1.0 Title Page

Statistical Analysis Plan

Study M14-031

A Phase 3, Multicenter, Randomized, Double Blind Study of Bortezomib and Dexamethasone in Combination with Either Venetoclax or Placebo in Subjects with Relapsed or Refractory Multiple Myeloma Who are Sensitive or Naïve to Proteasome Inhibitors

Date: 16 Jan 2019

Version 3.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analysis for Venetoclax (ABT-199/GDC-0199) study Protocol M14-031. It provides details on planned statistical analyses for efficacy and safety endpoints, interim analysis, and multiplicity control strategies.

Unless noted otherwise, all analyses will be performed using SAS version 9.3 or later (SAS Institute Inc., Cary, NC 27513) under the Unix operating system. EAST version 6.4 or later (Cytel Inc.®) will be used to determine stopping boundaries in the group sequential plan.

4.0 Study Background

4.1 Objective

Primary Objective:

The primary objective of the study is to compare progression free survival (PFS) based on the International Myeloma Working Group (IMWG) Criteria,[1,2] as determined by an Independent Review Committee (IRC), between treatment arms in subjects with relapsed or refractory multiple myeloma who are considered sensitive or naïve to proteasome inhibitors and received 1 to 3 prior lines of therapy for multiple myeloma.

Secondary Objectives:

The secondary objectives are to compare, between treatment arms, the following:

1. Overall Response Rate (ORR)
2. Very Good Partial Response (VGPR) or better response rate
3. Overall Survival (OS)
4. Physical Functioning
5. Worst Pain

6. PFS in subjects with high BCL-2 expression
7. Duration of response (DOR)
8. Time to progression (TTP)
9. Minimal residual disease (MRD) negativity rate
10. Global Health Status (GHS)/Quality of Life (QoL) (European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core [EORTC QLQ-C30])
11. Fatigue (Patient Reported Outcomes Measurement Information System [PROMIS] Cancer Fatigue Short Form [SF])
12. Safety

Tertiary Objectives:

The tertiary objectives of the study are to assess

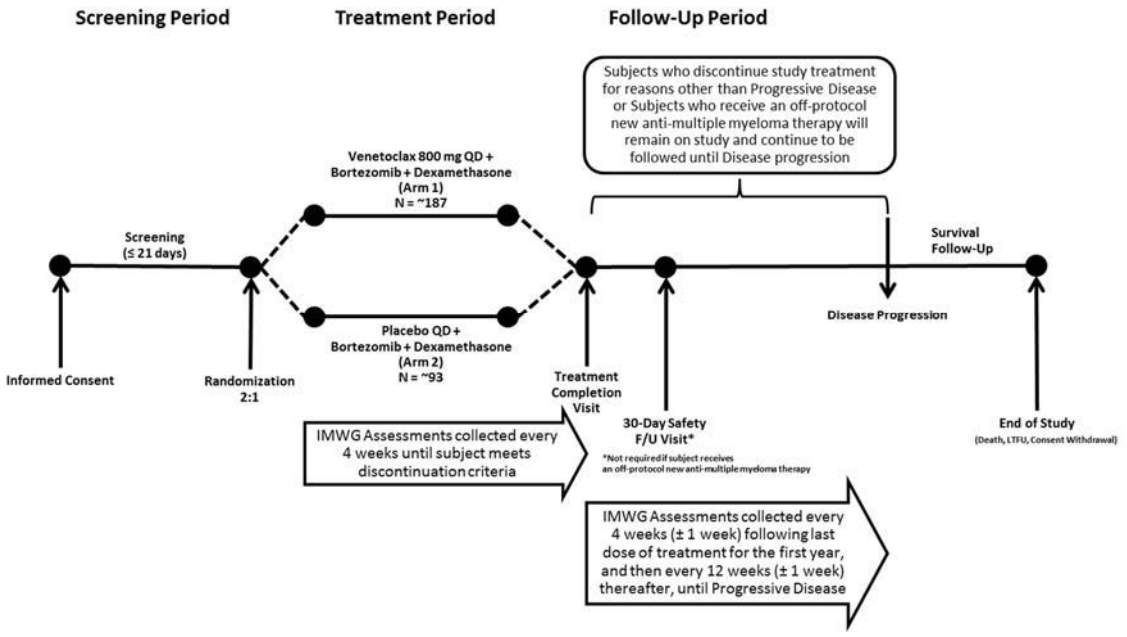
1. Clinical benefit rate (CBR)
2. Disease control rate (DCR)
3. Time to response (TTR)
4. Time to next treatment (TNT)
5. Other PRO endpoints (remaining subscales/items from BPI-SF, EORTC QLQ-C30, EORTC Quality of Life Questionnaire Multiple Myeloma Module [EORTC QLQ-MY20], and EuroQol EQ-5D-5L), and a preliminary assessment of potential biomarkers for association with PK, safety and efficacy.

4.2 Study Design

Study M14-031 is a Phase 3, multicenter, randomized, double blind study of bortezomib and dexamethasone in combination with either venetoclax or placebo in subjects with

relapsed or refractory multiple myeloma who are sensitive or naïve to proteasome inhibitors (PI).

Figure 1. Study Schema



There are two treatment arms:

- Arm 1: Venetoclax plus Bortezomib (B) and Dexamethasone (d)
- Arm 2: Placebo plus Bd

Once screening procedures are completed and eligibility is confirmed, subjects are randomized in 2:1 ratio to Arm 1 and Arm 2 using the following stratification factors:

1. Prior exposure to proteasome inhibitors (naïve versus sensitive), and
2. Number of prior lines of therapy (1 versus 2 or 3)

Separate randomization lists are used for the following groups of subjects:

1. Subjects from countries other than Japan
2. Subjects from Japan
    a. Run-in phase
    b. Rest of the subjects

In total 12 subjects are randomized in run-in phase in Japan to assess safety of the experimental treatment. The rest of the subjects in Japan are randomized once the experimental treatment is observed to be safe. Details on this are in the protocol.

Disease assessment will be based on IMWG response criteria for multiple myeloma. Each post-baseline IMWG assessment will be performed by the investigator and IRC (Independent Review Committee). Details regarding the IRC review are provided in the IRC Charter for Study M14-031. An independent data monitoring committee (IDMC) will periodically review safety data. Details regarding the IDMC's review are provided in the IDMC Charter for Study M14-031.

4.3 Endpoint

Unless noted otherwise, endpoints related to IMWG are per IRC assessments. Also, unless noted otherwise, for efficacy endpoints related to response, responses occurred on or before start date of new cancer therapy will be considered.

4.3.1 Primary Efficacy Endpoint

The primary endpoint is progression-free survival (PFS) based on IMWG criteria for multiple myeloma. The PFS is defined as the number of days from the date of randomization to the date of the first documented PD or death due to any cause, whichever occurs first.

PFS (days) = Date of first documented PD or death due to any cause –

Date of Randomization + 1

Details on event and censoring rules for PFS are provided in Table 1.

Table 1. Event and Censoring Date Used in PFS

Situation	Date of Censor or Event	Outcome
No/inadequate baseline assessment	Date of randomization	Censor
PD or death at scheduled assessment date or before the next scheduled assessment	If there is documented PD then date of PD. Otherwise date of death	Event
PD or death after exactly one missing assessment	If there is documented PD then date of PD. Otherwise date of death	Event
PD or death after two or more consecutive missing assessments	If no adequate assessment is available prior to PD/death then date of randomization. Otherwise, date of the last adequate assessment prior to PD/death	Censor
No PD and no death	If no adequate assessment is available then date of randomization. Otherwise, date of the last adequate assessment	Censor
Start of new anti-cancer therapy prior to PD or death	Date of last adequate assessment prior to start of new anti-cancer therapy. If no adequate assessment is available prior to start of new anti-cancer therapy then date of randomization.	Censor

All PFS events will be included, regardless of whether the event occurred while the subject was still taking study drug or had previously discontinued study drug. The primary efficacy analysis will be performed when approximately 136 PFS events per IRC assessments are observed.

4.3.2 Secondary Efficacy Endpoint

The following are secondary efficacy endpoints.

Overall Response Rate (ORR):

ORR is defined as the proportion of subjects with documented sCR, CR, VGPR or PR.

VGPR or better response rate:

VGPR or better response rate is defined as the proportion of subjects with documented sCR, CR, or VGPR.

Overall Survival (OS):

OS is defined as the number of days from the date of randomization to the date of death due to any cause. All events of death will be included, regardless of whether the event occurred while the subject was still taking study drug or after the subject discontinued study drug.

OS (days) = Date of death due to any cause – Date of Randomization + 1

If a subject is not known to have died, OS will be censored at the date of last contact, which is defined in Section 6.0.

Physical Functioning (EORTC QLQ-C30):

For Physical Functioning scale, scores will be computed according to procedures outlined in the EORTC QLQ-C30 scoring manual, available at http://groups.eortc.be/qol/manuals. The patient-reported outcome instruments are described in detail in Section 13.0. The endpoint is change from baseline.

Worst Pain (Brief Pain Inventory – Short Form [BPI-SF])):

For Worst Pain, scores will be computed according to procedures outlined in BPI-SF scoring manual, available at https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf. The patient-reported outcome instruments are described in detail in Section 13.0. The endpoint is change from baseline.

PFS in subjects with high BCL-2 expression:

PFS is defined in Section 4.3.1. Subjects with high or low BCL-2 expression are defined through central laboratory testing by immunohistochemistry (IHC) and based on the pre-

specified scoring algorithm in Table 2. Subjects who are randomized but have missing or indeterminate central laboratory IHC results for BCL-2 protein expression will be considered as missing BLC-2 expression status.

Table 2. Scoring Algorithm

Clinical Status	Clinical Score	Description
Low	0	No staining in tumor cells or < 50% tumor cells with cytoplasmic staining of any intensity
	1 +	≥ 50% of tumor cells with weak or higher cytoplasmic staining but < 50% of tumor cells with moderate or strong staining intensity
High	2 +	≥ 50% of tumor cells with moderate or higher cytoplasmic staining but < 50% of tumor cells with strong staining intensity
	3 +	≥ 50% of tumor cells with strong cytoplasmic staining

Duration of Response (DOR):

DOR is defined as the number of days from the date of first documented response (PR or better) to the date of first documented PD or death due to multiple myeloma, whichever occurs first.

DOR (days) = Date of first documented PD or death due to multiple myeloma – Date of First Documented Response (PR or better) + 1

Subjects who never achieve a documented response (PR or better) will be excluded from the analysis of DOR. Details on event and censoring rules for DOR are provided in Table 3.

Table 3. Event and Censoring Date Used in DOR

Situation	Date of Censor or Event	Outcome
PD or death due to multiple myeloma at scheduled assessment date or before the next scheduled assessment	If there is documented PD then date of PD. Otherwise date of death	Event
PD or death due to multiple myeloma after exactly one missing assessment	If there is documented PD then date of PD. Otherwise date of death	Event
PD or death due to multiple myeloma after two or more consecutive 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
Start of new anti-cancer therapy prior to PD/Death	Date of the last adequate assessment prior to start of new anti-cancer therapy	Censor

Time to Progression (TTP):

TTP is defined as the number of days from the date of randomization to the date of first documented PD or death due to multiple myeloma, whichever occurs first.

TTP (days) = Date of first documented PD or death due to multiple myeloma – Date of Randomization + 1

Details on event and censoring rules for TTP are provided in Table 4.

Table 4. Event and Censoring Date Used in TTP

Situation	Date of Censor or Event	Outcome
No/inadequate baseline assessment	Date of randomization	Censor
PD or death due to multiple myeloma at scheduled assessment date or before the next scheduled assessment	If there is documented PD then date of PD. Otherwise date of death	Event
PD or death due to multiple myeloma after exactly one missing assessment	If there is documented PD then date of PD. Otherwise date of death	Event
PD or death due to multiple myeloma after two or more consecutive missing assessments	If no adequate assessment is available prior to PD/death then date of randomization. Otherwise, date of the last adequate assessment prior to PD/death	Censor
No PD and no death due to multiple myeloma	If no adequate assessment is available then date of randomization. Otherwise, date of the last adequate assessment	Censor
Start of new anti-cancer therapy prior to PD/death	If no adequate assessment is available prior to start of new anti-cancer therapy then date of randomization. Otherwise, date of the last adequate assessment prior to start of new anti-cancer therapy	Censor

Minimal Residual Disease (MRD) negativity rate:

MRD negativity rate is defined as the proportion of subjects who have negative MRD by bone marrow aspirate at any time point after randomization and before progression or starting subsequent therapy. For the purpose of secondary efficacy endpoint, MRD negativity will be defined at 10^{-5} threshold (less than one residual myeloma cell per 10^5 total nucleated cells) as measured by centralized testing of bone marrow aspirate by NGS. MRD positive subjects include subjects of which all tested samples were found to be MRD positive or indeterminate. Subjects with missing or unevaluable MRD status will be considered as MRD positive. The MRD negativity rate will be calculated for each treatment group based on the ITT population (Check Section 5.0 for the definition of ITT population).

GHS/QoL (EORTC QLQ-C30):

The GHS/QoL scale specific scores will be calculated from the 22 items in EORTC QLQ-C30 for each subject based on the QLQ-C30 instruction. Higher scores are indicative of better HRQOL. The QLQ-C30 is described in detail in Section 13.0. The endpoint is score change from baseline.

Fatigue (PROMIS Cancer Fatigue SF):

For PROMIS Cancer Fatigue SF 7a, scores will be computed according to the procedures outlined in the PROMIS Fatigue scoring manual, available at https://www.assessmentcenter.net/Manuals.aspx. Higher scores are indicative of worse levels of fatigue. The Cancer Fatigue SF is described in detail in Section 13.0. The endpoint is score change from baseline.

4.3.3 Tertiary Efficacy Endpoints

The following are tertiary efficacy endpoints.

Clinical Benefit Rate (CBR):

CBR is defined as the proportion of subjects with documented MR or better.

Disease Control Rate (DCR):

DCR is defined as the proportion of subjects with documented MR or better or SD lasting at least 8 weeks.

Time to Response (TTR):

The TTR is defined as the number of days from the date of randomization to the date of first documented response (PR or better).

TTR (days) = Date of First Documented response (PR or better) – Date of Randomization + 1

Details on event and censoring rules for TTR are provided in Table 5.

Table 5. Event and Censoring Date Used in TTR

Situation	Date of Censor or Event	Outcome
Documented response at scheduled assessment date or before the next scheduled assessment	Date of response	Event
Start of new anti-cancer therapy prior to response	If no adequate assessment is available prior to start of new anti-cancer therapy then date of randomization. Otherwise, date of the last adequate assessment prior to start of new anti-cancer therapy	Censor
No documented response	If no adequate post baseline assessment is available then date of randomization. Otherwise, date of the last adequate assessment	Censor

Time to Next Treatment (TNT):

The TNT is defined as the number of days from the date of randomization to the starting date of new anti-cancer therapy (Other than study treatment).

TNT (days) = Starting date of any anti-cancer therapy – Date of Randomization + 1

Subjects not taking new anti-cancer therapy will be censored at the date of last contact.

Minimal Residual Disease (MRD) 10^{-4} and 10^{-6} negativity rates.

MRD negativity rate is defined as the proportion of subjects who have negative MRD by bone marrow aspirate at any time point after randomization and before progression or starting subsequent therapy. For the purpose of tertiary efficacy endpoints, MRD negativity will be defined at 10^{-4} and 10^{-6} thresholds (less than one residual myeloma cell per 10^4 and 10^6 total nucleated cells, respectively) as measured by centralized testing of bone marrow aspirate by NGS. MRD positive subjects include subjects of which all tested samples were found to be MRD positive or indeterminate. Subjects with missing or

unevaluable MRD status will be considered as MRD positive. The following tertiary efficacy analyses will be perform for MRD negativity.

1. MRD negativity based on 10^{-4} and 10^{-6} threshold: Rate of subjects with MRD negativity in the ITT analysis set
2. MRD negativity based on 10^{-4}, 10^{-5} and 10^{-6} threshold: Rate of subjects with MRD negativity along with CR/sCR in the ITT analysis set
3. MRD negativity based on 10^{-4}, 10^{-5} and 10^{-6} threshold: Rate of subjects with MRD negativity among the subjects with CR/sCR in the ITT analysis set

EORTC QLQ-MY20:

For EORTC QLQ-MY20, the endpoints are scores change from baseline. Scores will be summarized for four scales:

- Future perspective
- Body image
- Disease symptoms
- Side effects of treatment

Scores for each scale will be computed according to procedures outlined in EORTC QLQ-MY20 scoring manual, available at http://www.eortc.be/qol/files/ScoringInstructions/MY20_summary.pdf and Section 13.0.

EQ 5D-5L:

For EQ 5D-5L, the endpoints are scores change from baseline. Scores will be summarized for six scales:

- Mobility
- Self-care
- Usual activities

- Pain/Discomfort
- Anxiety/Depression
- EQ VAS score

Scores for each scale will be computed according to procedures outlined in EQ 5D-5L scoring manual, available at https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf.

Remaining Subscales of BPI-SF and EORTC QLQ-C30

Remaining subscales of BPI-SF are Pain Severity and Pain Interference. Scores for each subscale will be computed according to BPI-SF scoring manual, respectively. The endpoints are score change from baseline.

Remaining subscales of EORTC QLQ-C30 are

- Role Function
- Emotional Functioning
- Cognitive Functioning
- Social Functioning
- Fatigue
- Nausea and Vomiting
- Pain
- Dyspnea
- Insomnia
- Appetite Loss
- Constipation
- Diarrhea
- Financial Difficulties

Scores for each subscale will be computed according to EORTC QLQ-C30 scoring manual, respectively. The endpoints are score change from baseline.

4.3.4 Safety Endpoint

The following are safety endpoints.

- Adverse Events (AE)
- Serious Adverse Events (SAE)
- Deaths
- Electrocardiography (ECG)
- Vital sign parameters
- Laboratory results

4.4 Multiplicity Testing Procedures for Type-I Error Control

Statistical tests for selected secondary endpoints will be implemented in the testing strategy to maintain family-wise two-sided type I error rate at 0.05.[15,16] Each of these tests will be performed at an overall two-sided significance level of 0.05 (or equivalently one-sided significant level of 0.025). Statistical testing of the selected secondary endpoints will be performed only if the primary efficacy analysis of PFS is significant.

The following are selected secondary endpoints and testing order:

1. ORR
2. VGPR or better response rate
3. OS
4. Physical Functioning (EORTC QLQ-C30)
5. Worst Pain (BPI-SF)

The following testing procedure will be used for the selected secondary endpoints:

1. If the primary endpoint, PFS, is statistically significant then ORR will be tested;
2. If ORR is statistically significant then VGPR or better response rate will be tested;

3. If VGPR or better response rate is significant then OS will be tested using group sequential method. There will be two interim analyses of OS at the time of PFS analysis and at the time of observing approximately 75% of the total targeted OS events. The final OS analysis will be performed when approximately 116 OS events are available. The overall two-sided significance level for OS analysis will be kept at 0.05 (Equivalently one-sided significance level 0.025). Lan and DeMets α-spending method with O'Brien-Fleming type boundaries will be used to derive the significance level for OS at interim and final analyses.
4. If OS is significant then Physical Functioning (EORTC QLQ-C30) will be tested. This analysis will be based on data cut for the PFS analysis.
5. If Physical Functioning (EORTC QLQ-C30) is significant then Worst Pain (BPI-SF) will be tested. This analysis will be based on data cut for the PFS analysis.

4.5 Missing Data Imputation

No missing data imputation is planned for the study. Instead, missing data is addressed by adequate censoring as described in Section 4.3. In addition, missing data handling methods for PRO endpoints are provided in Section 13.0.

5.0 Analysis Populations

Intent-To-Treat (ITT) analysis set:

The ITT analysis set consists of all randomized subjects. The data from the ITT population will be analyzed by the treatment group assignment given at the time of randomization, even if the subject does not receive treatment, does not receive the correct treatment or is not compliant to the protocol procedures.

Safety analysis set:

The safety analysis set consists of all randomized subjects who take at least one dose of study drug (Venetoclax or Placebo). Subjects who received venetoclax at the first actual

dosing day will be considered as Venetoclax plus Bd group and otherwise they will be considered as Placebo plus Bd group. For IDMC review (refer to Section 4.2), the data from the safety population will be analyzed by the treatment group assignment given at the time of randomization.

6.0 Analysis Conventions

Data Cutoff Date and Data set

The cutoff date for the primary PFS analysis is Nov. 26, 2018. This cutoff date will be applied to all efficacy and safety analyses.

Only data with an assessment or event start date (e.g., vital sign assessment date or start date of an AE) prior to or on the cut-off date will be included in the analysis.

All events with a start date before or on the cut-off date and an end date after the cut-off date will be reported as 'continuing at the cut-off date.' The same rule will be applied to events starting before or on the cut-off date and not having a documented end date. This approach applies in particular to AE and concomitant medication reports. For these events, the end date will not be imputed and therefore will not appear in the listings.

If it is required to impute an end date to perform a specific analysis (e.g., for a dose administration record with a missing end date or an end date after the cut-off date, the cut-off date will be imputed as end date to allow the calculation of treatment exposure duration and dose intensity), the imputed date will be displayed and flagged in the listings.

Date of Randomization

The date of randomization is defined as the date that the Interactive Response Technology (IRT) issued a randomization number.

Definition of PFS event

The PFS event is derived from the following:

1. Death due to any cause is a PFS event.
2. A documented progressive disease is a PFS event if it is due to imaging per IMWG. The reason for any documented progressive disease will be provided by both IRC and Investigator, respectively.
3. A documented progressive disease is a PFS event if it is due to non-imaging per IMWG AND a consecutive assessment confirms the results. The reason for any documented progressive disease will be provided by both IRC and Investigator, respectively.

Date of Last Contact

The date of last contact will be derived from the following list of data panels:

- Vital signs
- Physical exam
- Lab variables, including SAE lab reports
- ECOG performance status
- PRO measures
- Study drug administration
- IMWG assessments
- Transfusions
- Electrocardiogram
- Adverse event
- Concomitant Medication
- Laboratory sample collection (e.g., biomarker)
- Survival follow-up

Definition of Study Drug and Study Treatment

1. Study drug refers to Venetoclax or Placebo.

2. Study treatment refers to
    - Venetoclax + Bortezomib + Dexamethasone
    - Placebo + Bortezomib + Dexamethasone
3. Study treatment components refer to
    - Venetoclax/Placebo or
    - Bortezomib or
    - Dexamethasone

Stratum information used for analyses

For all stratified analyses based on the ITT, stratification information will be used as provided by the IRT. The stratification factors are:

1. Prior exposure to proteasome inhibitors (naïve versus sensitive), and
2. Number of prior lines of therapy (1 versus 2 or 3)

Dealing with Multiple Values within the Same Visit

If there are multiple measurements/samples within the same post-baseline visit, the last measurement/sample within the visit (sorted by date and as available by time or repeat measurements) will be used in the analysis by visit or overall. For any analyses regarding outlier or abnormal assessments, all post-baseline values will be included (i.e., scheduled, unscheduled, repeat). For any other analyses, only scheduled visits will be included. This applies to quantitative and qualitative variables.

In addition, for PRO analysis, if more than one assessment is completed on the day of, or prior to a scheduled visit, the assessment performed closest to the scheduled visit will be used. In case the assessments are equidistant to the scheduled visit, the last assessment will be used for analyses. In case multiple assessments are performed on the same day and assessment time is not collected, the average of the assessment results will be used for analyses.

Definition of Baseline

For all analyses where baseline values are used, "baseline" refers to the last non-missing observation collected on or prior to the date of the first dose of any component of study treatment (for treated subjects) or the date of randomization (for non-treated subjects). Safety assessments that are related to a serious adverse event that occurred on the date of the first dose of any component of study treatment are excluded when applying this algorithm.

Definition of Final Observation

For PROs, laboratory, and vital signs variables, the final observation is defined as the last non-missing observation collected after baseline within 30 days following the last dose of study drug.

Definition of Study Rx Days (Days Relative to the First Dose of Study Treatment)

Study Rx Days are calculated for each time point of interest and provide a quantitative measure of days between the event and the first dose of any component of study treatment. If event date is earlier than the date of treatment start then Rx day = event date – treatment start date. Otherwise, Rx day = event date – treatment start date + 1.

Definition of Cycle Rx Days (Days Relative to the First Dose of Study Treatment in Each Cycle)

Cycle Rx Days for each cycle (cycle length for one cycle is defined as 21 days for Cycles 1 – 8, and as 35 days for Cycles 9 and beyond) are calculated for each time point relative to first dose of any component of study treatment in each cycle.

Definition of Analysis Windows

During the treatment period, all time points and corresponding time windows are based on Cycle Rx Days. For visit-wise longitudinal analyses of PROs the time windows are specified in Table 6.

Table 6. Time Windows for Longitudinal Analysis of PROs

Scheduled Visit	Nominal Cycle Rx Day	Time Window (Cycle Rx Day Range)
Cycle 1 Day 1	BASELINE	As baseline definition
Cycle 3 Day 1	1	(Day 2 of Cycle 1, Day 11 of Cycle 4)
Cycle 3 Day 1*	1	(Day 2 of Cycle 1, Day 21 of Cycle 3)
Cycle Y Day 1	1	(Day –10 of Cycle Y, Day 11 of Cycle Y + 1)
Cycle Y Day 1*	1	(Day –10 of Cycle Y, Day 21 of Cycle Y)
Cycle 9 Day 1	1	(Day -10 of Cycle 9, Day 17 of Cycle 10)
Cycle 9 Day 1*	1	(Day -10 of Cycle 9, Day 35 of Cycle 9)
Cycle Z Day 1	1	(Day –18 of Cycle Z, Day 17 of Cycle Z + 1)
Cycle Z Day 1*	1	(Day –18 of Cycle Z, Day 35 of Cycle Z)

Note: Y represents Cycles 5, and 7. Z represents odd cycles starting from Cycle 11.*: For subject who had last PRO assessment at the corresponding cycle or missing next cycle visit.

Definition of Creatinine Clearance (CrCl) and Body Surface Area (BSA)

CrCl = ((140 – age in years) × weight in kg × 0.85 if female) / (72 × serum creatinine in mg/dL)

BSA = $weight(kg)^{0.425} \times height(cm)^{0.725} \times 0.007184$

7.0 Subject Disposition

The ITT population will be used for the subject disposition summaries. The summary will be provided by treatment groups and overall population for all sites (pooled). Unless noted otherwise, there will be no statistical comparison between treatment groups. Number and percentage (%) of subjects for each of the following categories will be provided:

- Randomized subjects
- Treated subjects
- Discontinued study treatment
- Discontinued Venetoclax or Placebo

- Discontinued Bortezomib
- Discontinued Dexamethasone
- Discontinued from study
- Ongoing (Treatment, Progression Follow Up, Survival Follow Up)

In addition, the number and percentage of subjects who discontinued study treatment will be summarized by reason (primary and all reasons separately for Venetoclax or Placebo, Bortezomib and Dexamethasone, respectively).

8.0 Demographics, Baseline Characteristics, Medical History, and Previous/Concomitant Medications

The ITT population will be used for these summaries. Unless noted otherwise, there will be no statistical comparison between treatment groups for demographics, baseline characteristics, medical history or previous/concomitant medications.

8.1 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized by treatment groups and overall population. Continuous demographic data (e.g., age, height, and weight), shows in Table 7, will be summarized by the mean, standard deviation, median, minimum, and maximum values, which will be calculated using only the non-missing observations. Frequencies and percentages will be computed for categorical data, shows in Table 8.

Table 7. Demographic and Baseline Characteristics (Continuous Variables)

Continuous Variable
Age
Body weight
Body height
Time since diagnosis
Number of prior lines of therapies

Table 8. Demographic and Baseline Characteristics (Categorical Variables)

Categorical variable (possible values)
Age group (< 65 years/ ≥ 65 years)
Sex (Male/Female)
Race
Ethnicity
ECOG performance Status (0/1/2/ > 2/missing)
Prior lines of MM therapy (1/2/3/ > 3)
Prior exposure to proteasome inhibitors (PI) (sensitive/naïve)
Prior exposure to an immunomodulatory drug (IMiD) (naive/sensitive/refractory)
Prior exposure to bortezomib (sensitive/naïve)
Prior exposure to lenalidomide (naïve/sensitive/refractory)
Prior exposure to a PI and an IMiD (yes/no)
Prior exposure to an alkylating agent (yes/no)
Prior exposure to an anti-CD38 monoclonal antibody (yes/no)
Disease refractory to last line of therapy (yes/no)
Multiple Myeloma ISS stage (I/II/III)
Soft tissue plasmacytomas present at screening (yes/no)
Bone and lytic lesions present at baseline (yes/no)
Prior stem cell transplant (autologous/allogeneic/syngeneic)
t (11;14) status (negative/positive/unknown)
t (4;14) status (negative/positive/unknown)
t (14;16) status (negative/positive/unknown)
t (14;20) status (negative/positive/unknown)
17p status (deleted/not deleted/unknown)
1q gain (present/not present/unknown)
Ch9 gain OR Ch11 gain OR Ch15 gain (present/not present/unknown)
Geography (North America/EU/South America/Asia-Pacific)
Creatinine clearance (< 30/30 - < 60/60 - < 90/ ≥ 90 mL/min)
Type of measurable disease per CRF (IgG/IgA/Other)
IMWG consensus risk (High/Standard)
Chromosomal Abnormality (CA) risk by FISH (High/Standard)

Risk stratification per IMWG consensus will be defined according to the IMWG consensus on risk stratification in multiple myeloma:[4]

IMWG consensus High-risk:

- ISS staging at screening of Stage II or III AND
- t (4;14) positive OR 17p deletion

IMWG consensus Standard-risk:

- ISS staging at screening of Stage I OR
- ISS staging at screening of Stage II or III AND t (4;14) negative AND 17p not deleted

Risk stratification per chromosomal abnormality (CA) by FISH will be defined:[17]

CA High-risk:

- t (4;14) positive OR t (14;16) positive OR 17p deletion

CA Standard-risk:

- t (4;14) negative AND t (14;16) negative AND 17p not deleted

For risk stratification, subjects with missing ISS staging or missing or indeterminate t (4;14), t (14;16) or 17p status that do not allow for a categorization per above will be excluded from summary statistics.

8.2 Medical History

Medical history and ongoing conditions, including MM-related conditions and symptoms will be summarized and listed.

8.3 Prior/Concomitant Medications and Prior Multiple Myeloma Therapies

Concomitant Medications:

Concomitant medications are defined as medications (other than study treatment) taken after the first dose of any component of study treatment and within 30 days of the last dose of any component of study treatment. Concomitant medications will be reported by generic name assigned by the World Health Organization (WHO) dictionary.

The number and percentage of subjects who take at least one concomitant medication will be summarized. For summaries of concomitant medications, if an incomplete start date was collected for a medication, the medication will be assumed to be a concomitant medication unless there is evidence that confirms that the medication was not a concomitant medication (e.g., the medication end date was prior to the first dose of study drug).

Prior medication:

A prior medication is defined as any medication taken prior to the first dose of any component of study drug and collected in the CRF. The number and percentage of subjects who take at least one prior medication will be summarized.

8.4 Other

The number of patients randomized will be provided by country, investigational site and treatment group. The number and percentage of patients in each analysis set (ITT and safety) will be summarized by stratum and treatment group.

Moreover, frequencies will be provided for by treatment group and stratification factors. A summary for the primary reason(s) for screening failure will be presented.

All data collected at baseline, including the source of subject referral, childbearing potential, and pregnancy test results, will be listed. Demographic information on screening failures will be listed. Patients who have received treatment different from

intended treatment as determined by the IRT as well as patients who have been randomized according to incorrect stratum will be listed.

9.0 Study Treatment Exposure and Compliance

Analyses of study treatment exposure and compliance listed below will be performed on the safety analysis set by treatment groups. Unless noted otherwise, there will be no statistical comparison for these analyses.

9.1 Exposure to Study Treatment

Duration of study treatment exposure

Duration of study treatment exposure will be calculated as the number of days between the start and end of the study treatment.

Duration of study treatment exposure (days) = date of last administration of study treatment – date of first administration of study treatment + 1

For Risk Management Plan: Patient-years of exposure

Patient-years of exposure is computed as:

Patient-years of exposure (pt-yrs) = Sum of duration of study treatment for all subjects (day)/365.25

Duration of study treatment component exposure

Duration of study treatment component exposure will be calculated as the number of days between the start and end dates for the study treatment component

Duration of study treatment component exposure (days) =
Date of last administration study treatment component – Date of first administration study treatment component + 1

Cumulative dose

The cumulative dose is defined as the total dose given during the study treatment exposure and will be summarized by study treatment component and treatment group. For patients who did not take any drug the cumulative dose is by definition equal to 0.

Dose intensity (DI) and relative dose intensity (RDI)

The following DI or planned DI (PDI) (per protocol) calculations apply for both IV and oral dosing. In the case of IV dosing, the dose was given based on BSA calculation of DI or PDI involves BSA in the division. The RDI will be summarized separately for each of the study treatment components.

Venetoclax/placebo and Dexamethasone

$$\mathrm{DI(mg/day)} = \frac{\text{Cumulative dose [mg]}}{\sum_{c=1}^{k} \text{Actual duation of cycle c [day]}}$$

where k = total number of cycles.

PDI is defined as follows:

1. The last cycle length is not shorter than planned cycle length

$$\mathrm{PDI(mg/day)} = \frac{\text{Cumulative planned dose [mg]}}{\sum_{c=1}^{k} \text{Planned duration of cycle c [day]}}$$

2. The last cycles length is shorter than planned cycle length

$$\mathrm{PDI(mg/day)} = \frac{\begin{array}{c}\text{Total planned dose for first k} - 1 \text{ cycles [mg]} \\ +\text{total planned dose for actual duration of the last cycle k}\end{array}}{\sum_{c=1}^{k-1} \text{Planned duration of cycle c} + \text{Actual duration of the last cycle [day]}}$$

k is the actual total number of cycles.

RDI is defined as

$$\text{RDI}[\%] = \frac{\text{DI [mg/day]}}{\text{PDI [mg/day]}} \times 100$$

Bortezomib

Dose intensity (DI) over study period for bortezomib is defined as

$$\text{DI}(\text{mg/m}^2\text{/day}) = \frac{\sum_{c=1}^{k} \frac{\text{Total actual dose in cycle c (mg)}}{\text{BSA}_c(\text{m}^2)}}{\sum_{c=1}^{k} \text{Actual duration of cycle c (days)}}$$

where k = total number of cycles.

1. The last cycle length is not shorter than planned cycle length

$$\text{PDI}(\text{mg/m}^2\text{/day}) = \frac{\sum_{c=1}^{k} \text{Total planned dose for cycle c [mg/m}^2\text{]}}{\sum_{c=1}^{k} \text{Planned duration of cycle c (days)}}$$

2. The last cycle length is shorter than planned cycle length

$$\text{PDI}(\text{mg/m}^2\text{/day}) = \frac{\sum_{c=1}^{k-1} \text{Total planned dose for cycle c [mg/}m^2\text{]} + \text{Total planned dose for actual duration of the last cycle k [mg/m}^2\text{]}}{\sum_{c=1}^{k-1} \text{Planned duration of cycle c} + \text{Actual duration of the last cycle k (days)}}$$

where k = total number of cycles.

RDI is defined as

$$\text{RDI [\%]} = \frac{\text{DI [mg/}m^2\text{/day]}}{\text{PDI [mg/}m^2\text{/day]}} \times 100$$

The RDI will be summarized based on eCRF data on the drug administration pages. All analyses will be provided by treatment groups for each study treatment component.

Descriptive statistics, including the mean, standard deviation, median, minimum, and maximum will be presented for duration of exposure and RDI. Number and percentage of subjects who have dose reductions or interruptions, as well as their reasons, will be provided. Number of cycles of study treatment received will be provided. In addition summary statistics of cumulative dose and frequencies for RDI (%) will be provided by treatment group for the following categories:

- 0 to < 50%
- 50% to < 70%
- 70% to < 90%
- 90% to < 110%
- ≥ 110%

9.2 Compliance

To assess compliance, a listing of protocol deviations will be given.

10.0 Efficacy Analyses

10.1 General Considerations

Unless noted otherwise, all efficacy analyses will be performed by treatment group. The following stratification factors will be considered when applicable:

1. Prior exposure to proteasome inhibitors (naïve versus sensitive), and
2. Number of prior lines of therapy (1 versus 2 or 3)

If a subject has the blind prematurely broken before the cutoff date used for data analysis, the date of the break will be set as the cutoff date for this subject.

For time-to-event endpoints, survivorship functions will be estimated by using Kaplan-Meier product-limit methodology. Estimated survival curves will be presented. If

reached, median time to event and its two-sided 95% confidence interval will be presented.

10.2 Primary Efficacy Analysis

PFS per IRC assessment is the primary efficacy endpoint. The study is designed to test the following primary statistical hypothesis:

$$H_0: S(t)_{\text{Venetoclax + Bd}} = S(t)_{\text{Placebo + Bd}} \quad \text{vs.} \quad H_1: S(t)_{\text{Venetoclax +Bd}} \neq S(t)_{\text{Placebo + Bd}}$$

where S(t) is the survivorship function of PFS at time t. A stratified log-rank test with a two-sided type-I error rate of $\alpha = 0.05$ (one-sided type I error rate of $\alpha = 0.025$) will be used to test the primary null hypothesis. The hazard ratio (HR) for treatment effect will be estimated and its two-sided 95% confidence interval will be provided. The estimation will be based on a Cox proportional hazards model with treatment and stratification factors included.

As sensitivity analyses, the primary endpoint PFS will be analyzed with the following changes:

1. PFS based on assessment by investigator.
2. Not considering two or more consecutive missing assessment as a reason for censoring PFS events (per IRC).
3. Not considering new anticancer therapy as a reason for censoring PFS events (per IRC).

A robustness analysis will be performed including the following prognostic factors into the Cox proportional hazards model: treatment group, and other covariates including, but not limited to, age group (< 65 years/ ≥ 65 years), renal function per Creatinine Clearance, prior stem cell transplant (yes/no), and MM ISS stage (I/II/III) as covariates. The stratification factors will be reflected in the STRATA statement. HR and 2-sided 95% CIs

will be provided. Covariates with a large amount of missing data or small number of subjects within a category may be excluded or combined as appropriate.

10.3 Secondary Efficacy Analyses

VGPR or better response rate, MRD negativity rate and ORR will be analyzed by using stratified Cochran Mantel Haenszel test based on strata at randomization. Estimated rates along with corresponding exact 95% confidence intervals as derived by the Clopper-Pearson method will be presented.

DOR and TTP will be analyzed by using stratified log-rank tests. The HR for treatment effect will be estimated and two-sided 95% confidence intervals will be provided. The estimation will be based on a Cox proportional hazards model that includes treatment and stratification factors. DOR will be analyzed only based on data from responders (PR or better) out of the ITT analysis set.

OS will be analyzed by using a stratified, two-sided, 3-look group sequential log-rank test with a cumulative two-sided type-I error rate of $\alpha = 0.05$ (one-sided type I error rate of $\alpha = 0.025$). The HR for treatment effect will be estimated and two-sided 95% confidence intervals will be provided. The estimation will be based on a Cox proportional hazards model that includes treatment and stratification factors.

For PROs, change in score from baseline will be statistically compared between treatment arms for the following: Global Health Status/QoL (EORTC QLQ C30), Worst Pain (BPI-SF), Physical Functioning (EORTC QLQ-C30), and Fatigue (PROMIS-Fatigue). In order to analyze PRO variables, assessments will be time plotted using time-windows (Table 6). Change from baseline will be analyzed by a repeated measurement model. Stratification factors, age group and treatment group will be included as fixed factors and the baseline value as a covariate. Furthermore, time and treatment by time interaction will be included in the model. The repeated correlation structure in the timepoints will be assessed by using the Bayesian Information Criterion (BIC). The following covariance structures will be explored: Unstructured (TYPE = UN), compound symmetry (TYPE = CS) and

first-order autoregressive (TYPE = AR(1)). The type resulting in the lowest BIC will be used for analysis. Treatment and treatment by time interaction will tested simultaneously at two sided significance level 0.05 to test treatment effect. Also, unless noted otherwise, for EORTC and PROMIS endpoints, data collected on or before start date of new cancer therapy will be considered.

10.4 Tertiary Efficacy Analyses

CBR, DCR and MRD negativity rate will be analyzed using Cochran-Mantel-Haenszel test based on strata at randomization. Estimated rates along with corresponding exact 95% confidence intervals as derived by the Clopper-Pearson method will be presented.

TTR and TNT will be analyzed using stratified log-rank tests. The HR for treatment effect will be estimated and two-sided 95% confidence intervals will be provided. The estimation will be based on a Cox proportional hazards model that includes treatment and stratification factors.

For PROs, descriptive statistics will be used to summarize change from baseline in patient-reported outcomes based on the remaining subscales of BPI-SF, EORTC QLQ-C30, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Multiple Myeloma Module [EORTC QLQ-MY20], and Euroqol EQ-5D-5L.

10.5 Efficacy Subgroup Analyses

To evaluate the impact of demographic and baseline characteristics on the primary efficacy endpoint, the following subgroup analyses will be performed:

- Disease refractory to last line of therapy (yes/no)
- Prior exposure to proteasome inhibitors (naïve/sensitive)
- Prior exposure to an immunomodulatory drug (IMID) (naive/sensitive/refractory)
- Prior exposure to bortezomib (naïve/sensitive)

- Prior exposure to lenalidomide (naive/sensitive/refractory)
- Prior exposure to a proteasome inhibitor and an IMiD (yes/no)
- Prior exposure to an alkylating agent (yes/no)
- Number of prior lines of therapy (1/2 or 3)
- Age group (< 65 years/ ≥ 65 years)
- Sex (male/female)
- Race (White/Black/Asian/Other)
- Multiple Myeloma ISS stage (I/II/III)
- Renal function at baseline (Severe impairment (CrCl < 30 mL/min); moderate impairment (CrCl 30 - < 60 mL/min); mild impairment (CrCl 60 - < 90 mL/min); normal function (CrCl ≥ 90 mL/min))
- IMWG consensus risk (High/Standard)
- CA risk (High/Standard)
- t (11;14) status (negative/positive/unknown)
- Prior stem cell transplant (yes/no)
- Geographic region (North America/EU/South America/Asia-Pacific)
- BCL-2 expression (high/low) as defined through central laboratory testing by immunohistochemistry (IHC) and based on the pre-specified scoring algorithm in Table 2. Subjects who are randomized but have missing or indeterminate central laboratory IHC results for BCL2 expression will be excluded for the subgroup analysis.
- BCL2 gene expression (≥ cutoff/ < cutoff) as defined through central laboratory testing of bone marrow aspirates by quantitative polymerase chain reaction (qPCR) using the ΔCt method. High BCL2 expression will be defined as BCL2 expression normalized to TMEM55B ($2^{-\Delta Ct}$) greater than or equal to the 1st, 2nd, and 3rd quartile expression values calculated from randomized subjects with central laboratory qPCR results for BCL2 gene expression. ΔCt = Ct BCL2 (target gene) – Ct TMEM55B (reference gene). Subjects who are randomized but have missing or indeterminate central laboratory qPCR results for BCL2 gene expression will be excluded for the subgroup analysis.

- BCL2:BCL2L1 gene expression (≥ cutoff/ < cutoff) as defined through central laboratory testing of bone marrow aspirates by quantitative polymerase chain reaction (qPCR) using the ΔCt method. High BCL2:BCL2L1 expression will be defined as BCL2 expression relative to BCL2L1 ($2^{-\Delta\Delta Ct}$) greater than or equal to the 1st, 2nd, and 3rd quartile expression values calculated from randomized subjects with central laboratory qPCR results for BCL2 and BCL2L1 gene expression. ΔCt for BCL2 = Ct BCL2 (target gene) – Ct TMEM55B (reference gene). ΔCt for BCL2L1 = Ct BCL2L1 (target gene) – Ct TMEM55B (reference gene). ΔΔCt for BCL2:BCL2L1 = ΔCt BCL2 - ΔCt BCL2L1. Subjects who are randomized but have missing or indeterminate central laboratory qPCR results for BCL2 and BCL2L1 gene expression will be excluded for the subgroup analysis.
- BCL-2 and BCL-XL expression levels (positive profile/negative profile) as defined through central laboratory testing of bone marrow core biopsy tissue by immunohistochemistry (IHC) and the pre-specified scoring algorithm in Table 2. Two subgroups will be defined as follows:
  - Positive Profile: High BCL-2 and Low BCL-XL
  - Negative Profile: High BCL-2 and High BCL-XL OR Low BCL-2 and High BCL-XL OR Low BCL-2 and Low BCL-XL.

Subjects who are randomized but have missing or indeterminate central laboratory IHC results for BCL-2 and BCL-XL expression will be excluded for the subgroup analysis. HRs and 2-sided 95% CIs will be provided based on the Cox proportional hazards model. This model will include the factors treatment and the stratification factors using the BY statement to derive estimates by subgroup. Point estimates of HR and 2-sided 95% CI will be provided by subgroup in a forest plot.

11.0 Safety Analyses

11.1 General Considerations

Safety analyses will be performed on the Safety Analysis Set by treatment group. There will be no statistical comparison for these analyses.

For basophils, eosinophils, lymphocytes, monocytes, neutrophils and reticulocytes, reporting in tables will be done only for absolute values. If sites have provided only a % and no absolute value, conversion from % to absolute will be performed according to the following formulas:

- Basophils [absolute value] = basophils [%] • leukocytes [absolute value], same for eosinophils, lymphocytes, monocytes, and neutrophils
- Reticulocytes [absolute value] = reticulocytes [%] • erythrocytes [absolute value]

Reference intervals for absolute values will be used (i.e., reference intervals will not be converted).

For urea, reporting in tables will be done only as blood urea nitrogen (BUN). In case sites have provided only a urea value and no BUN value, conversion from Urea to BUN will be performed according to the following formula:

- BUN [mmol/L] = urea [mmol/l]/2.14

11.2 Analysis of Adverse Events

Analyses of adverse events will include only "treatment-emergent" events. Treatment-emergent adverse events (TEAE) are those which either begin or increase in severity after the first dose of study treatment. Operationally, TEAE is defined as any adverse event that has an onset on or after the day of the first dose of the study treatment. Analyses will not include those TEAE that have an onset greater than 30 days after the last dose of the study drug. Events for which the onset date is the same as the study treatment start date are assumed to be treatment-emergent. If an incomplete onset date was collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event was not treatment-emergent (e.g., the event end date was prior to the study treatment start date).

11.2.1 Treatment-Emergent Adverse Events

The number and percentage of subjects experiencing treatment-emergent adverse events will be summarized for the following categories:

- Adverse events by primary system organ class and preferred term
- Adverse events by primary system organ class and preferred term, by subgroups (Section 11.2.6)
- Adverse events with NCI-CTCAE toxicity grade 3 or 4 by primary system organ class and preferred term
- Adverse events with NCI-CTCAE toxicity grade 3, 4 or 5 by primary system organ class and preferred term
- Serious adverse events by primary system organ class and preferred term
- Adverse events with suspected relationship to Venetoclax/Placebo by primary system organ class and preferred term
- Adverse events with suspected relationship to Bortezomib by primary system organ class and preferred term
- Adverse events with suspected relationship to Dexamethasone by primary system organ class and preferred term
- Adverse events leading to death by primary system organ class and preferred term
- Adverse events leading to Venetoclax/Placebo discontinuation by primary system organ class and preferred term
- Adverse events leading to Bortezomib discontinuation by primary system organ class and preferred term
- Adverse events leading to Dexamethasone discontinuation by primary system organ class and preferred term
- Adverse events leading to Venetoclax/Placebo dose reduction by primary system organ class and preferred term
- Adverse events leading to Bortezomib dose reduction by primary system organ class and preferred term

- Adverse events leading to Dexamethasone dose reduction by primary system organ class and preferred term
- Adverse events leading to Venetoclax/Placebo temporary interruption by primary system organ class and preferred term
- Adverse events leading to Bortezomib temporary interruption by primary system organ class and preferred term
- Adverse events leading to Dexamethasone temporary interruption by primary system organ class and preferred term
- Any adverse event of neutropenia with NCI-CTCAE toxicity grade 3 or 4 and concurrent serious adverse event of infection
- Any adverse event of thrombocytopenia with NCI-CTCAE toxicity grade 3 or 4 and concurrent serious adverse event of hemorrhage
- Any adverse event of lymphopenia and concurrent serious infection
- Any adverse event of neutropenia with NCI-CTCAE toxicity grade 3 or 4 leading to Venetoclax or Placebo discontinuation
- Selected adverse events by preferred terms (Table 9)

Table 9. Selected Adverse Events

Risk	Search Criteria
Tumor Lysis Syndrome	SMQ – "Tumour lysis syndrome" (Narrow-scope)
Neutropenia Expanded Search	PT terms – "Neutropenia," "Neutrophil count decreased," "Febrile neutropenia," "Agranulocytosis," "Neutropenic infection," and "Neutropenic sepsis"
Neutropenia and Neutrophil Count Decreased	PT terms – "Neutropenia," "Neutrophil count decreased"
Serious Infection	SAEs in the SOC of "Infections and Infestations"
Second Primary Malignancy	SMQ – "Malignant tumours" (Narrow) and "Myelodysplastic syndromes" (Narrow)
Lymphopenia	PT terms – "Lymphopenia" and "Lymphocyte count decreased"
Anemia	PT terms – "Anaemia" and "Haemoglobin decreased"
Thrombocytopenia	PT terms – "Thrombocytopenia" and "Platelet count decreased"
Drug Induced Liver Injury (DILI)	SMQ – "Drug related hepatic disorders – comprehensive search"
Peripheral Neuropathy	Peripheral Neuropathy SMQ – "narrow"
Sepsis	PTs including the word 'sepsis' + PT 'septic shock'
Serious Opportunistic infections	Opportunistic infection _ CMQ (BCL2 inhibitor product specific)

An AE will be considered "concurrent" with a cytopenia event (eg, serious infection concurrent with neutropenia expanded search) if the onset of the AE was no more than 7 days prior to the onset of the cytopenia event and no more than 7 days after the end of the cytopenia event.

Adverse Events by Primary System Organ Class and Preferred Term

The number and percentage of subjects experiencing adverse events will be tabulated according to the primary MedDRA system organ class (SOC) and MedDRA preferred term (PT). Subjects reporting more than one adverse event for a given MedDRA PT will be counted only once for that term. Subjects reporting more than one type of adverse event within a MedDRA SOC will be counted only once for that SOC. Subjects reporting more than one type of adverse event will be counted only once in the overall total.

Adverse Events by SOC/PT and by NCI-CTCAE Toxicity Grade

Adverse events will also be summarized by SOC/PT and by maximum NCI-CTCAE toxicity grade. If a subject has an adverse event with an unknown NCI-CTCAE toxicity grade, then the subject will be counted in the category of "unknown," even if the subject has another occurrence of the same adverse event with a grade present. The only exception is if the subject has another occurrence of the same adverse event with the most extreme grade. In this case, the subject will be counted under this grade.

11.2.2 Serious Adverse Events

Serious adverse events will be summarized by System Organ Class and Preferred Term using the same methods as adverse events described above in Section 11.2.1.

11.2.3 Adverse Events Leading to Discontinuation, Dose Reduction, or Dose Interruption of Study Treatment and Leading to Death

The number and percentage of subjects experiencing adverse events leading to discontinuation, dose reduction, or dose interruption of Venetoclax or Placebo/Bortezomib/Dexamethasone will be summarized by SOC/PT separately for all combinations.

The number and percent of subjects experiencing adverse events leading to death will be tabulated according to the primary SOC and PT.

11.2.4 Deaths

The number of subject deaths will be summarized for:

- Deaths occurring while the subject was still receiving study drug
- Deaths occurring off treatment within 30 days after the last dose of study drug
- Deaths in this study regardless of the number of days after the last dose of study drug

11.2.5 Listing of Adverse Events

The following additional summaries of adverse events will be prepared:

- Listing of all adverse events
- Listing of all serious adverse events
- Listing of Grade 3 - 4 adverse events leading to death
- Listing of all adverse events leading to death
- Subject deaths occurring ≤ 30 days after last dose of study drug
- Subject deaths occurring > 30 days after last dose of study drug

11.2.6 Adverse Event Rates by Subgroup

Treatment-emergent adverse events may be assessed for the subgroups defined below:

- Age (< 65 years/ ≥ 65 years))
- Gender
- Race (white, black, Asian, other)
- Region (US, EU, and ROW)
- Renal Function at Baseline:
  - Normal renal function: Creatinine Clearance (CrCl) ≥ 90 mL/min;
  - Mild renal impairment: CrCl ≥ 60 to < 90 mL/min
  - Moderate renal impairment: CrCl ≥ 30 to < 60 mL/min
  - Severe renal impairment: CrCl < 30 mL/min
- Hepatic Function at Baseline:
  - Normal hepatic function: total bilirubin ≤ 1 mg/dL and Aspartate Amino Transferase (AST) ≤ 40 U/L
  - Mild hepatic impairment: total bilirubin ≤ 1 mg/dL and AST > 40 U/L, or total bilirubin > 1.0 to ≤ 1.5 mg/dL and any AST
  - Moderate hepatic impairment: total bilirubin > 1.5 to ≤ 3 mg/dL and any AST
  - Severe hepatic impairment: total bilirubin > 3 mg/dL and any AST

11.3 Analysis of Laboratory and Vital Signs

Changes from baseline may be summarized for each scheduled post-baseline visit and for the Treatment Completion Visit for hemoglobin, platelet count, neutrophil count, absolute leukocyte count, creatinine, and calcium. Post-baseline measurements more than 30 days after the last dose of randomized study drug will not be included. Subjects that do not have a baseline measurement or do not have any post-baseline measurements will not be included.

Descriptive statistics will include the mean, standard deviation, and median for baseline, and mean, standard deviation, median, minimum, and maximum for change from baseline for each scheduled post-baseline visit, and the Treatment Completion Visit by treatment group.

11.3.1 Analyses of Laboratory Data Using NCI CTCAE

For hematology and chemistry variables for which NCI CTCAE 4.03 criteria exist, baseline and post-baseline hematology and chemistry variable observations will be categorized as grade 0 to grade 4.

Criteria are specified for the assignment of grades with values between 1 and 4. The criteria are unidirectional: any one set of criteria constitute a screening either for low or high values of potential clinical significance.

For laboratory tests for which a normal range limit is one end of the grade 1 range then values that are either within the normal range or outside it in direction opposite the test will be classified as grade 0 values. For other tests, values outside the grade 1 range in the direction opposite that of the test will be classified as grade 0.

There can be instances in which the criteria for more than one grade apply to a lab test value. In those instances the highest applicable grade will be assigned.

For each graded variable, shifts from baseline grade to treatment-emergent maximum post-baseline grades (but no more than 30 days after the last dose of study drug) will be assessed.

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

11.3.2 Assessment of Drug-Induced Liver Injury

The number and percentage of subjects in each treatment group who have at least one observed post-baseline value meeting the following criteria will be tabulated

- Alanine Amino Transferase (ALT) > 3 × Upper Limit Normal (ULN)
- AST > 3 × ULN
- AST or ALT > 3 × ULN
- Total bilirubin value > 2 × ULN
- ALT or AST > 3 × ULN AND Total bilirubin > 2 × ULN within 72 hours.

A listing of all observed ALT, AST, and total bilirubin values will be generated for subjects with an observed value meeting any of these criteria.

11.3.3 Laboratory Assessments for Tumor Lysis Syndrome

To determine if a subject's laboratory values qualify for Tumor Lysis Syndrome (TLS), the Howard criteria[2] will be assessed. The Howard definition for laboratory TLS requires ≥ 2 of the metabolic abnormalities specified in Table 10 post-baseline and within 24 hours of each other.

Table 10. Laboratory Criteria for TLS

Element	Value
Uric Acid	> 476 µmol/L
Potassium	> 6.0 mmol/L
Inorganic Phosphorus	> 1.5 mmol/L
Calcium	< 1.75 mmol/L

The following summaries of laboratory criteria will be provided:

- Number and percentage of subjects meeting the definition of laboratory TLS (at least two values meeting the criteria in Table 10, occurring within 24 hours of each other).
- Listing of all values for these four analyses for each subject meeting the definition of laboratory TLS at least once during treatment.

11.3.4 Analyses of Vital Signs Using Criteria for Potentially Clinically Significant Values

Vital sign variables are: sitting systolic blood pressure, sitting diastolic blood pressure, pulse and temperature.

Vital signs values will be assessed for potential clinical significance (PCS) through the application of criteria developed at AbbVie as detailed in the following table:

Table 11. Criteria for Potential Clinical Significance in Vital Signs Variables

Vital Signs Variables	Criterion	Definition of Potentially Clinically Significant
Systolic Blood Pressure	High	Value ≥ 160 mmHg
Diastolic Blood Pressure	High	Value ≥ 100 mmHg
Pulse	Low	Value < 50 bpm
	High	Value ≥ 120 bpm
Temperature	Low	Value < 36°C
	High	Value ≥ 38.5°C

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

11.4 Analysis of ECG Parameters

A summary of the number and percentage of subjects who have at least one observation in the following categories will be provided:

- abnormal ECG – clinically significant (Yes/No)

A listing will also be prepared that will include details about all abnormal ECG findings.

12.0 Pharmacokinetics

Plasma concentrations of Venetoclax and Bortezomib (and Dexamethasone, Japan subjects only) will be tabulated for each subject by visit, and summary statistics will be computed for each sampling time. Samples with significant sampling time deviations will be excluded from summary statistics calculations.

For the Japan subjects during the run-in phase, pharmacokinetic parameter values for Dexamethasone will be tabulated for each subject by visit, and summary statistics will be computed for each parameter.

13.0 Appendix – PRO assessments

PRO assessments include: BPI-SF, EORTC QLQ-C30, EORTC QLC MY20, PROMIS Cancer Fatigue SF, and EQ-5D-5L.

BPI-SF

The BPI-SF is a pain-specific measure developed to assess patient-reported severity (or intensity) of pain (4 items) and the impact of pain on daily functioning (7 items) in patients with cancer pain.[7] The four pain severity items assess pain at its "worst," "least," "average," and "now" (current pain). For these items, patients are asked to rate their pain

on an 11-point numeric rating scale with anchors of 0 (no pain) and 10 (pain as bad as you can imagine). The BPI "worst" pain severity item has been shown to be reliable and valid for use as a single item.[8] The BPI-SF also includes questions to measure the interference of pain in the patient's daily life, including general activity, mood, ability to walk, normal work (both outside the home and housework), relationships with other people, sleep, and enjoyment of life. For these items, patients are asked to describe the extent to which pain has interfered on an 11-point numeric rating scale with anchors of 0 (does not interfere) to 10 (completely interferes).

Table 12. BPI-SF Scales

Scales	Items
Pain Severity*	BPI2-Pain Right Now
	BPI2-Pain at its Least in Last 24 Hours
	BPI2-Pain on the Average
	BPI2-Pain at its Worst in Last 24 Hours
Pain Interference**	BPI2-Pain Interfered Enjoyment of Life
	BPI2-Pain Interfered General Activity
	BPI2-Pain Interfered Walking Ability
	BPI2-Pain Interfered with Mood
	BPI2-Pain Interfered with Normal Work
	BPI2-Pain Interfered with Relations
	BPI2-Pain Interfered with Sleep

*Calculated as sum of scores of all 4 items and then dividing by 4. This gives the Pain Severity score out of 10.
**Calculated as sum of scores of all 7 items and then dividing by 7. This gives the Pain Severity score out of 10.

EORTC QLQ-C30

HRQoL, functioning, and symptoms will be assessed with the EORTC-QLQ-C30 version 3.[9] The QLQ-C30 is a 30-item subject self-report questionnaire composed of both multi-item and single scales, including five functional scales (physical, role, emotional, social, and cognitive), three symptom scales (fatigue, nausea and vomiting, and pain), a global health status/quality of life scale, and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). The subjects rate items on a

four point scale, with 1 as "not at all" and 4 as "very much." The QLQ-C30 was developed and validated for use in a cancer patient population, and its reliability and validity is highly consistent across different language cultural groups. The items and scales of the QLQ-C30 are as in Table 13:

Table 13. EORTC QLQ-C30 Scales

	Scale	Number of Items	Item Range	Item Numbers
Global Health Status/Quality of Life				
Global Health Status/Quality of Life	QL2	2	6	29, 30
Functional Scales				
Physical Functioning*	PF2	5	3	1 – 5
Role Function*	RF2	2	3	6 – 7
Emotional Functioning*	EF	4	3	21 – 24
Cognitive Functioning*	CF	2	3	20, 25
Social Functioning*	SF	2	3	26 – 27
Symptom Scales				
Fatigue	FA	3	3	10, 12, 18
Nausea and Vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite Loss	AP	1	3	13
Constipation	CO	1	3	10
Diarrhoea	DI	1	3	17
Financial Difficulties	FI	1	3	28

* Function Scales.

Note: Item range is the difference between the possible maximum and the minimum response to individual items.

Scoring algorithms for scales are as follows:

For all Scales, the Raw Score (RS) is the mean of the component items (I):

If items I_1, I_2…, I_n are included in a scale, the procedures are as follows:

$$\text{Raw Score (RS)} = (I_1 + I_2 + \ldots + I_n) / (\text{number of non-missing items}),$$

Then the Functional Scales:

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

where range is provided in Table 13.

For example, to calculate Emotional Functioning (EF),

$$RS_{EF} = (I_{21} + I_{22} + I_{23} + I_{24}) / 4$$

$$Score_{EF} = (1 - (R_{SEF} - 1) / 3) \times 100$$

If a subject completed more than 50% of the items in a scale, then the raw score of that subject will contribute to the summary statistics of that scale. If a subject completed less than 50% of the items in a scale, then the raw score of that subject will be dropped from the calculation of the summary statistics of that scale.

EORTC QLQ-MY20

The EORTC QLQ-MY20 was developed as an additional module for the QLQ-C30 and is composed of 20 items specific to multiple myeloma.[10] The QLQ-MY20 includes scales for disease symptoms, side effects of treatment, future perspective, and body image. Values for each scale range from 0 to 100. The subjects rate items on a four-point scale, with 1 as "not at all" and 4 as "very much." The QLQ-MY20 is a reliable and valid instrument for measuring quality of life in myeloma patients.

Table 14. QLQ-MY20 scales

	Scale	Number of Items	Item Range	QLQ-MY20 Item Numbers
Functional Scales/items				
Future perspective	MYFP	3	3	18 - 20
Body image	MYBI	1	3	17
Symptom Scales				
Disease symptoms	MYDS	6	3	1 - 6
Side effects of treatment	MYSE	10	3	7 - 16

PROMIS Cancer Fatigue SF

The PROMIS® is a system of highly reliable, precise measures of patient-reported health status for physical, mental, and social well-being.[11] PROMIS instruments measure concepts such as pain, fatigue, physical function, depression, anxiety and social function. Fatigue will be assessed using the PROMIS Cancer Fatigue SF that has been developed for use in oncology populations.[12,13] PROMIS Cancer Fatigue SF is a seven item questionnaire that assesses the impact and experience of fatigue over the past 7 days. All questions employ the following five response options: 1 = Not at all, 2 = A little bit, 3 = Somewhat, 4 = Quite a bit, and 5 = Very much. The score is calculated as below:

Score = Sum of score from all questions × 7/Total number of answered questions

If 4 or more questions were not being answered, then the score is considered as missing. The details of scoring method can be found in the scoring manual https://www.assessmentcenter.net/documents/PROMIS%20Fatigue%20Scoring%20Manual.pdf.

EQ-5D-5L

The EQ-5D-5L is a generic preference instrument that has been validated in numerous populations.[14] The EQ-5D-5L has five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These dimensions are measured on a five level

scale: no problems, slight problems, moderate problems, severe problems, and extreme problems. The EQ-5D-5L also contains a visual analog scale (VAS) to assess the subject's overall health. Detailed EQ-5D-5L scoring method can be found at https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf.

14.0 References

1. Rajkumar SV, Harousseau JL, Durie B, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. Blood. 2011;117(18):4691-5.
2. Durie BG, Harousseau JL, Miguel JS, et al. International uniform response criteria for multiple myeloma. Leukemia. 2006;20(9):1467-73.
3. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70(3):659-63.
4. Chng WJ, Dispenzieri A, Chim CS, et al. IMWG consensus on risk stratification in multiple myeloma. Leukemia. 2014;28(2):269-77.
5. Osoba D, Rodrigues G, Myles J, et al. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol. 1998;16(1):139-44.
6. Yost KJ, Eton DT, Garcia SF, et al. Minimally important differences were estimated for six patient-reported outcomes measurement information system-cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(5):507-16.
7. Cleeland CS. Pain assessment in cancer. In Effect of Cancer on Quality of Life. Osaba D, editor. Boca Raton: CRC Press; 1991:21.
8. Atkinson TM, Mendoza TR, Sit L, et al. The brief pain inventory and its "pain at its worst in the past 24 hours" item: clinical trial endpoint considerations. Pain Med. 2010;11(3):337-46.
9. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85(5):365-76.

10. Cocks K, Cohen D, Wisloff F, et al. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. Eur J Cancer. 2007;43(11):1670-8.
11. Cella D, Yount S, Rothrock N, et al. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. Med Care. 2007;45 (5 Suppl 1):S3-S11.
12. Garcia SF, Cella D, Clauser SB, et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. J Clin Oncol. 2007;25(32):5106-12.
13. Cella D, Riley W, Stone A, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010;63(11):1179-94.
14. EQ-5D-5L and EQ-5D-VA. Available from: http://www.euroqol.org.
15. Hung HM, Wang SJ, O'Neill R. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. J Biopharm Stat. 2007;17(6):1201-10.
16. Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med. 2010;29(2):219-28.
17. Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. J Clin Oncol. 2015;33(26):2863-9.