Statistical Analysis Plan

A PHASE 3, RANDOMIZED, OPEN-LABEL, CONTROLLED, MULTICENTER STUDY OF ZANDELISIB (ME-401) IN COMBINATION WITH RITUXIMAB VERSUS STANDARD IMMUNOCHEMOTHERAPY IN PATIENTS WITH RELAPSED INDOLENT NON-HODGKIN'S LYMPHOMA (INHL) – THE COASTAL STUDY

ZANDELISIB (ME-401)

NCT 04745832

August 26, 2020

11455 El Camino Real, Suite 250, San Diego, CA 92130 USA
Tel: +1 858 369 7100 Fax: +1 858 369 7101 www.meipharma.com

STATISTICAL ANALYSIS PLAN

A Phase 3, Randomized, Open-Label, Controlled, Multicenter Study of Zandelisib (ME-401) in Combination With Rituximab Versus Standard Immunochemotherapy in Patients With Relapsed Indolent Non-Hodgkin's Lymphoma (iNHL)

Protocol Number:	ME-401-004
Protocol Version:	Original: 21 July 2020
Sponsor:	MEI Pharma, Inc. 11455 El Camino Real, Suite 250 San Diego, CA 92130
Sponsor Representative:	Joanne Li, PhD Ex. Director, Biometrics MEI Pharma, Inc.
Document Date:	August 26, 2020
Document Version:	1.0

SIGNATURE PAGE

Sponsor:	MEI Pharma, Inc.
Protocol Title:	A Phase 3, Randomized, Open-Label, Controlled, Multicenter Study of Zandelisib (ME-401) in Combination With Rituximab Versus Standard Immunochemotherapy in Patients With Relapsed or Refractory Indolent Non-Hodgkin's Lymphoma (iNHL)
Protocol Number:	ME-401-004
Document Date: **Version:**	August 26, 2020 1.0
Joanne Li Sr. Director, Biometrics MEI Pharma Inc.	Signature:_____ Date:_____
Igor Gorbatchevsky VP, Clinical Development MEI Pharma Inc.	Signature:_____ Date:_____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
AESI	Adverse Event of Special Interest
B	bendamustine
C	cycle
CHOP	combination of cyclophosphamide, hydroxydoxorubicin, vincristine and prednisone
CMH	Cochran–Mantel–Haenszel
CR	Complete Response
CRR	Complete Response Rate
CTCAE	Common Terminology Criteria for Adverse Events
CVP	Combination of Cyclophosphamide, Vincristine, Prednisone
D	Day
DMC	Data Monitoring Committee
DOR	Duration of Response
DRS-P	Disease-Related Symptoms Subscale – Physical
EE	Efficacy Evaluable
EMA	European Medicines Agency
FDA	Food and Drug Administration
HR	Hazard Ratio
(i)NHL	(indolent) Non-Hodgkin Lymphoma
IPD	Important Protocol Deviation
IRRC	Independent Response Review Committee
ITT	Intent to Treat
KM	Kaplan-Meier
L	lenalidomide
mAb	monoclonal antibody
O	obinutuzumab
ORR	Overall Response Rate
OS	Overall Survival
PD	Progressive Disease
PFS	Progression-Free Survival
PK	pharmacokinetic(s)
PP	Per Protocol
PR	Partial Response
PRO	Patient Reported Outcome
QoL	Quality of Life
R-B	rituximab plus bendamustine
R-CHOP	combination of rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone
R-CVP	combination of rituximab, cyclophosphamide, vincristine, prednisone
R-L	lenalidomide plus rituximab combination therapy
SAE	Serious Adverse Event

Abbreviation	Definition
SAP	Statistical Analysis Plan
TEAE	Treatment-Emergent Adverse Event
TTNT	Time to Next Anti-Lymphoma Treatment
TTP	Time to Progression
WHODRUG	World Health Organization Drug

1. Introduction

The purpose of this statistical analysis plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol for Study ME-401-004 dated **21 July 2020**. The scope of this plan includes the safety, efficacy, and key Patient-Reported analysis (PRO). PKPD analysis plan will be provided in a separate PKPD analysis plan. Detailed PRO analysis will be provided in a separate supplement SAP.

2. Objective

Primary Objective:

- To demonstrate that zandelisib in combination with Rituxumab [R] is superior to standard immunochemotherapy in prolonging PFS as determined by the Independent Response Review Committee (IRRC) in previously treated subjects with follicular and marginal zone lymphoma

Secondary objectives

- To compare zandelisib + R to standard immunochemotherapy by the following efficacy measures
    - Complete response rate (CRR), Objective Response Rate (ORR), duration of response (DOR) as determined by the IRRC
    - Overall survival (OS)
- To evaluate the safety and tolerability of zandelisib in combination with R
- To evaluate Patient Reported Outcome (PRO) assessment with FlymSI-18

Exploratory objectives

- To evaluate:
    - PFS, CRR, ORR, and DOR as determined by the Investigator
    - ORR at week 24 by the Investigator and by the IRRC
    - Time to progression (TTP) by the Investigator and by the IRRC
    - Time to next anti-lymphoma treatment (TTNT)
- To correlate zandelisib exposure in plasma with efficacy and safety

3. Study Design

3.1. Synopsis of Study Design

This is an open label, two-arm Phase 3 study in subjects with relapsed or refractory FL or MZL to evaluate efficacy and safety of zandelisib in combination with rituximab compared with standard immunochemotherapy (R-B [rituximab + bendamustine], or R-CHOP [combination of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone]). The subjects must have relapsed after ≥1 previous line of therapy; prior therapy must have included anti-CD20 monoclonal antibody treatment (e.g., R, obinutuzumab [O]) in combination with a chemotherapeutic regimen containing an alkylating agent (B or CVP or CHOP or similar), or lenalidomide [L]. Subjects will be stratified based on the following criteria:

- Prior treatment regimen: anti-CD20 monoclonal antibody (mAb) in combination with non-bendamustine chemotherapy regimen or R-L vs. anti-CD20 mAb in combination with B
- NHL histology: FL vs. MZL
- Duration of treatment-free interval from the last lymphoma-directed therapy: ≤24 months vs. >24 months

Patients who meet the eligibility criteria will be randomly assigned in a 1:1 ratio to one of the treatment arms:

- Arm 1: Rituximab plus zandelisib
- Arm 2: Rituximab plus chemotherapy (CHOP or B)

During the study, an independent Data Monitoring Committee (DMC) will regularly review safety and efficacy data from all subjects as outlined in a separate DMC charter document.

3.2. Study Schema

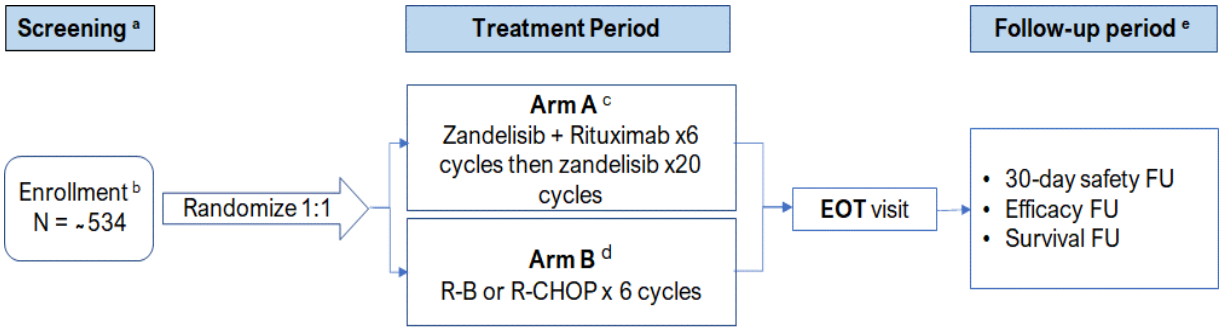
The study is composed of the following periods (Figure 1):

- **Screening**: Subjects will be screened during the 28-day period prior to Day 1 (D1) Cycle 1 (C1) to ensure they meet the entry criteria for the study.
- **Treatment**: After meeting eligibility criteria, subjects will be randomized 1:1 the 2 arms. Subjects randomized to the comparator arm will be allocated to either R-B or R-CHOP according to the treatments they may have received previously.
  - Arm 1 (Rituximab plus zandelisib): Rituximab will be administered for up to a total of 8 doses in 6 cycles. Zandelisib will be administered 60 mg qd for the first 2 cycles of therapy (56 days), then 60 mg for 7 days followed by 21 days off treatment in every 28-day cycle, **for up to 26 cycles**
  - Arm 2: R-B or R-CHOP: R-B will be administered every 4 weeks (q4w) **for 6 cycles.** R-CHOP will be administered once every 3 weeks (q3w) **for 6 cycles**.
- **Safety Follow-up:** 30 days from the last dose of study drug

- **Efficacy Follow-up:** Subjects who discontinue study treatment for any other reason than progressive disease (PD) or death will be followed for PFS till PD, death, withdrawal of consent, or start of new anti-cancer therapy for lymphoma.
- **Survival follow up**: After treatment discontinuation, all subjects will be followed for OS, and TTNT, if applicable, every 3 months for 5 years after the last subject started study treatment, except for subjects who withdraw consent.

The study duration for primary completion of PFS is approximately 64 months with 36 months enrollment period and 28 months follow-up.

Figure 1. Study Schematic



a. Screening from Day -28 to -1
b. Subjects should initiate therapy within 5 days of confirmation of eligibility
c. Zandelisib 60 mg QD for 2 cycles followed by IS schedule 7 days on 21 days off treatment; combination with rituximab is given for 6 cycles, followed by zandelisib for additional 20 cycles of therapy. When administered with zandelisib, R is given on Days 1, 8, 15, and 22 in Cycle 1 then on Day 1 of Cycles 3, 4, 5 and 6
d. R-B or R-CHOP therapy for 6 cycles; R given on Day 1 of each cycle of chemotherapy
e. Follow-up (FU) will include
   - Safety FU at 30 days (± 3 days) from the last dose of study drug. All SAEs will be reported and followed for up to 90 days after the last dose of study drug. After the 90 days only SAEs deemed related to study drug will be reported to the Sponsor
   - Efficacy – for subjects who discontinue treatment for reasons other than PD or death, response FU will continue until PD, death, withdrawal of consent, or start of new anti-cancer therapy for lymphoma
   - Survival FU every 3 months for 5 years until death or withdrawal of consent

Abbreviations: EOT, end of treatment; FU, follow-up; IS, intermittent dosing schedule; PD, progressive disease; R, rituximab; R-B, rituximab + bendamustine; R-CHOP, R + cyclophosphamide, hydroxydoxorubicin, vincristine, and prednisone

4. Analysis Population and Subgroups

4.1. Analysis Population

The following analysis populations will be evaluated and used for presentation and analyses of the data.

- The intent-to-treat (ITT) population: all subjects who are randomized.
- Efficacy Evaluable (EE) Population: all subjects who receive at least one dose of study drug and have at least one response assessment post baseline
- The safety population: all subjects who were randomized and received at least 1 dose of any study drug.
- The PK population is defined as all subjects who had at least two PK samples collected and analyzed.
- Per Protocol (PP) Population: all subjects who receive at least one dose of study drug, have at least one response assessment after Day 1, and do not have any major protocol deviation during the study.

4.2. Subgroups

Exploratory analyses to examine the consistency of the treatment effect for the primary endpoint and selected secondary endpoints will be performed within the subgroups defined by each stratification factor:

- Prior treatment regimen: anti-CD20 mAb in combination with any chemotherapy regimen other than bendamustine/Other vs. anti-CD20 mAb with bendamustine
- NHL histology: FL vs. MZL
- Duration of treatment-free interval from the last lymphoma-directed therapy: ≤24 months vs. >24 months

For these stratification factors, data collected by CRF will be used in subgroup analyses, if there are discrepancies between the CRF and stratification factors used in IWRS.

Additional subgroup analyses for the efficacy endpoints will include, but not limited to the following factors:

- Age (≥65, <65)
- Sex (M, F)
- Race (White, Asian, other)
- Geographical region
- # of prior anti-lymphoma therapies (1, >1)
- Relapse or refractory to last therapy/regimen prior to study enrollment
- Longest diameter of the longest node (>=60 mm, <60 mm) at baseline

- Longest diameter of the longest node (>=50 mm, <50 mm) at baseline
- Baseline ECOG status (0, 1)
- Ann Arbor disease stage (1-2 vs 3-4)
- Duration of treatment-free interval from the last lymphoma-directed therapy: ≤6 months, 6-24 months vs. >24 months

Safety endpoints including but not limited to SAEs and Grade 3 and above TEAEs will be descriptively summarized within subgroups defined by each stratification factor, and in addition defined by the following variables:

- Age (≥65, <65)
- Sex (M, F)
- Race (White, Asian, other)
- Geographical region
- # of prior anti-lymphoma therapies (1, >1)

In addition to support regional regulatory submissions, key efficacy and safety analyses will be performed including all patients in the corresponding regions or countries.

5. Endpoints and Definitions

5.1. Primary

- Progression Free Survival (PFS) as determined by the IRRC.

5.2. Secondary

- Efficacy: CRR, ORR, and DOR as determined by the IRRC
- OS
- TEAEs, Serious AEs (SAEs), and laboratory abnormalities
- PRO– time to deterioration in the 9-item disease-related symptoms subscale-physical (DRS-P) subset of FlymSI-18 of ≥5 points

5.3. Exploratory

- Efficacy:
    - PFS, CRR, ORR, and DOR as determined by the Investigator
    - ORR at 24 weeks by the Investigator and by the IRRC
    - TTP by the Investigator and by the IRRC
    - TTNT
- Estimates of zandelisib exposure by population PK method and association with efficacy and safety variables

5.4. Definitions

The definitions for the efficacy endpoints are based on Lugano criteria (**Error! Reference source not found.**, 2007) and FDA guidance (**Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.**, 2015) and (**Error! Reference source not found.**, 2018).

PFS

PFS time is measured from randomization date until the date of disease progression or death from any cause. Patients who have not progressed and are still alive at the time of analysis will be censored at the last disease assessment date indicating the absence of progression. Detailed censoring rules are provided in Section 6.7.1.

ORR

ORR is defined as the proportion of subjects achieving the best response rating of CR or PR based on the Lugano Classification over the entire duration of the study, including the efficacy follow-up period.

ORR at Week 24

ORR is defined as the proportion of subjects achieving the best response rating of CR or PR based on the Lugano Classification at the Week 24 assessment.

CRR

CRR is defined as the proportion of subjects achieving the best response rating of CR based on the Lugano Classification over the entire duration of the study, including the efficacy follow-up period.

CRR at Week 24

CRR is defined as the proportion of subjects achieving the best response rating of CR based on the Lugano Classification at the Week 24 assessment.

Duration of response (DOR)

DOR is defined as the time from first observed tumor response (CR or PR) until PD. Patients who achieved an objective response (PR or CR) and who did not have a PD at the time of analysis will be censored at the date of their last documented tumor assessment demonstrating an objective response. DOR will only be analyzed for patients with at least one CR or PR.

OS

Overall survival is defined as the time from randomization until death from any cause. Patients alive are censored at the last documented alive date. OS is measured in the intent-to-treat population.

Time to Progression (TTP)

TTP is defined as the time from randomization to PD. Patients without PD at the time of analysis will be censored at the date of their last documented tumor assessment demonstrating no evidence of PD.

Time to next anti-lymphoma treatment (TTNT)

Time to next anti-lymphoma treatment (TTNT) is defined as the time from date of randomization to date of first documented administration of a new anti-lymphoma treatment (including, but not limited to chemotherapy, radiotherapy, radioimmunotherapy or immunotherapy). Patients continuing in response or who are lost to follow-up will be censored on their last visit date.

Patients who died (due to any cause) before having received a new anti-lymphoma treatment will be censored at the date of death.

Time to deterioration in the 9-item DRS-P subset of FlymSI-18

Time to deterioration in the 9-item DRS-P subset of FlymSI-18 is defined as time from randomization to first increase of DRS-P score from baseline ≥5 points.

TEAE

A TEAE is defined as an AE starting or worsening per CTCAE grade after the first dose until 30 days after the last dose of study drug, or start of new anti-cancer therapy, whichever is earlier.

6. Statistical Method

6.1. Sample Size and Power

The sample size estimation is determined based on the primary endpoint PFS.

It is estimated that approximately 534 subjects will be randomized into this study.

Assuming a hazard ratio (HR) of 0.70, 330 PFS events at primary analysis are required to yield 87.5% power to detect superiority of zandelisib in combination with rituximab over standard immunochemotherapy.

An interim analysis is planned to take place after approximately 248 (75% of the 330 planned) PFS events have occurred. If the statistical significance for PFS is not achieved at the interim analysis, the primary PFS analysis will be performed after 330 PFS events have been observed.

Assuming a median time to PFS for the control group of 24 months, an accrual duration of 36 months, and an annual drop-out rate is 3%, a total sample size of approximately 534 is required with 28 months minimal follow-up time from the last patient enrolled to conduct the planned primary analysis.

OS is a secondary endpoint. The testing procedure for OS and all other secondary endpoints is specified in Section 6.7.2. There will be no additional interim looks driven specifically by total OS events, as OS will be analyzed at the same time when the interim and primary analysis for PFS take place.

6.2. General Method

Study disposition, patient demographic and baseline characteristic will be analyzed using the ITT population.

Efficacy analyses will be performed in the ITT population.

Primary efficacy endpoint and selected secondary efficacy analyses will be performed in the EE population and PP population.

Safety analyses will be performed on the Safety Population.

PK analysis will be performed for all subjects in the PK population.

Unless otherwise noted, all analyses will be performed by assigned treatment arm.

Because the objective of the study which is to test the hypothesis that zandelisib in combination with rituximab has better clinical activity and risk/benefit profile compared to standard 2nd line immunochemotherapy (R-CHOP/R-B) in patients with relapsed FL or MZL, all tests of treatment effects will be conducted at a 1-sided alpha level of 0.025, unless otherwise stated.

Primary efficacy and safety analyses will be conducted comparing the 2 treatment groups (pooled control versus pooled experimental); however, exploratory analyses will be conducted to assess any clinically meaningful differences in treatment effect between zandelisib plus rituximab vs. immunochemotherapy in each subgroup defined by prior therapy. If differences are present, sensitivity analyses will be performed to account for the heterogeneity in different subgroups in order to ensure the overall consistency and robustness of the primary analysis.

Continuous variables will be summarized by the non-missing sample size (n), mean, standard deviation, median, first and third quartiles, minimum, and maximum. Categorical variables will be summarized by the n and percentage in each category. Time to event endpoints will be summarized with hazard ratios, Kaplan-Meier (KM) curves, KM proportions at select time points, KM quartiles (when estimable), the number of subjects with events, the number of subjects censored, and the pattern of censoring. Point estimates for efficacy endpoints will be accompanied by 2-sided 95% confidence intervals including estimates of KM quartiles (Brookmeyer and Crowley, 1982), KM proportions (Kalbfleisch and Prentice, 1980), and binomial proportions (Clopper CJ and Pearson, 1934).

Hypotheses and/or Estimations

The null hypothesis is that there is no difference between treatment groups with respect to PFS versus the alternative hypothesis that rituximab plus zandelisib is superior to rituximab plus chemotherapy. The null hypothesis will be rejected if the p-value from a one-sided stratified log-rank test is less than the value specified by the alpha spending function specified in Section 6.7.1 at the given analysis (interim or primary).

6.3. Data Handling

6.3.1. Handling of missing data

Rules for handling missing data related to endpoints are described in the endpoint definitions (Section 5) or in the description of analyses. If dates are partially provided, provide the worst case scenario. If a date is missing, provide the last day of month. If a month is missing, provide the last month of year. If a year is missing, then the date will not be imputed.

6.4. Subject Disposition

The number and percent of subjects who were screened, randomized, received protocol-specified therapy along with the reasons for discontinuing protocol-specified therapy and discontinuing study will be summarized by treatment group. The number and percent of subjects randomized will be tabulated by study site. Key study dates for the first subject randomized, last subject randomized, and data cut-off date for analysis will be presented.

6.5. Important protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's visit and updated during the IPD reviews throughout the study prior to database lock.

These definitions of IPD categories, sub-category codes and descriptions will be used during the course of the study.

6.6. Demographic and baseline characteristics

Demographic (ie, age, age group [<65, ≥65], sex, race, and geographical region) and baseline disease characteristics (including but not limited to the randomization factors) will be summarized by treatment group and overall using descriptive statistics for the ITT population.

6.7. Efficacy Analyses

6.7.1. Analysis of primary efficacy endpoints

The primary endpoint in this study is PFS. PFS time is measured from randomization date until the date of disease progression, or death from any cause. Subjects who discontinue study treatment for any other reason than progressive disease (PD) will be followed for PFS. Detailed censoring rules are specified in Table 1 (FDA guidance [FDA, 2015] and [FDA, 2018]).

The PFS analysis to test the superiority of zandelisib in combination with rituximab to standard immunochemotherapy in prolonging PFS time will use the log-rank test stratified by the following randomization factors:

The Kaplan-Meier (KM) method will be used to estimate the median time to PFS with 95% CI. Landmark PFS rates at Years 1,2 and 3 will be estimated for each treatment group and compared by chi-squared test based on the LOGLOG transformation of the corresponding rates.

The Cox proportional hazard model will be used to estimate the hazard ratio and corresponding 95% confidence interval with Wald's test p-value after stratifying for the same randomization factors. Additional covariates will be considered in the Cox model in exploratory analyses, eg., demographical variables and baseline disease characteristics.

In addition, accounting for violation of proportional hazard ratio assumption and considering potential delayed treatment effect, stratified Fleming(0,1) test (Fleming and Harrington, 1981) which gives more weight to later events will be utilized as a sensitivity analysis:

```
proc lifetest data=datasetname ;
      time time_variable*censor_variable(list of censored values);
      strata strat_variable_1 strat_variable_2 strat_variable_3
      /group=tx_variable test=Fleming(0,1)
run;
```

Table 1. Censoring Rules for the Primary Analysis of Progression-free Survival

Situation	Date of Event or Censoring	Outcome
Progression documented	Date of earliest assessment which revealed progression determined by the IRRC	Event
Death before first PD assessment	Date of death	Event
Death between adequate assessment visits	Date of death	Event
Incomplete or no baseline tumor assessments	Randomization	Censored
No progression, nor death	Date of last adequate assessment with evidence of no progression by the IRRC	Censored
Treatment discontinuation for any reason other than documented PD or death	Date of last adequate assessment with evidence of no progression by the IRRC	Censored
New anti-lymphoma treatment started without evidence of PD by IRRC	Date of last adequate assessment with evidence of no progression by the IRRC before the start of new anti-lymphoma treatment	Censored
Death or progression after two or more consecutive missed scheduled visits	Date of last adequate assessment with evidence of no progression by the IRRC	Censored

Every effort will be made to minimize missing data for PFS assessment. The follow-up time and the reason for censoring will be summarized for the 2 treatment arms and the impact on PFS outcome will be assessed.

Sensitivity analysis will be conducted to evaluate the robustness of the primary analysis results, by considering all progressions and deaths as events per European Medicines Agency (EMA) guidance (EMA, 2013), regardless of whether they occurred after initiating next anti-lymphoma treatment, after 2 or more consecutive missed scheduled assessments, or being reported as clinical (non-IRRC reviewed radiographic) progression.

Below are the censoring rules for sensitivity analysis of PFS:

Table 2. Censoring Rules for the Sensitivity Analysis of PFS

Situation	Date of Event or Censoring	Outcome
Progression documented	Date of earliest assessment which revealed progression determined by the IRRC	Event
Death before first PD assessment	Date of death	Event
Death between adequate assessment visits	Date of death	Event
Incomplete or no baseline tumor assessments	Randomization	Censored
No progression, nor death	Date of last adequate assessment with evidence of no progression by the IRRC	Censored
Treatment discontinuation for any reason other than documented PD or death	Date of last adequate assessment with evidence of no progression by the IRRC	Censored
New anti-lymphoma treatment started *	Date of earliest assessment which revealed progression determined by the IRRC or death	Event
Death or progression after two or more consecutive missed scheduled visits *	Date of earliest assessment which revealed progression determined by the IRRC or death	Event
Investigator claim of clinical progression leading to treatment discontinuation *	Date of clinical progression	Event

* Outcome definitions are different from primary analysis of PFS

6.7.2. Analysis of secondary efficacy endpoints

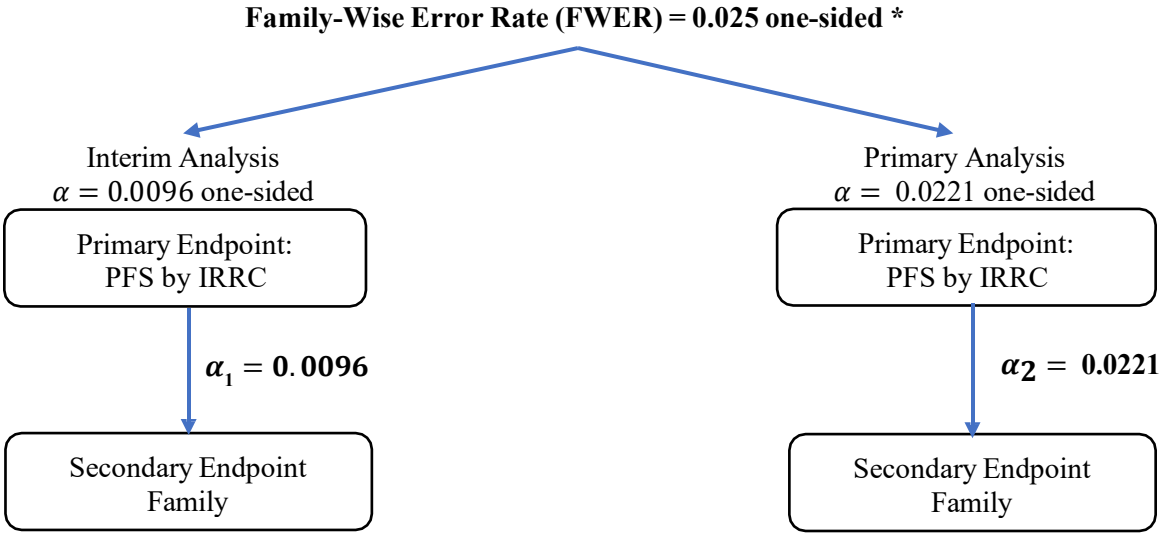
If the statistical significance is successfully met at the interim analysis or the primary analysis, the testing of secondary endpoints will proceed, by utilizing a graphical testing procedure.

A graphical multiple testing approach ([Bretz, 2009] and [Alosh, 2014]) will be implemented to control the overall Family-wise Error Rate (FWER) pre-specified at interim analysis and primary analysis respectively (Figure 2), for all primary and secondary efficacy endpoints (Figure 3). This graphical approach is a closed testing procedure; hence, it strongly controls the family-wise error rate across all endpoints. Each hypothesis is represented as a node in a graph. Directed

arrows between the nodes with associated weights represent how alpha is passed from its initial allocation to other nodes, i.e., proportion of the alpha which will be propagated to the corresponding endpoint(s). The weights being spent from each node (i.e., going out from each node) will sum up to 1. The testing scheme are fully pre-specified by the graph (including nodes, arrows, and weights) along with the initial alpha allocation. Figure 3 describes the graphical scheme, and pre-specified alpha at interim and primary analysis will be allocated to the primary endpoint PFS by IRRC initially.

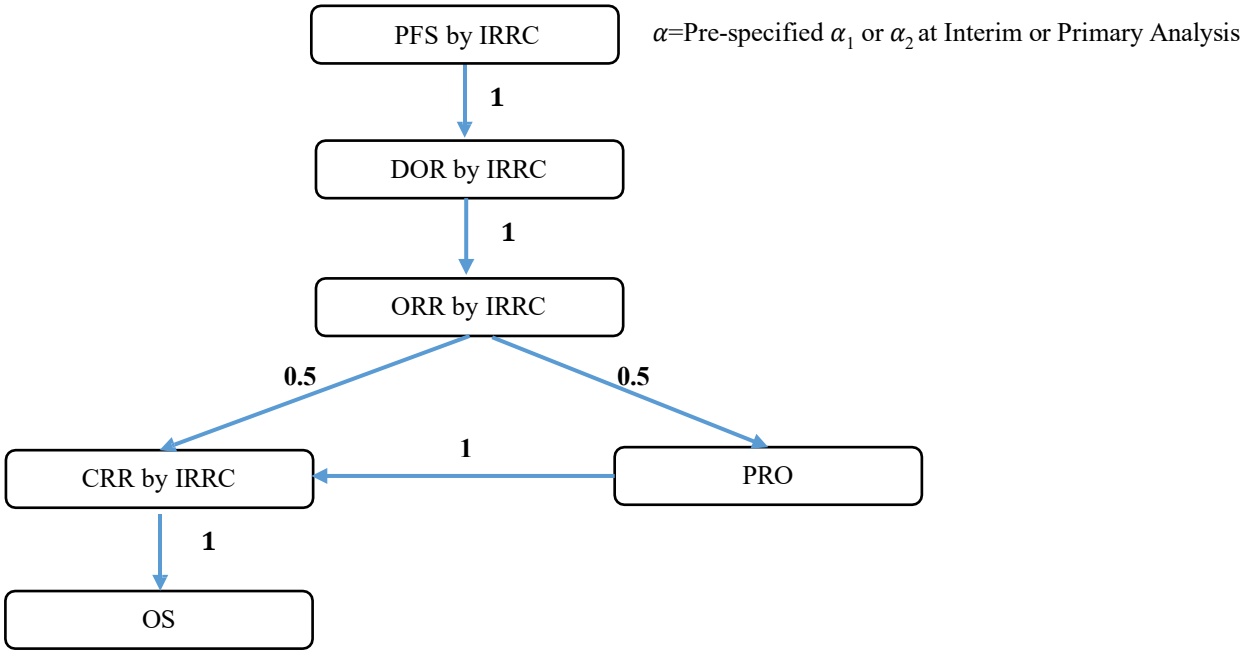
The testing scheme will be finalized before the database lock for interim analysis.

Figure 2. Illustrative of FWER spending within interim analysis and primary analysis, and between primary analysis



*Alpha-spending based on Lan-DeMets O'Brien-Fleming approximation spending function

Figure 3. Graphical testing scheme for FWER control withing secondary endpoint family



*PRO = time to deterioration in the 9-item disease-related symptoms subscale-physical (DRS-P) subset of FlymSI-18 of ≥5 points

ORR and CRR

Differences in ORR and CRR between treatment groups will be tested using Cochran–Mantel–Haenszel (CMH) method stratified by the randomization factors. In addition, the percentage of subjects in each treatment group with an OR (or CR) will be summarized with an exact binomial 95% confidence interval. Subjects missing post-baseline disease assessments will be considered not to have achieved OR (or CR). The primary analysis will be performed on the ITT Set. Sensitivity analyses may be performed on the EE set and Per Protocol Set.

DOR and OS

DOR and OS will be tested between the two treatment groups using stratified log-rank test. The Kaplan-Meier (KM) method will be used to estimate the median OS (or DOR) with 95% CI. Landmark overall survival rates at specific time points (Years 2,3,4,5) will be estimated from KM curves.

DOR will only include the progressions as determined by the IRRC as events. Subjects without progressions documented by the IRRC will be censored. Table 3 provides detailed censoring rules for DOR.

OS will be censored at the last date that the subject was known to be alive for subjects who were alive at the time of analysis and for subjects who were lost to follow-up. Last date known alive is defined as the last valid date of subject assessment prior to or on the data cutoff date in the clinical database. For subjects who have withdrawn consent during the study, the last date known alive will be the date of consent withdrawal from the study. For all other subjects, the last date known alive will be derived by searching through all valid assessment dates in all study datasets to identify the last valid subject assessment date available for each subject. If the last valid subject assessment date available is on or prior to the data cutoff date, it is used as the last date known alive. If the last valid subject assessment date available is after the data cutoff date, the data cut-off date is used as the last date known alive.

Patient-Reported Outcome (PRO) Analyses

Time to deterioration in the 9-item DRS-P subset of FlymSI-18 will be summarized by stratified log-rank test and KM method.

Time to deterioration in the 9-item DRS-P subset of FlymSI-18 is defined as time (in months) from randomization to the occurrence of one of the following, whichever occurs earliest:

- First increase of DRS-P score from baseline ≥3 points
- Disease progression as determined by the IRRC
- Death from any cause

For patients without any event of above but had at least one assessment of DRS-P, they will be censored at last assessment of DRS-P. For patients without any event of above and no assessment of DRS-P, they will be censored at randomization date.

Sensitivity analysis for time to deterioration in the 9-item DRS-P subset of FlymSI-18 may be defined by using different threshold (e.g, >=3, 4, 6, or 7 points).

Summary statistics will be provided for the 9-item DRS-P subset of FlymSI-18 by visits for both treatment arms.

Additional PRO analyses may be specified in a supplemental SAP.

Table 3. Censoring Rules for DOR

Situation	Date of Event or Censoring	Outcome
Progression determined by the IRRC	Date of earliest assessment which revealed progression determined by the IRRC	Event
No progression including • No progression determined by the IRRC • Death with no progression documented by the IRRC • Treatment discontinuation due to progression which was not determined by the IRRC	Date of last adequate assessment with evidence of no progression by the IRRC	Censored
Treatment discontinuation for toxicity or other reason	Date of last adequate assessment with evidence of no progression by the IRRC	Censored
New anti-lymphoma treatment started	Date of last adequate assessment with evidence of no progression by the IRRC before the start of new anti-lymphoma treatment	Censored
Progression after two or more consecutive missed scheduled visits	Date of last adequate assessment with evidence of no progression by the IRRC	Censored

6.7.3. Analysis of exploratory efficacy endpoints

Exploratory endpoints including TTP, TTNT will be summarized by stratified log-rank test, KM method, and Cox proportional hazard model.

The censoring rules for TTP are the same as the ones for DOR (Table 3).

The censoring rules for TTNT are provided in Table 4.

Table 4. Censoring Rules for TTNT

Situation	Date of Event or Censoring	Outcome
Administration of any new anti-lymphoma treatment (e.g., chemotherapy, radiotherapy, radio-immunotherapy, immunotherapy)	Date of first administration of the treatment	Event
Death without any new anti- lymphoma treatment	Date of death	Censored
No Death or new administration of any new anti-lymphoma treatment	Last day know alive	Censored

6.8. Safety Analyses

All safety summaries and analyses will be based on the Safety Population, defined as all enrolled patients receiving at least 1 dose of any study drug. Safety analyses will be performed by randomized group.

6.8.1. Adverse Events

Overall summary of subject incidence of AEs will include:

- TEAEs, including severity and possible relationship to study drug and/or study treatment
- AESI s
- SAEs
- TEAEs leading to study treatment discontinuation
- Deaths during treatment period (from first dose date to 30 days after last dose date of any study drug)
- Treatment-emergent changes in laboratory values
- Treatment-emergent changes in vital signs.

Time to AESIs will be calculated from the first day of study treatment to first occurrence of AESI, utilizing KM method.

Due to likely imbalance of treatment duration between 2 treatment arms, exposure duration adjusted analysis will be performed for AESIs and SAEs. The exposure duration adjusted rate is defined as the total number of specified events over the total person years, where the person year for each individual subject is calculated as the days from first dose till last dose date plus 30 days divided by 365.25. The confidence interval will be constructed using exact method with F distribution (Ulm, 1990).

Safety endpoints including but not limited to SAEs and Grade 3 and above AESIs will be descriptively summarized within subgroups defined in Section 4.2.

Listings will be provided for subjects who had SAEs, AEs leading to treatment discontinuation, and died during treatment period.

6.8.2. Laboratory Data

Shift tables between the worst post-baseline and baseline grades for select laboratory parameters. Plots or other summaries overtime will be presented for select laboratory parameters including but not limited to white blood cell, absolute lymphocyte count, absolute neutrophil count, platelets, hemoglobin, alanine transaminase, aspartate transaminase, and creatinine by treatment group.

6.8.3. Vital Signs

The number and percentage of subjects with abnormal changes in systolic blood pressure, diastolic blood pressure and heart beat rate will summarized by treatment group for subjects in the Safety Population.

6.8.4. Concomitant Medications

The number and proportion of subjects receiving concomitant medications from study day 1 through end of treatment will be summarized by preferred term as coded by the World Health Organization Drug (WHODRUG) dictionary by treatment group in the Safety Population.

In addition, the number and proportion of subjects receiving other anti-cancer therapies during efficacy and survival follow-up will be summarized by WHODRUG preferred term for each treatment group in the Safety Set.

6.8.5. Exposures

Descriptive statistics will be produced to describe the exposure to study drugs by treatment group for subjects in the Safety Population. For both treatment groups, the number of cycles of protocol-specified therapy administered will be summarized. In addition, the duration of therapy, the cumulative dose, and treatment compliance (proportion of actual dose /intended dose) will be summarized for both arms. The number and percent of subjects with dose modifications (eg, dose changes, dose interruptions) and reason for modification will be summarized for both treatment groups.

6.9. Interim Analysis

A group-sequential design by including 1 interim analysis and 1 primary analysis is planned for this study. At each analysis, the hypothesis for the primary endpoint PFS will be tested using a 1-sided with the family-wise error rate alpha of 0.025 controlled to adjust for multiplicity, utilizing Lan-DeMets O'Brien-Fleming approximation spending function.

The interim analysis is planned to take place after approximately 248 (75% of the 330 planned) PFS events have occurred. The objective of this interim analysis is to examine the efficacy to allow an early declaration of success. Based on our estimated recruitment velocity and PFS event rates, the planned interim analysis will occur after all subjects have been enrolled and a projected median time on study of approximately 30 months (range: 12-48). Also, it is anticipated that all subjects randomized to the control arm would have completed their treatment regimen at the planned interim analysis.

If the statistical significance for PFS is not achieved at the interim analysis, the primary PFS analysis will be performed after 330 PFS events have been observed. The alpha-spending for the interim and primary analyses will be 0.0096 and 0.0221 if the two analyses are performed at exactly 248 and 330, respectively. The actual boundary for the final analysis will be calculated based on the actual number of events observed at the time of each analysis using software that implements the alpha-spending scheme noted above (for example SAS or similar software).

The interim PFS analysis will be performed by the DMC.

Once statistical significance is successfully met at the interim analysis or the primary analysis, the study will be declared positive based on the primary endpoint of PFS, and the testing of secondary endpoints will proceed.

7. Changes to Planned Analyses

No changes in the planned analysis have been made at this time.

8. References

Alosh M, Bretz F, Huque M (2014). Advanced multiplicity adjustment methods in clinical trials. Statistics in Medicine, 33:693-713.

Bretz F, Maurer W, Brannath W, Posch M (2009) A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine, 28(4):586-604.

Brookmeyer R, & Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38(1): 29-41.

Cheson BD, Pfistner B, Juweid ME (2007). Revised response criteria for malignant lymphoma. J Clin Oncol. 25(5):579-86.

Clopper C, Pearson E (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika, 26(4): 404-413.

EMA (2013) Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man (CHMP/EWP/205/95 Rev.4): methodological considerations for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials. EMA/CHMP/27994/2008/Rev.1.

FDA 2015 Guidance for Industry: Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics

FDA 2018 Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

Fleming TR and Harrington DP (1981). “A Class of Hypothesis Tests for One and Two Samples of Censored Survival Data”. Communications in Statistics—Theory and Methods, 10: 763–794.

Kalbfleisch JD, Prentice RL (1980). The statistical analysis of failure time data. John Wiley & Sons.

Ulm K (1990). A simple method to calculate the confidence interval of a standardized mortality ratio. American Journal of Epidemiology, 131(2):373-375.