 35% at significance levels of 0.025 and 0.1 will be performed for Arms A and B, respectively.

Point estimates and exact (Clopper-Pearson) 95% confidence intervals of ORR for each of the two arms will be calculated.

6.2 Efficacy Analysis

PFS, DOR, and TTP will be analyzed with Kaplan-Meier methodology. Median PFS, DOR, and TTP will be estimated, along with corresponding 95% CIs.

7.0 Safety Analyses

7.1 Analysis of Adverse Events

AE analysis will only include treatment-emergent events (those that have an onset on or after the day of the first dose of any study drug), and will not include AEs that have an onset of more than 30 days after the last dose of any study drug.

Treatment-emergent AEs will be coded and summarized by system organ class and preferred term according to the Medical Dictionary for Regulatory Activities (MedDRA) AE coding dictionary. The percentage of subjects experiencing an AE at a given severity will be provided along with the NCI CTCAE version 4.03 toxicity grade and relationship to the study drugs. Serious AEs, AEs leading to discontinuation of treatment, and AEs leading to death will be summarized. All summaries will be done by dose. For the study as a whole, AEs will be evaluated and summarized.

7.2 Analysis of Laboratory Data

Where applicable, laboratory values will be categorized according to the NCI CTCAE version 4.03 grades. Changes from baseline to maximum grades, as well as final post-baseline grades will be assessed. The baseline and final grades will be defined as: the grade of the last measurement collected prior to the first dose of any study drug, and the grade of the last post-baseline measurement collected no more than 30 days after the last dose of any study drug, respectively. If multiple values are available for a post-baseline measurement, then the value with the highest NCI CTCAE grade will be used.

Detailed listings of data for subjects experiencing NCI CTCAE Grade 3 – 4 blood chemistry and hematology values will be provided. All collected measurements may be included in these listings, regardless of the number of days after the last dose of any study drug.

7.3 Analysis of Vital Signs and Weight

Detailed listings of data for subjects experiencing potentially clinically significant vital sign values (according to the AbbVie-defined criteria) will be provided. All collected measurements may be included in these listings, regardless of the number of days after the last dose of any study drug.

8.0 References

1. Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. Journal of the Royal Statistical Society: Series C (Applied Statistics). (2015);64(3):507-23.
2. Richardson PG, Siegel DS, Vij R, et al. (2014). Pomalidomide alone or in combination with low-dose dexamethasone in relapsed and refractory multiple myeloma: a randomized phase 2 study. Blood. 123(12)1826-32.
3. San Miguel J, Weisel K, Moreau P, et al. (2013). Pomalidomide plus low-dose dexamethasone versus high-dose dexamethasone alone for patients with relapsed and refractory multiple myeloma (MM-003): a randomised, open-label, phase 3 trial. The Lancet Oncology. 14(11)1055-66.

Appendix A. Bayesian Optimal Interval Design

A.1 Methodology

Let ϕ denote the pre-specified target toxicity rate. Let $\Delta_L > 0$ and $\Delta_U > 0$ be pre-specified lower and upper cutoffs, respectively, which satisfy $0 < \phi - \Delta_L < \phi + \Delta_U < 1$. The BOIN design consists of determining Δ_L and Δ_U such that the interval $(\phi - \Delta_L, \phi + \Delta_U)$ minimizes the probability of incorrect dose assignment decisions. To determine Δ_L and Δ_U, consider three hypotheses at dose level j:

$$H_{0j}: p_j = \phi, \; H_{1j}: p_j = \phi_1, \; H_{2j}: p_j = \phi_2$$

where p_j denotes the true toxicity rate at dose level j, ϕ_1 denotes the highest toxicity rate that is deemed subtherapeutic such that dose escalation should be made and ϕ_2 denotes the lowest toxicity rate that is deemed overly toxic such that dose de-escalation should be made. Thus, the correct dose assignment decisions under H_{0j}, H_{1j} and H_{2j} are retainment (i.e., stay at current dose), escalation and de-escalation, respectively.

Assuming an equal prior probability of the three hypotheses being true (i.e., = $\Pr(H_{0j}) = \Pr(H_{1j}) = \Pr(H_{2j}) = 1/3$), the optimal interval $(\phi - \Delta_L, \phi + \Delta_U)$ which minimizes the probability of incorrect dose assignment decisions is given by:

$$\phi - \Delta_L = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)} \text{ and } \phi + \Delta_U = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)}$$

Suppose the current dose is j where $j = 1, \ldots, J$. Let $\hat{p}_j = \left(y_j / n_j\right)$ denote the estimated toxicity rate based on the accumulated information on dose combination j with y_j toxicities and n_j subjects. Define an admissible dose escalation set as $A_E = \{(j+1))\}$ and an admissible dose de-escalation set as $A_D = \{(j-1)\}$. Dose assignment decisions under the BOIN design proceeds as follows:

1. Treat the first cohort at the starting dose combination.

2. Suppose the current cohort is treated at dose combination *j*; then for the next cohort of patients:
   a. If $\hat{p}_j \leq \phi - \Delta_L$, escalate to the next higher dose combination that belongs to A_E. If $j = J$, the current dose combination is retained.
   b. If $\hat{p}_j \geq \phi + \Delta_U$, de-escalate to the next lower dose combination that belongs to A_D. If $j = 1$, the current dose combination is retained.
   c. Otherwise, if $\phi - \Delta_L < \hat{p}_j < \phi + \Delta_U$, the current dose combination is retained.
3. This process continues until the total sample size is exhausted or the study is terminated early due to excess toxicity (as described below).

The BOIN design also imposes the following safety rule: dose combinations which satisfy $Pr\{p_j > \phi | y_j\} \geq \lambda$ are eliminated, where λ is a pre-specified threshold probability.

A.2 Application to Study M16-085

If possible, only the venetoclax dose will be increased. The pomalidomide and dexamethasone doses will be fixed in both available cohorts during the dose escalation part of the study.

For Study M16-085, the toxicity rates and safety threshold are pre-specified as follows:

$\phi = 0.30$ (target toxicity rate)

$\phi_1 = 0.6\phi$ (highest toxicity rate which is subtherapeutic)

$\phi_2 = 1.4\phi$ (lowest toxicity rate which is overly toxic)

$\lambda = 0.95$ (safety threshold for dose combination elimination)

The above pre-specified values yield an optimal interval of (0.236, 0.359). Approximately 6 – 12 subjects may be enrolled in cohorts of at least 3. The

corresponding BOIN design decision rules (based on the cumulative number of patients who experience a DLT at the current dose combination) are presented in Table 5.

Table 5. Dose-Escalation Decision Rules

Action	Number of DLT-Evaluable Subjects Treated at Current Combination							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Escalate if number of subjects with DLT ≤	0	0	1	1	1	1	2	2
Stay at current combination if number of subjects with DLT =	1	1	-	2	2	2	3	3
De-escalate if number of subjects with DLT ≥	2	2	2	3	3	3	4	4
Eliminate[a] if number of subjects with DLT ≥	3	3	4	4	5	5	5	6

DLT = dose-limiting toxicity

a. Eliminate current and higher combinations (i.e., venetoclax dose and pomalidomide dose ≥ current dose level).