Title: A Phase 3, Randomized, Double-Blind, Multicenter Study Comparing Oral Ixazomib Plus Lenalidomide and Dexamethasone Versus Placebo Plus Lenalidomide and Dexamethasone in Adult Patients With Relapsed and/or Refractory Multiple Myeloma

NCT Number: NCT01564537

SAP Approve Date: 16 December 2014

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

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

- Named persons or organizations associated with the study.
- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.
- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

STATISTICAL ANALYSIS PLAN

A Phase 3, Randomized, Double-Blind, Multicenter Study Comparing Oral Ixazomib Plus Lenalidomide and Dexamethasone Versus Placebo Plus Lenalidomide and Dexamethasone in Adult Patients With Relapsed and/or Refractory Multiple Myeloma

Protocol #: C16010

SAP Version:	Date of Statistical Analysis Plan:
Version 1.0	18 March 2014
Final	16 December 2014

Approval Signatures

PPD

Purpose for this amendment:

The main purpose of this amendment is to incorporate changes from Clinical Study Protocol C16010 Amendment 3, dated 08 July 2014.

- to add an additional progression-free survival (PFS) analysis, which will be considered the final PFS analysis. The original planned first interim analysis (IA [which was the final analysis (FA) for PFS]) will be conducted at a later date when approximately 262 PFS events have occurred. The added PFS analysis (the new FA) is to be conducted when approximately 365 PFS events have occurred. The alpha allocation for PFS will be based on the O'Brien-Fleming alpha spending function (the Lan-DeMets method).

Details of major changes are specified in Section 1.1, Section 5.1, and Section 4.1.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND GLOSSARY OF TERMS

Abbreviation	Term
AE	adverse event
CI	confidence interval
CL	clearance
CMH	Cochran-Mantel-Haenszel
CR	complete response
del	deletion
DLT	dose-limiting toxicity
DNA	deoxyribonucleic acid
DOR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EDC	electronic data capture
EORTC	European Organization for Research and Treatment of Cancer
EOT	End of Treatment (visit)
EQ-5D	EuroQol 5-Dimensional Health Questionnaire
EU	European Union
FA	final analysis
FDA	United States Food and Drug Administration
GCP	Good Clinical Practice
GLP	Good Laboratory Practices
HIV	human immunodeficiency virus
HU	health utilization
IA	interim analysis
IB	Investigator's Brochure
IC_{50}	concentration producing 50% inhibition
ICF	informed consent form
ICH	International Conference on Harmonisation
IDMC	independent data monitoring committee
IEC	independent ethics committee
IMiD	immunomodulatory drug
IMMPACT	Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
IMWG	International Myeloma Working Group
IRB	institutional review board
IRC	independent review committee
ITT	intent-to-treat
IV	intravenous; intravenously
IVRS	interactive voice response system
K-M	Kaplan-Meier
MedDRA	Medical Dictionary for Regulatory Activities
MID	minimally important difference

Abbreviation	Term
Millennium	Millennium Pharmaceuticals, Inc., and its affiliates
MM	multiple myeloma
MR	minor response
MRI	magnetic resonance imaging
MTD	maximum tolerated dose
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NDMM	newly diagnosed multiple myeloma
NFKB	nuclear factor kappa-light chain enhancer of activated B cells
NSAID	nonsteroidal anti-inflammatory drugs
OME	oral morphine equivalent
ORR	overall response rate
OS	overall survival
PD	progressive disease (disease progression)
PET-CT	positron emission tomography - computed tomography
PFS	progression-free survival
PI	package insert
PK	pharmacokinetic(s)
PN	peripheral neuropathy
PO	*per os*; by mouth (orally)
PR	partial response
PRO	patient-reported outcome
PSMB1	Proteasome subunit beta type-1
QALY	quality-adjusted life year
QLQ	Quality of Life Questionnaire (EORTC)
QOL	quality of life
QTc	rate-corrected QT interval (millisec) of electrocardiograph
RP2D	recommended phase 2 dose
RRMM	relapsed and/or refractory multiple myeloma
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous
sCR	stringent complete response
SD	stable disease
SMA	safety management attachment
SQ	subcutaneous
t	translocation
$t_{1/2}$	half-life
TEAE	treatment-emergent adverse event
TRAF	Tumor necrosis factor receptor-associated factor
TTP	time to progression
UK	United Kingdom

Abbreviation	Term
US	United States
VGPR	very good partial response
WBC	white blood cell
WHO	World Health Organization

1. INTRODUCTION

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

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

1.1 Study Design

This is a phase 3, randomized, double-blind, multicenter study to evaluate the safety and efficacy of ixazomib versus placebo in patients with relapsed and/or refractory multiple myeloma (MM) who are treated with lenalidomide and dexamethasone (LenDex) as their standard therapy. The patient population will consist of adult men and women who have a confirmed diagnosis of MM, who have received 1 to 3 prior lines of therapy, and who meet other outlined eligibility criteria.

Following the Screening period, patients who will be enrolled and treated with lenalidomide plus dexamethasone will be randomized to receive either ixazomib or placebo in a double-blind fashion. Eligible patients will be randomized in a 1:1 ratio into those 2 treatment arms, stratified by: 1 versus 2 or 3 prior therapies, proteasome-inhibitor (PI) exposed versus PI naïve, and International Staging System (ISS) stage at screening of 1 or 2 versus stage 3.

Patients will receive ixazomib 4.0 mg or matching placebo capsule on Days 1, 8, and 15 plus lenalidomide (25 mg) on Days 1 through 21 and dexamethasone (40 mg) on Days 1, 8, 15, and 22 of a 28-day cycle. Patients may continue to receive treatment until disease progression (PD) or unacceptable toxicity, whichever comes first.

Patients will be assessed for disease response and progression by an independent review committee (IRC). Response will be assessed according to the International Myeloma Working Group (IMWG) criteria. All patients will be followed for survival after PD. In survival follow-up, patients will be contacted every 12 weeks until death or termination of the study by the sponsor.

1.2 Study Objectives

The primary objective is:

- To determine whether the addition of oral ixazomib to the background therapy of lenalidomide and dexamethasone improves progression-free survival (PFS) in patients with relapsed and/or refractory multiple myeloma (RRMM)

The key secondary objectives are:

- To determine whether the addition of oral ixazomib to lenalidomide and dexamethasone improves overall survival (OS)
- To determine whether the addition of oral ixazomib to lenalidomide and dexamethasone improves the OS in high-risk patients carrying deletion del(17)

Other secondary objectives are:

- To determine overall response rate (ORR), including partial response (PR), very good partial response (VGPR), and complete response (CR)
- To determine CR + VGPR
- To determine duration of response (DOR)
- To determine time to progression (TTP)
- To determine the safety of the addition of ixazomib to lenalidomide and dexamethasone
- To determine pain response rate, as assessed by the BFI-SF and analgesic use
- To assess change in global health status, functioning, and symptoms as measured by the patient-reported outcome (PRO) instrument European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) and MY20 module
- To determine the PFS and OS in high-risk cytogenetic patient groups such as translocations t(4;14), t(14;16), +1q, del(13), or del(17)

- To evaluate the potential relationship between response or resistance to ixazomib treatment and proteasome and NFKB-related genes, such as PSMB1 and TRAF-3 in blood samples
- To collect pharmacokinetic (PK) data to contribute to population PK analyses

2. POPULATIONS FOR ANALYSIS

2.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population is defined as all patients who are randomized. Patients will be analyzed according to the treatment they were randomized to receive, regardless of any errors of dosing.

The ITT population will be used for the primary and secondary efficacy analyses, and resource utilization and patient-reported outcome analysis.

2.2 Safety Population

The Safety population is defined as all patients who receive at least 1 dose of any study drug. Patients will be analyzed according to the treatment actually received. That is, those patients who are randomized to Arm B but received the regimen in Arm A will be included in Arm A; those patients who are randomized to Arm A but received the regimen in Arm B will be included in Arm B for safety analyses. More specifically, patients who received any dose of ixazomib will be included in the ixazomib + LenDex arm and patients who did not receive any dose of ixazomib will be included in the placebo plus LenDex arm, regardless of their randomized treatment.

The Safety population will be used for all safety-related analyses such as AEs, concomitant medications, laboratory tests, and vital signs.

2.3 Response-Evaluable Population

The Response-Evaluable population consists of patients who receive at least 1 dose of study drug, have measurable disease at baseline, and at least 1 postbaseline response assessment. The Response-Evaluable population will be used for the analyses of response rates, time to response, and DOR (More specifically, TTR and DOR is defined in patients with confirmed response and will be summarized descriptively). Patients must have measurable disease defined by at least 1 of the following 3 measurements:

- Serum M-protein ≥ 1 g/dL (≥ 10 g/L)
- Urine M-protein ≥ 200 mg/24 hours
- Serum free light chain assay: involved free light chain level ≥ 10 mg/dL (≥ 100 mg/L) provided the serum free light chain ratio is abnormal

2.4 Per-Protocol Population

The Per-Protocol (PP) population consists of all ITT patients who do not violate the terms of the protocol in a way that would affect the study outcome significantly as determined by the medical monitor, who is blinded to study drug assignment. The PP population includes patients who

- Do not have major inclusion/exclusion violation (inclusion criteria #3, #4; exclusion criteria #1, #8), and
- Do not have low drug compliance (compliance rate ≤ 70%) or overdose (compliance rate ≥120%) or dispensing error, and
- Have not taken excluded concomitant medication (ie, systemic treatment with strong inhibitors of CYP1A2 [fluvoxamine, enoxacin, ciprofloxacin], strong inhibitors of CYP3A [clarithromycin, telithromycin, itraconazole, voriconazole, ketoconazole, nefazodone, posaconazole] or strong CYP3A inducers [rifampin, rifapentine, rifabutin, carbamazepine, phenytoin, phenobarbital], or use of Ginkgo biloba or St. John's wort).

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

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

3. HYPOTHESES AND DECISION RULES

3.1 Statistical Hypotheses

There is 1 primary endpoint, PFS, in this study (see Section 5.7.2 for study treatment arms).

The null and alternative hypotheses for PFS are:

H_0: PFS in Arm Ixazomib + LenDex = PFS in Arm LenDex

H_a: PFS in Arm Ixazomib + LenDex > PFS in Arm LenDex

There are 2 key secondary efficacy endpoints in this study, OS and OS in high-risk patients carrying deletion del (17).

The null and alternative hypotheses for OS are:

H_0: OS in Arm Ixazomib + LenDex = OS in Arm LenDex

H_a: OS in Arm Ixazomib + LenDex > OS in Arm LenDex

The null and alternative hypotheses for OS in high-risk patients carrying deletion del(17) are:

H_0: OS in high-risk patients carrying deletion del(17) in Arm Ixazomib + LenDex = OS in high-risk patients carrying deletion del(17) in Arm LenDex

H_a: OS in high-risk patients carrying deletion del(17) in Arm Ixazomib + LenDex > OS in high-risk patients carrying deletion del(17) in Arm LenDex

3.2 Statistical Decision Rules

A closed sequential testing procedure will be used to control type I error rate at a 2-sided alpha level of 0.05 for the primary endpoint (PFS) and key secondary endpoints (OS, OS in high-risk patients carrying del(17)). That is, only if PFS is significant based on the O'Brien-Fleming alpha spending function (the Lan-DeMets method) at the first interim analysis (IA) or the second IA (whichever occur first), the test of OS will be conducted on its own alpha spending functions.

The first IA will be performed when approximately 262 PFS events have occurred based on IRC assessment. This will be the first analysis for PFS for statistical testing purposes. Based on the O'Brien-Fleming stopping boundary, the alpha level at the first IA and second IA on PFS would be 0.0163 and 0.0337, respectively, if the number of PFS events at the first IA is exactly 262.[1] Correspondingly, if the observed p value is less than 0.0163 and 0.0451 at the first IA and the second IA, respectively, the test for PFS will be claimed to be statistically significant. If the test for PFS is statistically significant at the first IA, a non-inferential analysis of PFS will be performed at the second IA where the PFS data is considered mature.

If the test for PFS is not statistically significant at the first IA, the study will continue and PFS testing will be conducted at the second IA when approximately 365 PFS events have occurred, which will also be the final analysis of PFS for statistical testing purpose. If the test for PFS is not statistically significant at second IA, the study will be claimed as unsuccessful and no further testing will be conducted.

The key secondary endpoint, OS, will be tested at the 3 planned IAs with approximately 154, 222, and 322 death events and again at the final analysis (FA) with approximately 486 events if necessary. The O'Brien-Fleming boundary will be calculated using the Lan-DeMets method. The information fraction for the IA is equal to the number of death events observed divided by 486. If there are exactly 154 death events at the first IA, the null hypothesis for OS will be rejected if the observed p value of the stratified log-rank test is less than 0.0001. If there are exactly 222 death events at the second IA, the null hypothesis for OS will be rejected if the observed p value of the stratified log-rank test is less than 0.0018. If there are exactly 322 death events at the second IA, the null hypothesis for OS will be rejected if the observed p value of the stratified log-rank test is less than 0.0112. The trial will be stopped for futility if the observed p value is greater than 0.366 at the third IA; otherwise, the OS hypothesis will be tested again at the FA. If the observed p value of the stratified log-rank test is less than 0.0462 at the FA (corresponding to a nominal alpha of 0.0382), the null hypothesis for OS will be rejected. However, the study will not be stopped after the first IA based on the test for OS.

The hypothesis for the additional key secondary endpoint, OS in high-risk patients carrying del(17), will be tested sequentially when the OS null hypothesis is rejected either at the IAs or at the FA. The testing levels of significance will be the same as those used for OS. For instance, if OS is significant at the 0.0382 level at FA, OS in high-risk patients carrying del(17) will also be tested at 2-sided alpha level of 0.0382 at the FA.

The rejection boundary for OS in high-risk patients carrying del(17) at the FA will be calculated on the basis of the actual correlation between the test statistics at the second IA and the FA. Examples of critical values at the FA based on different correlations between the test statistics at the IA and the FA using a 2-sided alpha of 0.0379 are presented below. The formula to calculate the critical values is also shown in the proof of strong control of error rate in Section 10.1.

Information fraction (n_1/n_2)	Correlation ($\sqrt{n_1/n_2}$)	Critical value
1/2	0.71	2.02
2/3	0.82	1.99
5/6	0.91	1.97
1	1	1.96

Where n_1 and n_2 are the number of death events at the second IA and FA, respectively.

Because the testing of the key secondary endpoints is gated by the significance of the primary endpoint, and the type I error rate for all key secondary endpoints is controlled at a

2-sided alpha level, the overall type I error rate for all endpoints, ie, 1 primary and 2 key secondary endpoints, is also controlled strongly at the 2-sided 0.05 alpha level.

4. INTERIM ANALYSIS

4.1 Interim Analysis

There are 3 planned IAs. The first IA will be performed when approximately 262 of the disease progression/death events have occurred. This IA is expected to occur approximately 24.5 months after the first patient is enrolled. This is the first analysis for PFS for statistical testing purpose and first IA for OS. If the test for PFS is not statistically significant at a 2 sided alpha level of 0.0163, the study will continue and PFS testing will be conducted again at the second IA when approximately 365 PFS events have occurred, which will also be the final analysis of PFS for statistical testing purposes and the second IA for OS. Based on the projected number of PFS events, the statistical significance of PFS could be claimed if the observed p value is less than 0.0451 (corresponding to nominal alpha of 0.0337). The third IA will be performed when approximately 322 deaths (two thirds of the total expected deaths) have been observed. The third IA is expected to occur approximately 44 months after the first patient is enrolled.

The test significance for the interim and final analyses of OS will be determined using O'Brien-Fleming boundaries (the Lan-DeMets method). The alpha level of OS at the first IA will be 0.00014 if there are 154 death events. However, the study will not be stopped after the first IA based on the test for OS. Based on the projected number of death events, the trial will be stopped for overwhelming efficacy if the observed p value is less than 0.0018 and 0.0112 at the second and third IA, respectively. Also the trial will be stopped for futility if the observed p value is greater than 0.366 at the third IA. The final analysis will be tested at the 2-sided alpha level of 0.0462 (corresponding to nominal alpha of 0.0382).

4.2 Independent Data Monitoring Committee (IDMC)

An independent data monitoring committee (IDMC) will periodically review safety and efficacy data at regularly scheduled meetings prespecified in the IDMC charter. All activities of the IDMC (eg, type of data reviewed, frequency of meetings, location of meetings) have been summarized in the charter. At the time of the IAs, the IDMC will be responsible for providing a recommendation regarding study continuation based on the safety and efficacy parameters.

4.3 Independent Review Committee (IRC)

An independent review committee (IRC) will review all disease evaluation data from the study and determine disease status (response and progression).

5. STATISTICAL METHODOLOGY

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

5.1 Sample Size Justification

The primary endpoint, PFS, along with the key secondary endpoints of OS and OS in high-risk patients carrying del(17), will be sequentially tested in this study.

The total sample size was calculated to provide 80% power for the OS endpoint. The study is also adequately powered to test PFS. There are 3 planned IAs and 1 FA.

The parameters used for sample size calculation on OS are a 2-sided test at the significance level of $\alpha = 0.05$, power of 80%, a control arm median OS of 30 months, and testing arm median OS of 39 months (assuming exponential distribution, hazard ratio 0.77). A total of approximately 703 patients will need to be randomized in a 1:1 ratio into those 2 arms, assuming an average enrollment rate of approximately 13 patients/month for the first 6 months, approximately 45 patients/month thereafter, and an approximately 10% dropout rate at month 35. The FA of OS is estimated to occur approximately 80 months from the enrollment of the first patient, including a 20-month randomization period and an additional 60-month follow-up from the last patient enrolled. With an observed hazard ratio of 0.833 (eg, median OS of 30 months for control vs 36 months for treatment, 20% improvement), statistical significance can be claimed at the FA with 486 death events.

Assuming a hazard ratio of 0.728 (median PFS of 15 months in the control arm versus 20.6 months in the treatment arm), 365 PFS events will be needed (85% power and 2-sided alpha of 0.05) with 2 planned IAs and the second IA as the final PFS analysis. The first IA will be performed when approximately 262 PFS events have occurred. With an assumption of approximately 13 patients/month for the first 6 months, 45 patients/month thereafter, and approximately 10% dropout rate at month 30, 262 PFS events is expected to occur at

approximately 24.5 months after the first patient is enrolled and approximately 703 patients have been enrolled. This is the first IA of the study with the opportunity to claim PFS benefit. With an observed hazard ratio of 0.743 (eg, median PFS of 15 months in the control arm versus 20.2 months in the treatment arm), statistical significance can be claimed at the analysis for PFS with 262 progression/death events. If the test for the PFS at the first IA is statistically significant, the PFS test at the second IA will be non-inferential .

If the test for PFS is not statistically significant at the first IA, the study will continue and PFS testing will be conducted at the second IA when approximately 365 PFS events have occurred, which will also be the final analysis of PFS for statistical testing purposes. An observed hazard ratio of 0.810 (median PFS of 15 months in thr control arm versus 18.5 months in the treatment arm) or better will lead to statistical significance for the PFS endpoint.

5.2 Randomization and Stratification

The randomization scheme will be generated by an independent statistician at Millennium who is not on the study team. Before dosing, a randomization number will be assigned to each patient. The randomization assignment will be implemented by an interactive voice response system (IVRS).

Eligible patients will be randomized in a 1:1 ratio into those 2 treatment arms, stratified by 1) 1 versus 2 or 3 prior therapies; 2) proteasome inhibitor exposed versus proteasome inhibitor naïve; and 3) ISS stage at screening of 1 or 2 versus stage 3. Randomization error may occur during study conduct; if this happens in the study, original stratification factors will be used in the primary analyses.

5.3 Blinding and Unblinding

This is a double-blind study; all study personnel, including the investigators, site personnel, study clinicians, and the sponsor, will be blinded to the treatment assignments for the duration of the study. Only the independent statistical center (ISC) and IDMC will have access to unblinded individual patient data in the electronic data capture system. The periodic safety analyses will be generated for the IDMC by an ISC. Three formal efficacy IAs will also be conducted by ISC for the IDMC.

Refer to Section 4.2 for the roles and responsibilities of IDMC.

5.4 Data Handling

5.4.1 Methods for Handling Missing Data

All available efficacy and safety data will be included in data listings and tabulations. Data that are potentially spurious or erroneous will be examined under the auspices of standard data management operating procedures.

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

For PRO data, primarily missing data imputation will be based on published instrument specific methods.

5.4.1.1 Missing/Partial Dates in Screening Visit

The following rules apply to dates recorded in the Screening visits:

- If only the day component is missing, the first day of the month will be used if the year and the month are the same as those for the first dose of study drug; otherwise, the 15th will be used.

- If only a year is present, and it is the same as the year of the first dose of study drug, the 15th of January will be used unless it is later than the first dose, in which case the date of the first of January will be used, unless other data indicate that the date is earlier.

- If only a year is present, and it is not the same as the year of the first dose of study drug, the 15th of June will be used, unless other data indicate that the date is earlier.

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

Every effort will be made to avoid missing/partial dates in on-study data.

Adverse events with stop dates that are completely or partially missing will be imputed as follows:

- If the stop date has a month and year but the day is missing, the last day of the month will be imputed.

- If the stop date has a year but the day and month are missing, the 31st of December will be imputed.

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

Adverse events with start dates that are completely or partially missing will be imputed as follows:

- If the start date has a month and year but the day is missing, the first day of the month will be imputed.
    - If this date is earlier than the first dose date, then the first dose date will be used instead.
    - If this date is later than the stop date (possibly imputed), then the stop date will be used instead.
- If the start date has a year, but the day and month are missing, the 15th of June will be imputed.
    - If this date is earlier than the first dose date, then the first dose date will be used instead.
    - If this date is later than the stop date (possibly imputed), then the stop date will be used instead.

If the start date of an event is completely missing, then it is imputed with the first dose date.

Concomitant therapies with start dates that are completely or partially missing will be analyzed as follows:

- If the start date has a month and year but the day is missing, the therapy will be included in the summary table if the month and year of the start date of the event are:
    - On or after the month and year of the date of the first dose of study drug

      and

  - On or before the month and year of the date of the last dose of study drug plus 30 days

- If the start date has the year but the day and month are missing, the therapy will be included in the summary table if the year of the start date of the event is:

  - On or after the year of the date of the first dose of study drug

    and

  - On or before the year of the date of the last dose of study drug plus 30 days

If the start date of an event is completely missing, then the therapy will be included in the summary table.

Subsequent therapies with start dates that are completely or partially missing will be analyzed as follows:

- When the month and year are present but the day is missing:

  - If the onset month and year are the same as the month and year of the last dose of study drug, the day of the last dose + 1 will be imputed.

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

- When only a year is present:

  - If the onset year is the same as the year of the last dose of study drug, the date of last dose + 1 will be imputed.

  - If the onset year is not the same as the year of the last dose of study drug, the first day of the year is imputed.

- If no components of the onset date are present, the date of the last dose of study drug + 1 will be imputed.

5.4.2 Definition of Baseline Values

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

5.4.3 Windowing of Visits

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

5.4.4 Justification of Pooling

All data from all sites will be pooled. Study center or treatment-by-center interaction will not be included in any statistical analysis.

5.4.5 Withdrawals, Dropouts, Loss to Follow-up

Time-to-event parameters will be censored if patients withdraw, drop out, or are lost to follow-up before documentation of the events (PD/death). Rules for censoring are detailed in Section 5.8.

5.5 Patient Disposition

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

A listing will present data concerning patient disposition.

5.6 Demographics and Baseline Disease Characteristics

5.6.1 Demographics

Demographics will be summarized by treatment group in a descriptive fashion for the ITT population. Baseline demographic data to be evaluated will include age, sex, race, ethnicity, height, weight, and other parameters as appropriate. Patient enrollment by region and country will also be summarized by treatment groups.

5.6.2 Medical History

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

5.6.3 Baseline Disease Status

Baseline disease characteristics including Eastern Cooperative Oncology Group (ECOG) performance status; type of myeloma; ISS stage; serum M-protein; urine M-protein; serum FLC; β_2-microglobulin by category (ie, < 3.5, 3.5-5.5, ≥ 5.5 mg/L); serum creatinine and its category (≤ 2, > 2 mg/dL); creatinine clearance by category (ie, 30-60, ≥ 60 mL/min); serum albumin by category (ie, < 3.5, ≥ 3.5 g/dL); corrected calcium; hemoglobin; Durle-Salmon stage; Lytic bone lesions; and extramedullary disease will be summarized for all patients.

A patient's type of myeloma is determined by the combination of heavy chain type (IgG, IgA, IgM, IgD, and other) and light chain type (kappa, lambda, and biclonal). In descriptive summaries, the disease will be categorized by the heavy chain type first, then within each of these categories, patients will be further classified according to their light chain type.

Patients are to be classified into 3 stages on the basis of the ISS. The derivation is based on the baseline β_2-microglobulin and baseline albumin level measurements.

Stage	Criteria
Stage I	Serum β_2-microglobulin < 3.5 mg/L Serum albumin ≥ 3.5 g/dL
Stage II	Neither Stage I nor Stage III
Stage III	Serum β_2-microglobulin ≥ 5.5 mg/L

If corrected calcium is not reported directly, it can be calculated using either of the following 2 formulas according to the units of calcium and albumin:

$$\text{corrected calcium (mg/dL)} = \text{serum calcium (mg/dL)} + 0.98 \times (4 - \text{serum albumin[g/dL]})$$
$$\text{corrected calcium (mmol/L)} = \text{serum calcium (mmol/L)} + 0.0246 \times (40 - \text{serum albumin[g/L]})$$

Creatinine clearance is to be calculated using the Cockcroft-Gault formulas as follows:

For male patients:

$$\text{creatinine clearance} = \frac{(140 - \text{Age[yrs]}) \times \text{weight[kg]}}{72 \times (\text{serum creatinine[mg/dL]})}$$

For female patients:

$$\text{creatinine clearance} = 0.85 \times \frac{(140 - \text{Age[yrs]}) \times \text{weight[kg]}}{72 \times (\text{serum creatinine[mg/dL]})}$$

Integer values will be used.

A separate table will summarize the numbers and percentages of patients who had prior systemic therapy, prior transplant, prior surgery, and prior radiation therapy. This table will also include regimens and the number of lines of prior systemic therapies, best hematologic responses to prior systemic therapy, relapsed and/or refractory status to prior systemic therapy, and months since progression from last prior systemic therapy and months since diagnosis for each treatment group in the ITT population. A by-patient listing will also be presented for prior systemic therapy.

Months from diagnosis to the randomization date for each treatment is calculated by:

$$\frac{\text{randomization date - date of diagnosis}}{365.25/12}$$

Distribution of stratification factors will also be summarized.

5.7 Treatments and Medications

5.7.1 Concomitant Medications

Concomitant medications will be coded by Preferred Term using the World Health Organization (WHO) Drug Dictionary. The number and percentage of patients taking concomitant medications from the first dose through the end of the on-treatment period will be tabulated by Anatomical Therapeutic Chemical (ATC) classification pharmacological subgroup and WHO drug generic term for each treatment group in the Safety population. A by-patient listing will also be presented for concomitant medications.

Concomitant procedures will not be coded but will be presented in a data listing for the Safety population.

5.7.2 Study Treatments

Following the Screening period, patients who will be enrolled and treated with lenalidomide plus dexamethasone will be randomized to receive a study drug in a double-blind fashion, either ixazomib or placebo. Eligible patients will be randomized in a 1:1 ratio into those 2 treatment arms.

Arm Ixazomib + LenDex: Patients will receive ixazomib (4.0 mg) orally (PO) on Days 1, 8, and 15 plus lenalidomide (25 mg) on Days 1 through 21 and dexamethasone (40 mg) on

Days 1, 8, 15, and 22 of a 28-day cycle. Patients may continue to receive treatment until PD or unacceptable toxicity, whichever comes first. Dose modifications may be made on the basis of toxicities. Patients with creatinine clearance ≤ 60 mL/min (or ≤ 50 mL/min based on regional label) will receive a reduced lenalidomide dose of 10 mg once daily on Days 1 through 21 of a 28-day cycle. The lenalidomide dose may be escalated to 15 mg once daily after 2 cycles if the patient is not responding to treatment and is tolerating the treatment.

Arm Placebo + LenDex: Patients will receive a placebo capsule on Days 1, 8, and 15 plus lenalidomide (25 mg) on Days 1 through 21 and dexamethasone (40 mg) on Days 1, 8, 15, and 22 of a 28-day cycle. Patients may continue to receive treatment until PD or unacceptable toxicity, whichever comes first. Dose modifications may be made on the basis of toxicities. Patients with creatinine clearance ≤ 60 mL/min (or ≤ 50 mL/min based on regional label) will receive a reduced lenalidomide dose of 10 mg once daily on Days 1 through 21 of a 28-day cycle. The lenalidomide dose may be escalated to 15 mg once daily after 2 cycles if the patient is not responding to treatment and is tolerating the treatment.

5.7.2.1 Duration of Follow-up

The duration of follow-up is defined as time from randomization to the date of death or last known visit. If a subject dies, the duration will equal the date of death minus the date of study start + 1, with a censor variable = 1 (censored for follow-up). If a subject is alive, the duration will equal the date when the subject was last known to be alive minus the date of study start + 1, with a censor variable = 0 (event for follow-up).

5.7.2.2 Extent of Exposure

An overall summary of drug exposure will be presented, including number of treated cycles, and numbers and percentages of patients who had ≥ 1, ≥ 2, …, and ≥ 18 treated cycles, for each treatment group in the Safety population. An aggregate summary of numbers and percentages of patients who had 1 to 6, 7 to 12, 13 to 18, and ≥ 19 treated cycles will also be presented in the same table.

Additionally, exposure to dexamethasone will be characterized by total amount of dose taken in mg, total number of dose taken, number of treated cycles, numbers and percentages of patients who had ≥ 1, ≥ 2, …, and ≥ 18 treated cycles, and relative dose intensity (%) for each treatment group in the Safety population. An aggregate summary of numbers and percentages of patients who had 1 to 6, 7 to 12, 13 to 18, and ≥ 19 treated cycles will also be presented in the same table.

Ixazomib/placebo and lenalidomide exposure will be summarized similarly for the applicable treatment group/option.

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

A treated cycle for a specific drug is defined as a cycle in which the patient received any amount of the specific drug.

$$\text{relative dose intensity (\%)} = 100 \times \frac{\text{total dose received}}{\text{sum of prescrbed dose over treated cycles}}$$

The number of patients with 100% relative dose intensity, 80% to < 100%, 50% to < 80%, and < 50% will be summarized by treatment group and for all patients.

Percent compliance is calculated as 100% * (study drug taken in mg) / (study drug expected to be taken in mg). Only reduced prescribed or increased prescribed or dose hold or discontinued permanently will change the expected prescribed dose.

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

5.7.2.3 Treatment Modifications

Action on each study drug will be summarized by each of Cycles 1 through 18, sum of the remainder cycles, Cycles 1 through 6, Cycles 7 through 12, 13 through 18, and ≥ 19 cycles and total for each treatment group in the Safety population.

5.8 Efficacy Analyses

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

5.8.1 Primary Efficacy Endpoint

There is 1 primary endpoint, PFS, defined as the time from the date of randomization to the date of first documentation of PD based on central laboratory results and IMWG criteria as evaluated by an IRC, or death due to any cause, whichever occurs first. PD confirmation is required by 2 consecutive evaluations (at least 1 week apart) as per IMWG criteria. Confirmation is not required if PD occurs at the date of the last response assessment (special cases, e.g. PR->PD->NE, will be considered as confirmed PD if NE occurs at the last visit). Patients without documentation of PD will be censored at the date of the last response

assessment that is stable disease (SD) or better. The details regarding the handling of missing assessments and censoring for the PFS analysis are presented Table 5-1.

Table 5-1 Handling of Missing Assessments and Censoring for PFS Primary Analysis Based on FDA guidance

Situation	Date of Progression or Censoring	Outcome
No baseline and/or no post baseline assessment, no subsequent anticancer therapy after study treatment, no death	Date of Randomization	Censored
Disease progression documented between scheduled visits	Date of next scheduled visit	Progressed
No documented death or disease progression	Date of last adequate assessment[a]	Censored
Lost to follow-up, withdraw consent before any documented death or disease progression	Date of last adequate assessment[a]	Censored
Death or progression after more than 1 missed visit[b]	Date of last adequate assessment[a]	Censored
Alternate antineoplastic therapy started prior to disease progression	Date of last adequate assessment prior to starting alternate antineoplastic therapy	Censored
Death before first assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed

a Adequate disease assessment is defined as there is sufficient data to evaluate a patient's disease status.
b Death or progression occur more than 90 days from previous adequate assessment.

5.8.1.1 Primary Efficacy Analysis

PFS will be analyzed when approximately 262 and 365 PFS events have occurred for the first IA and for the second IA, respectively. A 2-sided, stratified log-rank test will be used to compare the treatment groups with respect to PFS at a 2-sided alpha level of 0.0163 and 0.0337, which corresponds to a nominal p value of 0.0451 based on the O'Brien-Fleming stopping boundary (the Lan-DeMets method). In addition, an unadjusted stratified Cox model will be used to estimate the hazard ratio and its 95% CIs for the treatment effect using the stratification factors. The Kaplan Meier (K-M) survival curves and K-M medians (if estimable), along with their 2-sided 95% CIs, will also be provided for each treatment group.

Sensitivity analyses for PFS include:

1. PFS assessed by investigator will be analyzed in the ITT population.

2. PFS assessed by IRC will be analyzed in the PP population.

Progression-free survival assessed by IRC using different censoring mechanisms will be analyzed in the ITT population, for example, not censoring for patients who discontinue treatment and go on transplant or alternative neoplastic therapy. The details of the handling of missing assessments and censoring for additional sensitivity analyses are presented in Table 5-2.

Table 5-2 Handling of Missing Assessments and Censoring for PFS Sensitivity Analysis Based on EMA Guidance

Situation	Date of Progression or Censoring	Outcome
Disease progression documented between scheduled visits	Date of documented disease progression	Progressed
Alternate antineoplastic therapy started prior to disease progression	Date of documented disease progression	Progressed
Death or disease progression after more than one missed visit	Date of death or disease progression	Progressed

In addition, a stepwise Cox model will be implemented to identify potential predictive factors using relevant demographic or diagnostic covariates, with the entry level fixed at 0.25 and a stay level fixed at 0.10. Besides treatment and the stratification factors, the model may include the following prognostic factors including, but not limited to, age; race (white, non-white); prior therapy (immunomodulatory drug [IMiD]-exposed vs IMiD-naïve); baseline ECOG score; cytogenetic test (high-risk group [del(17); t(4;14) and t(14;16)] vs standard); and corrected serum calcium.

Subgroup analyses will be performed for PFS relative to baseline stratification factors, demographic data such as sex, race, and age, and disease characteristics such as type of prior regimen. The details on subgroups are presented in the following:

Subgroup	Definition of Group
Age	≤ 65 years, >65 and ≤ 75 years, > 75 years
Sex	male vs female
Race	white, black-African American, Asian, other
Region	North America, Europe, Asian, other
Western Countries	North America + Western Europe
Cytogenetic risk	Standard-risk, high-risk [(del17); t(4;14); t(14;16)], not available
ISS stage	In additional to stratification factors , also define as 1 or 2 or 3
Lines of prior therapies	In additional to stratification factors , also define as 1 or 2 or 3
Prior proteasome inhibitor therapy	exposed vs naïve
Prior IMiD therapy	exposed vs naïve
Thalidomide Refractory	yes vs no
Refractory to any line of prior therapy	yes vs no
Refractory to last line of prior therapy	yes vs no
Relapsed and/or refractory	Relapsed, efractory, Relapse and refractory
Prior Velcade therapy	Exposed vs naïve
Renal function based on baseline creatinine clearance	< 60 mL/min, and ≥60 mL/min
Liver Function based on baseline ALT or AST	≥1.5× ULN vs. <1.5 × ULN
ECOG performance status	0 or 1 vs 2

If it is determined that the study will continue after PFS is significant at a 2-sided alpha level of 0.0163 at the first IA, a non-inferential analysis of PFS will be performed at the second IA.

5.8.2 Key Secondary Efficacy Endpoints

There are 2 key secondary endpoints: OS and OS in high-risk patients carrying del(17).

Overall Survival

Overall survival is defined as the time from the date of randomization to the date of death. Patients without documentation of death at the time of the analysis will be censored at the date when they were last known to be alive. Overall survival will be analyzed using the ITT population.

Overall Survival in High-Risk Patients With del(17)

The high-risk patients with del(17) subgroup is defined as the cases reported as positive for del(17) by the central laboratory combined with those cases that lack a central laboratory

result for del(17) and were reported positive for del(17) by local laboratory. Overall survival within the high-risk patients with del(17) subgroup is defined the same as OS in the overall population.

5.8.2.1 Key Secondary Efficacy Analysis

Two key secondary efficacy endpoints will be tested sequentially in the order of 1) OS, and 2) OS in high-risk patients carrying del(17). The second key secondary endpoint will be tested only if the test for OS is statistically significant at the IAs or FA at the same significance level as OS to ensure strong type I error control for both key secondary endpoints.

Overall Survival

Overall survival will be tested only after statistical significance is achieved for PFS. A 2-sided, stratified log-rank test will be used to compare the treatment groups with respect to OS. The test significance level at the IAs and FA is decided by the O'Brien-Fleming alpha spending function (the Lan-DeMets method). In addition, an unadjusted stratified Cox model will be used to estimate the hazard ratio and its 95% CIs for the treatment effect using the stratification factors. The K-M survival curves and K-M medians (if estimable), along with their 2-sided 95% CIs, will also be provided for each treatment group.

In addition, a stratified Cox regression model will be used to further evaluate the treatment effects on OS after adjusting for some prognostic factors. Besides treatment and the stratification factors, the model may include, but will not be limited to, the following prognostic factors simultaneously: age; race (white, non-white); prior therapy (IMiD-exposed vs IMiD-naive); baseline ECOG score; cytogenetic results (high risk vs normal); and corrected serum calcium.

To adjust for the potential confounding effects of subsequent therapies after patients discontinue study treatment, the following 2 methods will be used:

- Marginal Structural Models (MSMs) by Robins and Finkelstein[(2)]
- Inverse Probability of Censoring Weighted (IPCW) method by Robins and Finkelstein[(3)]

In the MSM and IPCW analyses, in order to derive weights adjusting for the time-fixed and time-varying confounding effects due to taking alternative therapies, the covariates that

affect disease progression and post-progression treatment, and the OS endpoint will be used. Baseline covariates include region (North America, others), age (< 75, ≥ 75), race (white, non-white), ECOG score (0 or 1, 2), relapse and/or refractory relapsed, refractory, relapse and refractory), type of myeloma (IgA, other), percentage of plasma cells (≤ 30, > 30; missing), presence of extramedullary plasmacytomas (yes, no), presence of lytic bone lesions (yes, no), cytogenetic abnormalities (high risk, others), prior lenalidomide (yes, no), prior thalidomide (yes, no), prior proteasome inhibitor (yes, no), number of line therapies (1 vs 2 or 3), baseline hemoglobin, baseline platelets, creatinine clearance (median), albumin (median), LDH (median), β_2 microglobulin (median), and corrected calcium (median). Time-varying covariates include duration of exposure, disease progression status at each study visit, hemoglobin value at each study visit and progression/relapse, platelets value at each study visit and progression/relapse, M-protein value at progression/relapse, type of subsequent therapy with proteasome inhibitor, types of subsequent therapy with IMIDs. The final criteria for selected covariates would need to be statistically have a p-value of less than or equal to 0.1 in the multivariate Cox regression models for weight calculations, and in the final Cox model for OS. If there are more than 5% missing in the covariate, then this covariate will be dropped from the weighting calculation and final OS model. For both MSM and IPCW analyses, logistic regression models on repeated measurements will be used to approximate the Cox models in the weight derivations[4] from which stabilized weights will be derived per subject per observation. Adjusted K-M curves[5] will also be presented along with hazard ratios (HRs), 95% confidence intervals for HRs, and adjusted p-values based on MSM and IPCW approaches.SAS proc PHREG procedure with counting process type of data input, which takes multiple observations per subject, will be used as the final Cox model for OS for both MSM and IPCW approaches.

Furthermore, a versatile test for the equality of the 2 survival functions based on weighted differences of K-M curves proposed by Uno, Tian, Claggett, and Wei[6] will be used to test the treatment differences as a sensitivity analysis.

Subgroup analyses will be performed for OS relative to baseline stratification factors and demographic data such as sex, race, age, and disease characteristics, such as type of prior regimen.

Additional exploratory analyses may be performed if deemed necessary.

Overall Survival in High-Risk Patients With del(17)

Overall survival within the high-risk patients with del(17) subgroup will be analyzed using a similar method as the OS in the overall population.

5.8.3 Other Secondary Efficacy Endpoints and Analyses

Other secondary efficacy parameters include overall response rate (ORR); the combined CR + VGPR rate; time to progression; DOR; time to subsequent antineoplastic therapy; OS and PFS in high-risk subgroups harboring del(17), translocation t(4;14), or t(14;16); and pain response rate.

The high-risk patients are defined as the cases reported as positive by the central laboratory combined with those cases that lack a central laboratory result and were reported positive by a local laboratory.

Disease response-related endpoints will be analyzed using the IRC-assessed response rate.

Overall Response Rate

The ORR is defined as the proportion of patients who achieved PR or better relative to the ITT population. The ORR is determined from the start of the study treatment until the end of treatment, taking into account any requirement for confirmation. To calculate the confirmed response, 1 visit after end of treatment (EOT) is used to confirm the assessment at EOT. A stratified Cochran-Mantel-Haenszel (CMH) test will be used to compare ORR between the 2 treatment arms. A logistic regression model will be used to estimate the treatment effect in terms of odds ratio. The odds ratio and its associated 95% CIs will be presented.

The IRC-assessed response data will be used for the analysis. Investigator-assessed response data will be used for the sensitivity analysis.

Combined Complete Response + Very Good Partial Response Rate

The combined CR + VGPR rate will be analyzed using the same method as ORR.

Time to Progression

Time to progression is defined as the time from the date of randomization to the date of first documentation of PD. Patients without documentation of PD at the time of the analysis will be censored at the date of their last response assessment that is SD or better. Patients with no response assessment will be censored at the day of randomization. Patients who do not

experience progression and start new systemic therapy for multiple myeloma will be censored at the date of their last response assessment that is SD or better. Time to progression will be analyzed using the ITT population using a similar method as PFS.

Duration of Response

Duration of response is defined as the time from the date of first documentation of a PR or better to the date of first documentation of PD for responders. Responders without documentation of PD will be censored at the date of their last response assessment that is SD or better. Duration of response will be summarized descriptively using the K-M method. Kaplan Meier (K-M) survival curves and K-M medians (if estimable) will be provided for each treatment group.

OS and PFS in High Risk Cytogenetic Patient Groups

Overall survival, PFS, and ORR in the high-risk subgroups will be analyzed using a similar method as those in the ITT population. The following high-risk populations will be analyzed:

- Individual abnormality group: patients carrying 1 of the following cytogenetic abnormalities: del(17), translocation t(4;14), or t(14;16)
    - the del17 will include pts with del17 alone along with pts where the del17 is associated to t(4;14) or t(14;16)
    - the t(4;14) group will include ONLY pts with t(4;14) ALONE
    - the t(14;16) group will include ONLY pts with t(14;16) ALONE
- Cytogenetic high-risk group: patients carrying any of the following cytogenetic abnormalities: del(17), translocation t(4;14), or t(14;16)

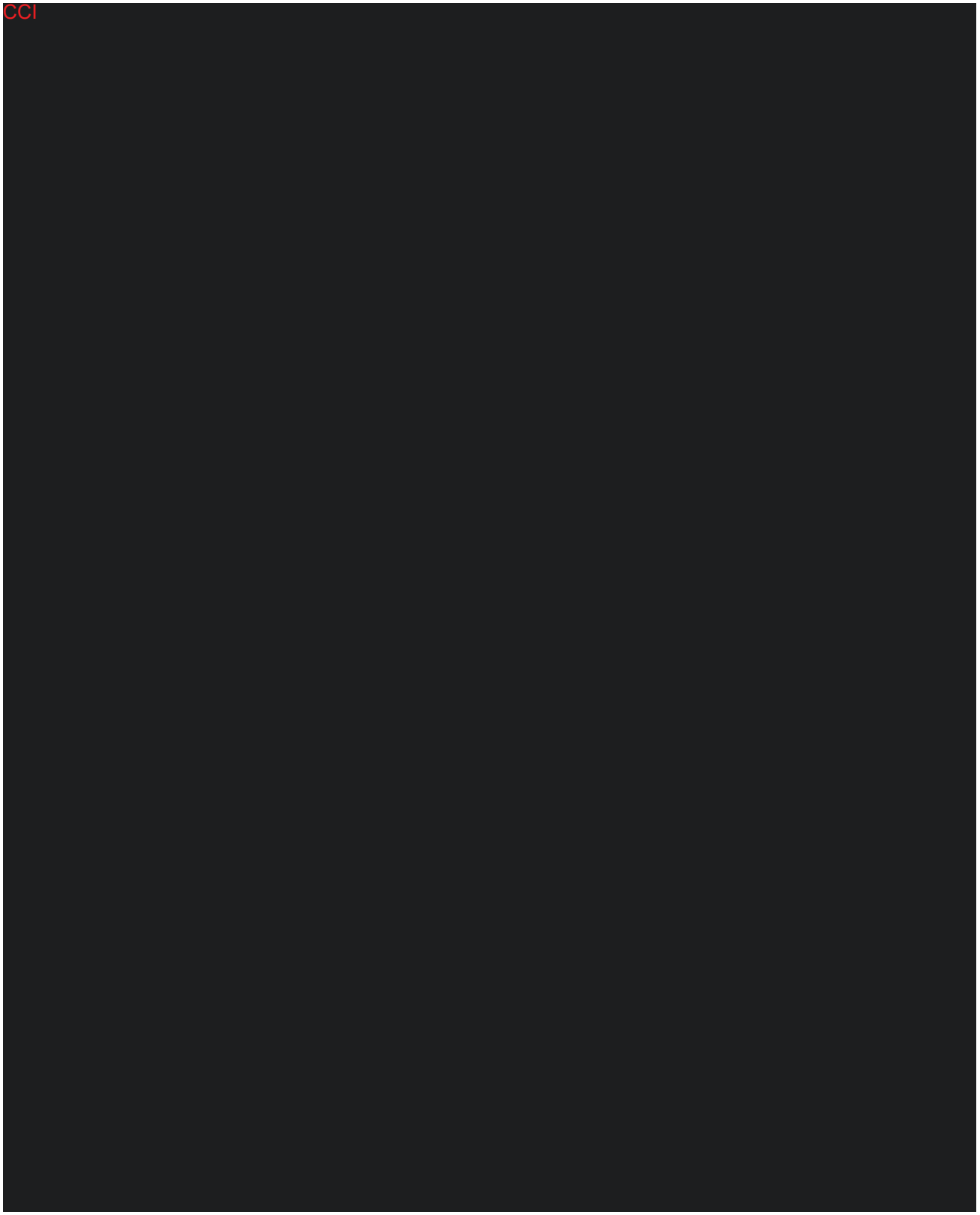
5.9 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses

5.9.1 Pharmacokinetic Analyses

Plasma concentration-time data will be presented in listings. Pharmacokinetic data will be used to perform population PK analysis using a nonlinear mixed effects modeling approach and to assess the effect of various covariates on PK after including data from other studies, if possible. The analysis plan for the population PK analysis will be separately defined, and the results of these analyses will be reported separately.

5.9.2 Pharmacodynamic Analyses

Not applicable.



5.10 Resource Utilization and Patient-Reported Outcome Analysis

5.10.1 Patient-Reported Outcomes (PROs)

Patient-reported outcome assessments using the EORTC QLQ-C30 and MY20 will be analyzed using the ITT population. The descriptive statistics of actual values and changes from baseline of the subscale scores for the EORTC QLQ-C30 and MY20 will be summarized by treatment group over time. Additionally, the descriptive statistics of actual values and changes from baseline of global health status/quality of life (QOL) will be summarized by treatment group over time for responders and then nonresponders. The subscales of the EORTC QLQ-C30 and QLQ-MY20 are defined in Table 5-3 and Table 5-4, respectively.

Table 5-3 Definition of Subscale Scores of EORTC QLQ-C30

Subscale	Individual Items
Physical functioning	1-5
Role functioning	6-7
Emotional functioning	21-24
Cognitive functioning	20, 25
Social functioning	26-27
Quality of life	29-30
Fatigue	10, 12, 18
Nausea and vomiting	14-15
Pain	9, 19
Dyspnea	8
Insomnia	11
Appetite loss	13
Constipation	16
Diarrhea	17
Financial difficulties	28

Table 5-4 Definition of Subscale Scores of EORTC QLQ-MY20

Subscale	Individual Items
Future perspective	18-20
Body image	17
Disease symptoms	1-6
Side effects of treatment	7-16

Differences between treatment groups in the EORTC QLQ-C30 and MY20 subscale scores will be evaluated using published minimally important difference (MID) values of 10 points. Patients with a change from baseline score ≥ MID in a direction reflecting improved QOL/functioning or decreased symptoms at a given time point will be defined as improvement. The number and percentage of patients with an improvement from baseline in subscale scores will be summarized by treatment group over time. Specific areas of interest include physical functioning, global QOL, fatigue, nausea/vomiting, pain, dyspnea, appetite loss, and constipation/diarrhea. The main endpoint for the PRO analysis will be the global health status/QOL subscale of the EORTC QLQ-C30 and functional scales and symptom scales of the MY20. The other PRO endpoints include the remaining EORTC QLQ-C30 and MY20 subscale scores. The change from baseline in subscale scores at any postbaseline visit will be presented using cumulative distribution function (CDF) figures. The patient's largest improvement from baseline will be determined first, and CDF curves will then be used to display the cumulative percentage of patients experiencing that improvement by treatment group. Additionally, CDF curves will be generated for global health status/QOL score by treatment group among responders and nonresponders, respectively.

The change from baseline in subscale scores will be also analyzed using the repeated-measures linear mixed-effects (random-intercept only) models, including treatment group, time (a discrete variable), the interaction between treatment group and time, baseline score, and stratification factors as covariates. The repeated-measures analysis will use measurements collected from all available time points specified in the Schedule of Events in the protocol. The estimation of variance-covariance matrix and statistics such as Akaike information criteria (AIC) and Bayesian information criteria (BIC) will be included in evaluating the linear mixed-effects model. The estimated mean difference in the changes from baseline between the 2 treatments and the 95% CIs will be provided.

Details of scoring and initial handling of missing data are included in the EORTC QLQ-C30 and MY20 scoring guidelines.

Missing data patterns will be examined. For sensitivity analyses, different imputation methods for missing data including LOCF, the random slope model, and the pattern mixture model may be performed if appropriate after examining missing data patterns.

5.10.2 Health Economics (Health Care Resource Use)

EuroQol 5 Dimensional Health Questionnaire scores will be summarized in descriptive statistics for treatment arms over time.

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

Further modeling will be performed separately at post hoc analyses.

5.10.3 Pain

Pain response rate will be analyzed in patients with a baseline pain score ≥ 4 in the ITT population. Pain response rate is the proportion of patients who have a pain response and will be summarized by treatment groups. The CMH test will be used to compare the 2 treatment groups by stratification factors. In addition, the absolute treatment difference in pain response rate will be provided, along with the 95% CI. A CDF plot will be generated for pain response.

Pain response is defined as the occurrence of a 30% reduction from baseline in BPI-SF worst pain score over the last 24 hours without an increase in analgesic use at 2 consecutive evaluations.

Additional analyses on pain include:

- Time to pain response/progression, as assessed by the time from randomization to initial response/progression classification
- Duration of pain response, measured as the time from the first documented pain response to the first documented pain progression classification

Time to pain progression and duration of pain response will be compared using the stratified log-rank test between the 2 treatment groups. An unadjusted stratified Cox model will be used to estimate the hazard ratio and its 95% CIs for the treatment effect using the stratification factors. The K-M survival curves and K-M medians (if estimable), along with their 2-sided 95% CIs, will also be provided for each treatment arm. Time to pain response will be summarized descriptively among patients with pain response.

Pain progression is defined as the occurrence of 1 of the following and confirmed by 2 consecutive evaluations. To qualify as progression, the patient must have a worst BPI-SF score ≥ 4 during progression.

- A ≥ 2 point and 30% increase from baseline in BPI-SF worst pain score without an increase in analgesic use, or
- A 25% or more increase in analgesic use from baseline without a decrease in BPI-SF worst pain score from baseline

Analgesic use can be stable or increased according to the following definitions:

- Stable analgesic use is defined as less than a 25% change of the oral morphine equivalent (OME) dose from baseline
- Increased analgesic use is defined as an increase of 25% or more in OME from baseline

Confirmation is not required if surgical treatment for pain, palliative radiation for pain, or subsequent antineoplastic therapy has been received before a confirmatory assessment.

In addition, the actual value and change from baseline in BPI-SF pain scores will be summarized by treatment group over time. The change from baseline in worst pain score will be also analyzed using the repeated-measures linear mixed-effects (random-intercept only) models, including treatment group, time (a discrete variable), the interaction between treatment group and time, baseline score, prior therapies, ISS stage at screening, sex, race, and age as covariates.

Compliance rates for EORTC QLQ-C30, MY20, BPI-SF, and EQ-5D will be summarized over time by treatment group.

5.11 Safety Analyses

Safety evaluations will be based on the incidence, intensity, type of AEs, clinically significant changes in the patient's physical examination findings, electrocardiograms (ECGs), vital sign measurements, and clinical laboratory results.

These analyses will be performed using the Safety population. All analyses will be performed for each treatment arm.

5.11.1 Adverse Events

5.11.1.1 Adverse Events

Adverse events will be coded using MedDRA version 14.0 or higher. All AEs will be presented in a by-patient listing. Treatment-emergent AEs are AEs that occur after administration of the first dose of any study drug and through 30 days after the last dose of any study drug.

Adverse events will be tabulated according to the MedDRA by System Organ Class, High Level Term, and Preferred Term, and will include the following categories:

- Treatment-emergent AEs
- Drug-related treatment-emergent AEs
- Grade 3 or higher treatment-emergent AEs (also report Grade 3 and 4 separately)
- Grade 3 or higher drug-related treatment-emergent AEs (also report Grade 3 and 4 separately)
- The most commonly reported treatment-emergent AEs (ie, those events reported by ≥ 10% of patients in either treatment group)
- Serious AEs (SAEs)

Patients with the same AE more than once will have that event counted only once within each body system, once within each High Level Term, and once within each Preferred Term.

Drug-related treatment-emergent AEs will also be summarized by the National Cancer Institute Common Toxicity Criteria (NCI CTCAE) version 4.03. Patients with the same AE more than once will have the maximum intensity of that event counted within each body system, once within each High Level Term, and once within each Preferred Term.

The most commonly reported treatment-emergent AEs (ie, those events reported by ≥ 10% of any treatment arm) will be tabulated by Preferred Term. Patients with the same AE more than once will have that event counted only once within each Preferred Term.

An overall summary AE table will include numbers and percentages of patients who had any AE, drug-related AE, Grade 3 or higher AE (also Grade 3 and 4 AE), Grade 3 or higher

drug-related AE (also Grade 3 and 4 AE), SAE, drug-related SAE, AE resulting in discontinuation, and on-study deaths. On-study death is defined as the death that occurs between the first dose of any study drug and within 30 days of the last dose of any study drug.

Incidence rates will be calculated for the safety population based on AEs and SAEs. Incidence rate is defined as the number of events divided by the total number of person-years at risk (per 100) in the Safety population.

5.11.1.2 Serious Adverse Events

The number and percentage of patients experiencing at least 1 treatment-emergent SAE will be summarized by MedDRA primary System Organ Class, High Level Term, and Preferred Term. Drug-related SAEs will be summarized similarly.

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

5.11.1.3 Deaths

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

5.11.1.4 Adverse Events Resulting in Discontinuation of Study Drug

A by-patient listing of treatment-emergent AEs resulting in discontinuation of study drug will be presented.

5.11.2 Laboratory Data

For the purposes of summarization in both the tables and listings, all laboratory values will be converted to standardized units. If a lab value is reported using a non-numeric qualifier (eg, less than (<) a certain value, or greater than (>) a certain value), the given numeric value will be used in the summary statistics, ignoring the non-numeric qualifier. However, for the bone marrow plasma cell percentage, the convention of (x-1)% (mainly for < 5% for CR) will be used.

Laboratory test results from the central laboratory will be used when they are available. Laboratory test results from local laboratory will be used only when no central laboratory test result exists at the same scheduled sample collection time point.

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

Laboratory test results will be summarized according to the scheduled sample collection time point. Change from baseline will also be presented. Unscheduled laboratory test results will be listed and included in laboratory shift tables. The parameters to be analyzed are as follows:

- Hematology: hemoglobin, hematocrit, absolute neutrophil count (ANC), absolute lymphocyte count (ALC), monocytes, platelets, and white blood cell (WBC) count
- Serum chemistry: blood urea nitrogen, creatinine, total bilirubin, uric acid, lactate dehydrogenase, albumin, alkaline phosphatase, aspartate aminotransferase (AST), alanine aminotransferase (ALT), glucose, calcium, sodium, potassium, chloride, carbon dioxide, magnesium, and phosphate.

Shift tables will be constructed for laboratory parameters to tabulate changes in NCI CTCAE for toxicity (version 4.03) from baseline to worst postbaseline CTC grade. Parameters to be tabulated will include:

- Hematology: ALC, ANC, hemoglobin, platelets, and WBC
- Serum chemistry: ALT, AST, alkaline phosphatase, creatinine, total bilirubin, calcium, magnesium, potassium, sodium, and phosphate

Summary statistics will also be presented for shift from baseline urinalysis values.

Mean laboratory values and box plots over time for key lab parameters will be produced including, but not limited to, ANC, platelets, and liver function tests (ALT, AST, alkaline phosphatase, and total bilirubin).

By-patient listings to be presented include hematology, serum chemistry, urinalysis, urine total protein, and urine creatinine.

5.11.3 Electrocardiograms

Descriptive statistics for the actual values and changes from baseline in ECGs will be tabulated by time point.

Corrected QT interval (QTc) will be calculated using Bazett's correction and Fridericia's correction, if necessary. The formulas are:

$$QTc\ (Bazett) = QT / (RR^{0.5})$$

$$QTc\ (Fridericia) = QT / (RR^{0.33})$$

where RR = 60 / heart rate (bpm)

In addition, a categorical analysis of QTc will be performed for each time point. The number and percentage of patients in each QTc category (< 450 msec, 450-480 msec, 481-500 msec, and ≥ 500 msec) will be summarized at baseline and at each subsequent time point. Categories of changes from baseline (≥ 30 msec and ≥ 60 msec) will also be summarized.

Maximum QTc and maximum changes from baseline will be summarized similarly in a separate display.

Electrocardiogram abnormalities will be presented in a data listing.

5.11.4 Vital Signs

The actual values of vital sign parameters including oral temperature, pulse rate, systolic and diastolic blood pressure, and weight will be summarized over time for each treatment arm. Change from baseline will also be presented.

A by-patient listing will also be presented.

5.11.5 Eastern Cooperative Oncology Group (ECOG) Performance Status

Eastern Cooperative Oncology Group performance status and change from baseline will be summarized. Shifts from baseline to the worst postbaseline score will be tabulated by treatment arm.

5.11.6 Other Safety Assessments

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

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

6. MAJOR PROTOCOL DEVIATIONS

A listing will be presented for major protocol deviation, for example, if 1 of the following criteria is met:

1. Major inclusion/exclusion violation (any violations)
2. Have major dosing compliance errors (ie, ≤ 70%) or overdose (≥ 120%) or dispensing error
3. Take excluded concomitant medication (ie, systemic treatment with strong inhibitors of CYP1A2 [fluvoxamine, enoxacin, ciprofloxacin], strong inhibitors of CYP3A [clarithromycin, telithromycin, itraconazole, voriconazole, ketoconazole, nefazodone, posaconazole] or strong CYP3A inducers [rifampin, rifapentine, rifabutin, carbamazepine, phenytoin, phenobarbital], or use of Ginkgo biloba or St. John's wort);
4. Missed pregnancy testing;

7. CHANGES TO PLANNED ANALYSES FROM PROTOCOL

Reference materials for this statistical plan include Clinical Study Protocol C16010 Amendment 3 (Protocol dated 08 July 2014). Additional major changes inlcude:

- Futhur clarify per protocol populaiton using programatic approach

8. PROGRAMMING CONSIDERATIONS

8.1 Statistical Software

SAS version 9.1 (or higher) will be used for all analyses, except for RPSFT and Uno's method, which may be done using R (version 3.0.1 or higher)

8.2 Rules and Definitions

Patient populations are defined in Section **Error! Reference source not found.Error! Reference source not found.**.

Baseline values are defined in Section 5.4.2.

9. REFERENCES

1. Lan K, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659-63.
2. Robins JM, Hernan M, Brumback B. Marginal strucural models and causal inference in epidemiology. Epidemiology 2000; 11(5): 550-560.
3. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000;56(3):779-88.
4. D'Agostino R.B., Lee M.L., Belanger A.J., Cupples L.A., Anderson K., and Kannel W.B. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. Statistics in Medicine, 9(12): 1501–1515, 1990.
5. Xie J. and Liu C. (2005). Adjusted Kaplan-Meier Estimator and Log-Rank Test with Inverse Probability of Treatment Weighting for Survival Data. Statistics in Medicine, 10;26(10):2276, 2007.
6. Uno H, Tian L, Claggett B, Wei LJ. A versatile test for equality of two survival functions based on weighted differences of Kaplan-Meier curves. Harvard University Biostatistics Working Paper Series 2013;Paper 159.

10. APPENDIX

10.1 Proof of Strong Control of Type I Error Rate for Key Secondary Endpoints

Proof of strong control of Type I error rate for both key secondary endpoints (OS and OS in high-risk patients carrying del(17)):

With the proposed testing procedure of the 2 key secondary endpoints (OS and OS in high-risk patients carrying del(17)), this is to prove the strong control of overall Type I error rate for both endpoints at one-sided 0.025 level.

To facilitate the probability presentation, we introduce the following notations. Let P, S indicate the first and second key secondary, OS and OS in high-risk patients carrying del(17). Denote the family of null hypotheses of interest as: $H_0^P : \theta^P = 0$ (OS is not efficacious); $H_0^S : \theta^S = 0$ (OS in high-risk patients is not efficacious). Let T_1^P, T_2^P be the log-rank test statistic for OS at the second interim and FA, and T_1^S, T_2^S be the log-rank test statistic for OS in high risk patients at the second interim and FA. Denote z_α as the upper α quantile from the standard normal distribution.

We use the fundamental multiple testing technique, partitioning principle to prove the overall Type I error rate control for both endpoints.

Table 10-1 Partition Hypotheses for Testing Both Key Secondary Endpoints

Index	Partition Hypothesis
1	$\theta^P = 0, \theta^S = 0$
2	$\theta^P = 0, \theta^S > 0$
3	$\theta^P > 0, \theta^S = 0$

By partition principle, as long as each of the partition hypothesis is tested at level 0.025, the overall Type I error rate is also controlled at the same level.

1. Under partition hypothesis $\theta^P = 0, \theta^S = 0$, a false rejection is {reject P or reject S}. Due to the closed sequential testing between P and S at both second IA and FA, rejecting S implies rejecting P, therefore probability of false rejection is simply the probability of rejecting P, i.e.

$$P(T_1^{\mathrm{P}} > z_{\alpha_1} \text{ or } T_2^{\mathrm{P}} > z_{\alpha_2})$$

where α_1 and α_2 are calculated based on the actual information fraction at second IA from O'Brien-Fleming boundary (Lan-DeMets method). Base on this boundary, the probability is no greater than 0.025.

2. Under partition hypothesis $\theta^P = 0, \theta^S > 0$, the probability of false rejection: Pr{reject P} is again controlled at 0.025 by O'Brien-Flemming boundary.
3. Under partition hypothesis $\theta^P > 0, \theta^S = 0$, a false rejection is {reject S}={(reject P at IA and reject S at IA) or (failed to reject P at IA and reject P at FA and reject S at FA)} which is a subset of {reject S at IA or reject S at FA}. Therefore, the probability of false rejection is bounded by:

$$P(T_1^{\mathrm{S}} > z_{\alpha_1} \text{ or } T_2^{\mathrm{S}} > c) = P(T_1^{\mathrm{S}} > z_{\alpha_1}) + \mathrm{P}(T_1^{\mathrm{S}} \le z_{\alpha_1} \text{ and } T_2^{\mathrm{S}} > c) \qquad (1)$$

The critical value c will be calculated as follows

$$\mathrm{P}(T_1^{\mathrm{S}} \le z_{\alpha_1} \text{ and } T_2^{\mathrm{S}} > c) = 0.025 - \alpha_1$$

to maintain the probability equation (1) at 0.025 level using following asymptotic multivariate normal distribution:

$$\begin{pmatrix} T_1^S \\ T_2^S \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\frac{n_1}{n_2}} \\ \sqrt{\frac{n_1}{n_2}} & 1 \end{pmatrix} \right)$$

where n_1 and n_2 are the size of the patient population at the second IA and FA, respectively.