STATISTICAL ANALYSIS PLAN

Study Title:	A Phase 2, Randomized, Double-Blind, Placebo-Controlled Study to Assess the Efficacy and Tolerability of Entospletinib, a Selective SYK Inhibitor, in Combination with Systemic Corticosteroids as First-Line Therapy in Subjects with Chronic Graft Versus Host Disease (cGVHD)
Name of Test Drug:	Entospletinib
Study Number:	GS-US-406-1840
Protocol Version (Date):	Amendment 3.0 (30 January 2017)
Analysis Type:	Interim and Final
Analysis Plan Version:	Version 2.0
Analysis Plan Date:	11 May 2018
Analysis Plan Author(s):	PPD

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BORR	best overall response rate
CI	confidence interval
cGVHD	chronic Graft Versus Host Disease
CMH	Cochran-Mantel-Haenszel
CPK	creatine phosphokinase
CR	complete response
CRF	case report form(s)
DMC	Data Monitoring Committee
DOR	duration of response
EBV	Epstein-Barr Virus
ECG	electrocardiogram
eCRF	electronic case report form(s)
ENTO	entospletinib
FSH	follicle-stimulating hormone
GGT	gamma-glutamyl transferase
GVHD	Graft Versus Host Disease
Hb	hemoglobin
HBV	hepatitis B virus
HCV	hepatitis C virus
HIV	human immunodeficiency virus
ITT	intent-to-treat
IWRS	interactive web response system
KM	Kaplan-Meier
LLOQ	lower limit of quantitation
LOQ	limit of quantitation
LSS	Lee Symptom Scale
MedDRA	Medical Dictionary for Regulatory Activities
MMF	mycophenolate mofetil
MMRM	mixed model repeated measures
NCAA	NIH cGVHD Activity Assessment
NIH	National Institute of Health
PCP	Phencyclidine
PFT	pulmonary function tests
PK	pharmacokinetic
PPI	proton pump inhibitor
PTM	placebo-to-match
Q1	first quartile
Q3	third quartile

RNA	ribonucleic acid
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SOC	System Organ Class
SYK	spleen tyrosine kinase
ULN	upper limit of the normal range
WBC	white blood cell
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical analysis methods and data presentations to be used in tables, figures, and listings (TFLs) for the planned interim and final analyses for Study GS-US-406-1840. This SAP is based on Protocol Amendment 3 dated 30 January 2017 and the electronic case report form (eCRF). The SAP was finalized before the final analysis. The changes made after the finalization of the SAP will be documented in the CSR.

1.1. Study Objectives

The primary objective of this study is:

- To evaluate the effect of entospletinib (ENTO) on the best overall response rate (BORR) as assessed by the NIH cGVHD Activity Assessment (NCAA) by 24 weeks in the setting of add-on to systemic corticosteroids as part of first-line therapy for cGVHD

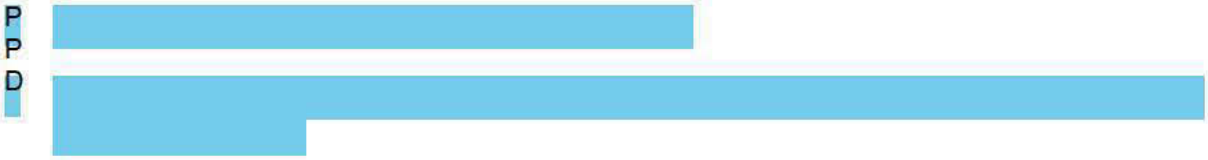
The key secondary objectives of this study are:

- To evaluate the effect of ENTO on the skin domain of the Lee Symptom Scale (LSS) at 24 weeks
- To evaluate the effect of ENTO on the mouth domain of the LSS at 24 weeks
- To evaluate the effect of ENTO on the eyes domain of the LSS at 24 weeks
- To evaluate the effect of ENTO on the total score of the LSS at 24 weeks

Additional secondary objectives are:

- To evaluate the effect of ENTO on duration of response (DOR)
- To evaluate the effect of ENTO on dose reduction of corticosteroids
- To evaluate the effect of ENTO on initiation of second-line therapy for cGVHD
- To evaluate the effect of ENTO on progression of cGVHD
- To evaluate the safety and tolerability of ENTO

Exploratory objectives include:

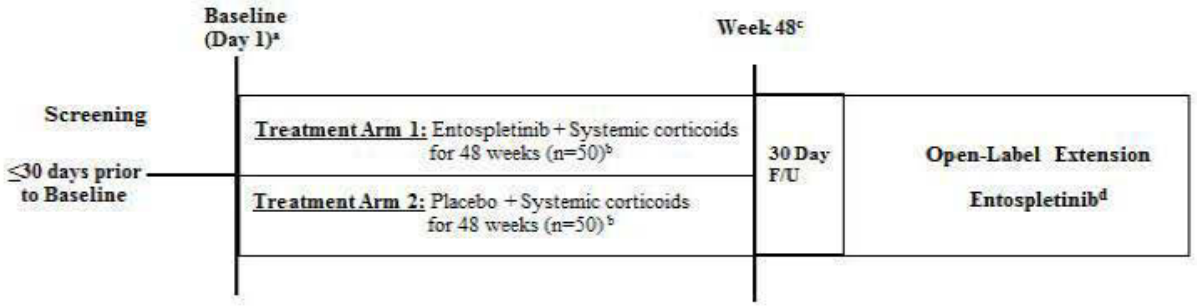


P
P
D

1.2. Study Design

This is a randomized, double-blind, placebo-controlled, multicenter study to evaluate the efficacy and safety of ENTO (Treatment Arm 1) vs. placebo (Treatment Arm 2) in combination with systemic corticosteroids in subjects with cGVHD. An interim analysis for futility is planned when approximately 50% of subjects have been on study for 24 weeks, or have discontinued the study before 24 weeks, or have achieved a response prior to reaching 24 weeks. The primary endpoint will be evaluated when all subjects complete 24 weeks of treatment. Duration and type of response (complete or partial) will be evaluated for up to 48 weeks of blinded treatment. After 48 weeks, subjects who are still on investigational treatment (ENTO or placebo-to-match [PTM]) and are eligible to remain on study will have the option to participate in an open-label extension (OLE) to receive ENTO and attend visits per the schedule of assessments (Appendix 1) for an additional 96 weeks. Subjects who complete the study through week 48 and do not wish to participate in the OLE will be required to return to the clinic after completion of study drug for a 30-Day Follow-up Visit. This OLE offers patients who are deriving a benefit from ENTO and have no alternative treatments the opportunity for further therapy for a serious and life threatening disease.

Figure 1. Study Schema



a Following the Baseline/ Day 1 visit subjects will return to the clinic for visits at weeks 2, 4, 8, 12, 16, 20, 24, 30, 36, 42 and 48.
b For subjects receiving tacrolimus, cyclosporine or MMF the blood levels of tacrolimus, cyclosporine or the active metabolite of MMF (MPA) will be monitored at baseline/Day 1 (prior to starting ENTO or PTM), on day 2 or day 3 and day 6 or day 7 after initiating ENTO or PTM, and at the end of week 2 and Week 50 study visit. Beyond these time points, concentration monitoring for tacrolimus, cyclosporine and MPA should follow the treating institutions protocol and therapeutic range for monitoring these medicines
c Subjects who complete the study through week 48 and do not wish to continue to participate in the OLE phase of the trial will be required to return to the clinic 30 days after the completion of study drug for a 30-Day Follow-up visit
d After Week 48, all subjects will have the option to receive OLE for an additional 96 weeks. Subjects will attend visits at Weeks: 50, 52, 56, 60, 64, 72 and every 12 weeks thereafter.

The study consists of the following visits:

- Screening
- Baseline/Day 1
- Treatment Assessments (weeks 2-48 and OLE)
- Unscheduled
- Early Study Drug Discontinuation (ESDD)
- 30 Day Follow-Up

Randomization will be stratified by 1) calcineurin inhibitor or mycophenolate mofetil (MMF) use (no vs. yes) and 2) moderate vs. severe cGVHD as determined by the NIH cGVHD Diagnosis and Staging Criteria (NCDSC) at the Screening.

Stratification Factor	**Stratification Levels**
Calcineurin inhibitor or MMF use	Yes vs. No
Disease severity as determined at Screening by the NIH cGVHD Staging and Diagnosis Criteria (NCSDC) Global Severity Score	Moderate cGVHD: • ≥ 1 organ with score = 2, **or** • ≥ 3 organs with score = 1, **or** • lung score = 1 Severe cGVHD: • any organ score = 3 **or** • lung score ≥ 2

- Treatment Arm 1: ENTO 400 mg twice daily: 2 X 200 mg tablets twice daily, for subjects currently on proton pump inhibitors (PPI)
- Treatment Arm 2: Placebo-to-match (PTM) twice daily

Both treatment arms will be administered orally without food (in a fasted state) for 48 weeks. For subjects **NOT** currently taking a PPI, the dose will be adjusted to ENTO 200 mg twice daily: 1 X 200 mg tablet orally twice daily without food (in a fasted state) for 48 weeks.

1.3. Sample Size and Power

With 40 subjects per treatment group, there is 75% power to detect a 30% improvement in response rate at a 2-sided 0.05 significance level using Fisher's exact test, assuming a placebo response rate of 50%. A total of 100 subjects (50 per treatment group) will be enrolled assuming a 20% dropout rate.

2. TYPE OF PLANNED ANALYSIS

2.1. Data Monitoring Committee Analyses

An external multidisciplinary Data Monitoring Committee (DMC) will review the progress of the study and perform interim reviews of safety and efficacy data for futility in order to protect subject welfare and preserve study integrity. The DMC is to recommend to the sponsor whether the nature, frequency, and severity of adverse effects associated with study treatment warrant the early termination of the study in the best interests of the participants, whether the study should continue as planned, or whether the study should continue with modifications.

The DMC will meet at designated intervals every 5 month after the first 20 subjects have been on study for 8 weeks to review accumulated safety data. The DMC will also review 1 formal interim analysis for futility when approximately 50% of subjects have been on study for 24 weeks or have discontinued the study before 24 weeks, or have achieved a response prior to reaching 24 weeks.

The DMC's role and responsibilities and the scope of analysis to be provided to the DMC are detailed in a mutually agreed upon charter, which defines the DMC membership, meeting logistics, and meeting frequency.

2.2. Interim Analysis

The DMC is primarily responsible for reviewing safety data. Efficacy data will be reviewed only during the formal interim futility analysis after approximately 50% of subjects have been on study for 24 weeks, or have discontinued the study before 24 weeks, or have achieved a response prior to reaching 24 weeks. A non-binding futility rule will be implemented which is based on the predictive probability of success.

The predictive probability of success is defined as the probability of achieving statistical significance (two sided p-value < 0.05) and a $\geq 30\%$ difference in the BORR between the ENTO and placebo group at the study completion given the data observed at the interim.

If the predictive probability of success is $> 30\%$, the study will continue. If the predictive probability of success is $\leq 30\%$, the totality of the data including the DOR, response rate at Week 24, and safety data will be reviewed by the DMC to make a thorough risk-benefit assessment. The VP of Clinical Research and the VP of Biometrics at Gilead may review the interim futility analysis results to make subsequent decisions based on the DMC recommendations.

A 2-sided alpha of 0.001 will be allocated to this interim futility analysis; following the Haybittle-Peto approach, the week 24 analysis alpha level will be 0.05 (2-sided).

2.3. Final Analysis

Due to early termination of the study by sponsor for lack of efficacy based on the interim futility analysis results, the final analysis of the data will be performed after all enrolled subjects have finished their follow-up/early-termination visit after the last dose of study drug, outstanding data queries have been resolved or adjudicated as unresolvable, and data have been cleaned and finalized.

3. GENERAL CONSIDERATIONS FOR DATA ANALYSES

Analysis results will be presented using descriptive statistics. For categorical variables, the number and percentage of subjects in each category will be presented; for continuous variables, the number of subjects (n), mean, standard deviation (SD) or standard error (SE), median, first quartile (Q1), third quartile (Q3), minimum, and maximum will be presented. Time-to-event endpoints will be summarized using median and quartiles of the Kaplan-Meier (KM) estimates.

3.1. Analysis Sets

Analysis sets define which subjects are included in an analysis. The assignment of subjects to analysis sets will be done before the study is unblinded for analysis. A summary of the number and percentage of subjects in each analysis set will be provided by treatment group and in total as part of the subject disposition summary.

3.1.1. Intent-to-Treat (ITT) Analysis Set

The Intent-to-Treat (ITT) Analysis Set includes all subjects who were randomized into the study, and will be analyzed according to treatment randomized. This is the primary analysis set for all efficacy analyses and summaries of demographics and baseline characteristics, unless otherwise specified.

A subset of the ITT Analysis Set will be utilized for the interim futility analysis, which includes subjects who have been on study for 24 weeks or have discontinued the study before 24 weeks, or have achieved a response prior to reaching 24 weeks.

3.1.2. Safety Analysis Set

The Safety Analysis Set includes data from all subjects who received at least 1 dose of study drug, with treatment assignments designated according to actual treatment received. All data collected up to earlier of the last dose of study drug plus 30 days or first dose of OLE will be included in the safety summaries. The Safety Analysis Set will be used for analyses and summaries of exposure of study drug, corticosteroids, prior and concomitant medication, laboratory data, vital signs, and AEs.

3.1.3. Pharmacokinetics (PK) Analysis Set

The Pharmacokinetic (PK) Analysis Set will include all randomized subjects who took at least 1 dose of study drug and have at least 1 non-missing post-dose concentration value reported by the PK laboratory. This is the primary analysis set for all PK analyses.

3.2. Subject Grouping

For analyses based on the ITT Analysis Set, subjects will be grouped according to the treatment to which they were randomized. For analyses based on the Safety Analysis Set, subjects will be grouped according to the randomized treatment except when their actual treatment differs from randomized treatment for the entire treatment duration. In this case, subjects will be grouped based on actual treatment received.

For the PK Analysis Set, subjects will be grouped according to the actual treatment they received.

3.3. Strata and Covariates

Subjects will be randomly assigned to treatment groups via the interactive web response system (IWRS) in a 1:1 ratio using a stratified randomization schedule. Stratification will be based on the following factors:

- calcineurin inhibitor or mycophenolate mofetil (MMF) use (no vs. yes) at Screening
- moderate vs. severe cGVHD as determined by the NIH cGVHD Diagnosis and Staging Criteria (NCDSC) at the Screening

If there are discrepancies in stratification factor values between the IWRS and the clinical database, the baseline values recorded in the clinical database will be used for analyses. Both stratification factor values can be defined based on questionnaire data collected at the screening visit, and concomitant medicine data.

Efficacy endpoints will be evaluated using stratification factors as covariates or stratification variables for analyses.

3.4. Examination of Subject Subgroups

Due to early termination of the study by the sponsor, no pre-specified subgroup analyses for efficacy and safety will be conducted.

3.5. Multiple Comparisons

Due to early termination of the study by the sponsor, no formal statistical comparisons will be done.

3.6. Missing Data and Outliers

3.6.1. Missing Data

Missing data can have an impact upon the interpretation of the trial results. In general, values for missing data will not be imputed unless methods for handling missing data are specified. The handling of missing or incomplete dates is described in Section 7.1.5.2 for AE onset, in Section 6.3 for efficacy, in Section 7.4 for prior and concomitant medications, and in Section 7.1.5.2 for missing last dose date.

Values for missing safety laboratory data will not be imputed; however, a missing Baseline/Day 1 result will be replaced with a screening result, if available. If no pre-treatment safety laboratory value is available, the Baseline/Day 1 value will be assumed to be normal (ie, no grade) in the summary of graded laboratory abnormalities. Values for missing vital signs data will not be

imputed; however, a missing Baseline/Day 1 result will be replaced with a screening result, if available.

- If safety laboratory results for a subject are missing for any reason at a time point, the subject will be excluded from the calculation of summary statistics for that time point. If the subject is missing a pre-dose value, then the subject will be excluded from the calculation of summary statistics for the pre-dose value and the change from pre-dose values.

SAS code to calculate the LSS score and handle missing data are provided in Appendix 6.

3.6.2. Outliers

Outliers will be identified during the data management and data analysis process. All data will be included in the data analysis.

3.7. Data Handling Conventions and Transformations

Where appropriate, safety data for subjects that did not complete the study will be included in summary statistics.

- If a subject received study drug, the subject will be included in the summary of AEs according to the treatment received; otherwise, if the subject is not dosed then they will be excluded from the summary.

Laboratory data that are continuous in nature but are less than the lower limit of quantitation (LLOQ) or above the upper LOQ will be imputed as follows:

- A value that is 1 unit less than the lower LOQ will be used for calculation of descriptive statistics if the data is reported in the form of "<x" (x is considered the LOQ). For example, if the values are reported as <50 and <5.0, then values of 49 and 4.9 will be used for calculation of summary statistics, respectively.

- A value that is 1 unit above the upper LOQ will be used for calculation of descriptive statistics if the data is reported in the form of ">x" (x is considered the LOQ). For example, if the values are reported as >50 and >5.0, then values of 51 and 5.1 will be used for calculation of summary statistics, respectively.

- The LOQ will be used for calculation of descriptive statistics if the data is reported in the form of "≤ x" or "≥ x" (x is considered as the LOQ).

PK concentration values that are below the limit of quantitation (BLQ) will be presented as "BLQ" in the concentration data listing. Values that are BLQ will be treated as 0 at pre-dose time points, and one-half the value of the LLOQ at post-baseline time points, where LLOQ is corrected for the dilution factor (ie, reported LLOQ/dilution factor) for determination of summary and order statistics.

The following conventions will be used for the presentation of summary and order statistics:

- If at least 1 subject has a concentration value of BLQ for the time point, the minimum value will be displayed as "BLQ"
- If more than 25% of the subjects have a concentration value of BLQ for a given time point, the minimum and Q1 values will be displayed as "BLQ"
- If more than 50% of the subjects have a concentration value of BLQ for a given time point, the minimum, Q1, and median values will be displayed as "BLQ"
- If more than 75% of the subjects have a concentration value of BLQ for a given time point, the minimum, Q1, median, and Q3 values will be displayed as "BLQ"
- If all subjects have concentration values of BLQ for a given time point, all order statistics (minimum, Q1, median, Q3, and maximum) will be displayed as "BLQ"

PK parameters that are BLQ will be imputed as one-half LLOQ before log transformation or statistical model fitting.

3.8. Analysis Visit Windows

3.8.1. Definition of Study Day

Study day is the day relative to the randomization date. Study Day 1 is defined as the randomization day.

Study day will be calculated from the date of randomization and defined as:

- Post-randomization: (assessment date) – (date of Study Day 1) + 1
- Prior to randomization: (assessment date) – (date of Study Day 1).

3.8.2. Analysis Visit Windows

Subject visits might not occur on protocol-specified days. Therefore, for the purpose of analysis, observations will be assigned to analysis windows.

The analysis windows are provided in Table 1.

Table 1 Analysis Visit Windows for Double Blind Phase for All Except ECG

Nominal Visit	Analysis Visit	Nominal Day	Lower Limit	Upper Limit
Baseline	Baseline	1	n/a	1
Week 2	Week 2	14	2	21
Week 4	Week 4	28	22	42
Week 8	Week 8	56	43	70
Week 12	Week 12	84	71	98

Nominal Visit	Analysis Visit	Nominal Day	Lower Limit	Upper Limit
Week 16	Week 16	112	99	126
Week 20	Week 20	140	127	154
Week 24	Week 24	168	155	189
Week 30	Week 30	210	190	231
Week 36	Week 36	252	232	273
Week 42	Week 42	294	274	315
Week 48	Week 48	336	316	343

Table 2. Analysis Visit Windows for Double-Blind Phase ECG Only

Nominal Visit	Analysis Visit	Nominal Day	Lower Limit	Upper Limit
Baseline	Baseline	1	n/a	1
Week 24	Week 24	168	2	252
Week 48	Week 48	336	253	420

Vital signs, weight, ECG and safety laboratory data collected after last dose date of any study drug will be summarized up to last dose date plus 30 days, and labeled as "FU-30 Day". For any subjects on ENTO from baseline to Week 48 who decided to enroll into the OLE phase, there wil not be a 30 day follow up visit. An unscheduled visit prior to the first dosing of study drug may be included in the calculation of the baseline value, if applicable. The nominal day of "FU-30 Day" is the last dose date plus 30 days. The lower limit of window for "FU-30 Day" is the upper limit of last on drug visit window plus 1, while the upper limit is Follow-Up Day 30.

Follow-Up Day 1 is the date of the last dose of study drug plus 1.

Nominal Visit	Analysis Visit	Nominal Follow-Up Day	Lower Limit in Follow-Up Day	Upper Limit in Follow-Up Day
FU-30 Day	FU-30 Day	30	1	30

3.8.3. Selection of Data in the Event of Multiple Records in an Analysis Visit Window

Depending on the statistical analysis method, single values may be required for each analysis window. For example, change from baseline by visit usually requires a single value, whereas a time-to-event analysis would not require 1 value per analysis window.

If multiple valid, nonmissing, continuous measurements exist in an analysis window, records will be chosen based on the following rules if a single value is needed:

- In general, the baseline value will be the last nonmissing value on or prior to the first dose date of study drug, unless otherwise specified. If multiple measurements occur on the same day, the last nonmissing value prior to the time of first dose of study drug will be considered as the baseline value. If these multiple measurements occur at the same time or the time is not available, the average of these measurements (for continuous data) will be considered the baseline value.

- For postbaseline visits:

  — The record closest to the nominal day for that visit will be selected.

  — If there are 2 records that are equidistant from the nominal day, the later record will be selected.

  — If there is more than 1 record on the selected day, the average will be taken, unless otherwise specified.

If multiple valid, nonmissing, categorical measurements exist in an analysis window, records will be chosen based on the following rules if a single value is needed:

- For baseline, the last available record on or prior to the date of the first dose of study drug will be selected. If there are multiple records with the same time or no time recorded on the same day, the value with the lowest severity will be selected (eg, normal will be selected over abnormal for safety ECG findings).

- For postbaseline visits:

  — The record closest to the nominal day for that visit will be selected.

  — If there are 2 records that are equidistant from the nominal day, the later record will be selected.

  — If there is more than 1 record on the selected day, the value with the worst severity will be taken (eg, abnormal will be selected over normal for safety ECG findings), unless otherwise specified.

4. SUBJECT DISPOSITION

4.1. Subject Enrollment and Disposition

A summary of subject enrollment will be provided by treatment group for each country, investigator and overall. The summary will present the number and percentage of subjects enrolled. For each column, the denominator for the percentage calculation will be the total number of subjects analyzed for that column.

A similar enrollment table will be provided by randomization stratum. The denominator for the percentage of subjects in the stratum will be the total number of enrolled subjects within that stratum. If there are discrepancies in the value used for stratification assignment between the IWRS and the clinical database, the value collected in the clinical database will be used for the summary. A listing of subjects with discrepancies in the value used for stratification assignment between the IWRS and the clinical database at the time of data finalization will be provided.

A listing of subject enrollment will be provided to describe site, subject number, subject stratification, informed consent date, first screening date, and randomization date.

A summary of subject disposition will be provided by treatment group and overall. This summary will present:

- Number of subjects screened
- Number of subjects randomized (ITT Analysis Set)
- Number and percentage of subjects randomized and treated with study drug (Safety Analysis Set)
- Number and percentage of randomized subjects in the PK Analysis Sets
- Number and percentage of randomized subjects who discontinued study drug, and the reasons for study drug discontinuation
- Number and percentage of randomized subjects who discontinued from the study, and the reasons for study discontinuation

No inferential statistics will be generated.

A listing of subject disposition for all randomized subjects will be generated to describe subject number, randomized study drug (ENTO or placebo) assignment, actual study drug assignment, first dose date, last dose date, duration of study drug treatment, reason for study drug and study discontinuation. A separate listing of reasons for screen failure will be provided by screening ID number in ascending order.

4.2. Extent of Study Drug Exposure and Adherence

Extent of exposure to study drug will be examined by assessing the total duration of exposure to study drug and the level of adherence to the study drug specified in the protocol.

4.2.1. Duration of Exposure to Study Drug

Duration of exposure is defined as (last dose date – first dose date + 1), regardless of temporary interruptions in study drug administration, and will be expressed in weeks (recorded to 1 decimal place, eg, 4.5 weeks).

Duration of exposure to study drug will be summarized using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) and as the number and percentage of subjects exposed for at least 4, 8, 12, 24, 36 and 48 weeks.

Summaries will be provided by treatment group for the Safety Analysis Set. No inferential statistics will be provided.

4.2.2. Adherence to Study Drug

A separate by-subject listing of study drug administration and drug accountability will be provided by subject ID number (in ascending order) and visit (in chronological order).

Adherence will be determined by the reporting of doses taken as documented on the study drug administration eCRF.

Adherence to study drug (%) is defined as (total days on drug) / [Last dose date – First dose date + 1]. Subjects are defined as treatment compliant if they achieved at least 80% adherence. The total number of doses administered will also be summarized.

Descriptive statistics for adherence and total number of doses taken (n, mean, SD, median, Q1, Q3, minimum, and maximum) along with the number and percentage of subjects belonging to adherence categories (eg, $< 80\%$, $\geq 80\%$) will be provided by treatment group for the Safety Analysis Set. Adherence will also be listed.

4.3. Protocol Deviations

Subjects who did not meet the eligibility criteria for study entry, but enrolled in the study will be summarized regardless of whether they were exempted by the sponsor or not. The summary will present the number and percentage of subjects who did not meet at least 1 eligibility criterion and the number of subjects who did not meet specific criteria by treatment group based on all enrolled subjects. A by-subject listing will be provided for those subjects who did not meet at least 1 eligibility (inclusion or exclusion) criterion. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that subjects did not meet and related comments, if collected.

Protocol deviations occurring after subjects entered the study are documented during routine monitoring. The number and percentage of subjects with important protocol deviations by deviation reason will be summarized by treatment group for the ITT Analysis Set. A by-subject listing will be provided for those subjects with any protocol deviation.

5. BASELINE CHARACTERISTICS

5.1. Demographics

Subject demographic variables (ie, age, sex, race, ethnicity and region) will be summarized by treatment group and overall using descriptive statistics for age, and using number and percentage of subjects for sex, race, ethnicity and region. Age at baseline is calculated in years at randomization date was made. The summary of demographic data will be provided for the ITT Analysis Set.

5.2. Other Baseline Characteristics

In addition to the randomization stratification factors listed in Section 3.3, the following demographic and baseline characteristics will be summarized by treatment group for the overall population:

- Age and Age group (≥ 65 years or < 65 years)
- Gender (male or female)
- Race
- Region
- Ethnicity (Hispanic/Latino, not Hispanic/Latino, or not permitted)
- Height
- Weight
- BMI
- Clinician symptom severity score (Appendix 3, Form A, row 1, box 2)
- Health care provider global ratings (Appendix 3, Form A, row 1, box 1)
- Total score of Lee Symptom Scale
- Skin domain score of the Lee Symptom Scale
- Mouth domain score of the Lee Symptom Scale
- Eyes domain score of the Lee Symptom Scale

5.3. Medical History

General medical history data will be collected at screening and listed but not coded.

6. EFFICACY ANALYSES

6.1. Definition of the Primary Efficacy Endpoint

The primary endpoint is BORR by 24 weeks, defined as the proportion of subjects who achieve a complete or partial overall response (CR or PR) as assessed by the NCAA (Appendix 3) within 24 weeks, in the setting of add-on therapy to systemic corticosteroids as part of first-line therapy for cGVHD.

At each assessment time point, organ-specific responses will be determined according to Appendix 2. An overall time point response across all involved organs will then be determined according to Table 3. The best overall response across all time points up to 24 weeks will be used to calculate BORR by 24 weeks.

Table 3. Criteria for time point response determination

Response Category	Response Criteria
Complete response (CR)	Resolution of all cGVHD manifestations in each organ or site as described in the NCAA organ response determination table
Partial response (PR)	Improvement of at least 1 organ or site, without progression in any other organ or site as described in the NCAA organ response determination table.
Stable disease (SD)	Neither sufficient improvement to qualify for a PR nor sufficient decline to qualify for a PD
Progressive disease (PD)* * A flare is defined as acute worsening of signs and symptoms in any organ or site that occurs at least 4 weeks after PR/CR is achieved, which lasts ≤ 4 weeks.	• Worsening of signs and symptoms in any organ or site as described in-the NCAA organ response determination table, taking the baseline as reference. **OR** • Increase in systemic corticosteroid dose to ≥ baseline dose (1 mg/kg/day) **OR** • Initiation of systemic immunosuppressive therapy other than systemic corticosteroids

6.2. Statistical Hypothesis for the Primary Efficacy Endpoint

The primary analysis will test the null hypothesis (H_0) that there is no difference in BORR between ENTO and placebo versus the alternative hypothesis (H_1) that there is a difference. Formally,

$$H_0: R_1 = R_0 \text{ vs. } H_1: R_1 \neq R_0,$$

where R_1 and R_0 are the BORR rate for ENTO and placebo treatment groups, respectively.

6.3. Analysis of the Primary Efficacy Endpoint

The primary analysis for BORR will be performed in the ITT Analysis Set, which includes all subjects who were randomized. Response rate will be summarized by count and percent of subjects with each ordinal response (CR, PR, CR or PR, SD and PD). A stratified Cochran-Mantel-Haenszel (CMH) Chi-square test will be performed to compare BORR (CR + PR) between ENTO and placebo-to-match (PTM) by 24 weeks. The difference in response rates between treatment groups and the corresponding 95% CIs will be presented.

If the entire NIH questionnaire is missing, then the organ and subject level statuses are missing. Missing individual NIH domain scores, for example, skin score, will lead to missing status for skin domain. In such a case, the subject level status will be based on the remaining available domain statuses. This is equivalent to imputing the missing domain status as ND or NA.

6.4. Secondary Efficacy Endpoints

6.4.1. Definition of Secondary Efficacy Endpoints

The key secondary endpoints are:

- Change from baseline in the skin domain of the LSS at 24 weeks
- Change from baseline in the mouth domain of the LSS at 24 weeks
- Change from baseline in the eyes domain of the LSS at 24 weeks
- Change from baseline in the total score of the LSS at 24 Weeks

Other secondary endpoints include:

- Duration of response, defined as the time from the documentation of BORR to the documentation of PD as defined in Table 3. Note that a flare does ***not*** constitute an end to response. A sensitivity analysis for DOR could consider a flare does constitute an end to response.
- Proportion of subjects who ever achieve at least 50% reduction in systemic corticosteroid dose relative to baseline, where % reduction is calculated as (systemic corticosteroid dose at 24 weeks – baseline systemic corticosteroid dose) / baseline systemic corticosteroid dose.
- Proportion of subjects who initiate second-line therapy for cGVHD, defined as receiving any therapy besides systemic corticosteroids or study drug for the treatment of cGVHD. Inhaled and topical steroids are not considered second-line therapy.
- Failure-free survival (FFS), defined as the time from randomization to the earliest of first documentation of systemic therapy change, non-relapse mortality, or recurrent malignancy

- Safety and tolerability of ENTO

6.4.2. Analysis Methods for Secondary Efficacy Endpoints

The key secondary endpoints will be analyzed using a two sample z-test at week 24.

Other secondary endpoints will be analyzed as follows:

DOR and FFS will be analyzed using the Kaplan-Meier (KM) method and stratified logrank test. Kaplan-Meier summary statistics (median, 95% CI for median, and 25th, 75th percentiles) and number and percentage of events will be presented by treatment group. A stratified Cox proportional hazards model will be used to estimate the hazard ratio and corresponding 95% CI. A KM survival plot will also be provided. Data for subjects who do not experience an event will be censored at the date of last visit. DOR will be analyzed only on subjects who responded.

Proportion of subjects with at least 50% reduction in systemic corticosteroid dose relative to baseline and proportion of subjects who initiated a second-line therapy for cGVHD will be analyzed similarly to the primary endpoint.

Line plot of change from baseline in each domain and total of LSS will be done.

6.5. Exploratory Endpoints

6.5.1. List of Exploratory Endpoints

PPD

6.5.2. Analysis Methods for Exploratory Endpoints

PPD

6.6. Changes From Protocol-Specified Efficacy Analyses

PPD


PPD

7. SAFETY ANALYSES

7.1. Adverse Events and Deaths

7.1.1. Adverse Event Dictionary

Clinical and laboratory adverse events (AEs) will be coded using the current version of MedDRA 20.1. System organ class (SOC), high-level group term (HLGT), high-level term (HLT), preferred term (PT), and lower-level term (LLT) will be provided in the AE dataset.

7.1.2. Adverse Event Severity

Adverse events are graded by the investigator as Grade 1, 2, 3, 4, or 5 according to toxicity criteria specified in the protocol. The severity grade of events for which the investigator did not record severity will be categorized as "missing" for tabular summaries and data listings. The missing category will be listed last in summary presentation.

Severity of adverse events will be determined by the investigator as mild, moderate, or severe. The severity of events for which the investigator did not record severity will be categorized as "missing" for tabular summaries and data listings. The missing category will be listed last in summary presentation.

7.1.3. Relationship of Adverse Events to Study Drug

Related AEs are those for which the investigator selected "Related" on the AE case report form (CRF) to the question of "Related to Study Treatment." Relatedness will always default to the investigator's choice, not that of the medical monitor. Events for which the investigator did not record relationship to study drug will be considered related to study drug for summary purposes. However, by-subject data listings will show the relationship as missing.

7.1.4. Serious Adverse Events

Serious adverse events (SAEs) will be identified and captured as SAEs if AEs met the definitions of SAE specified in the study protocol. Serious adverse events captured and stored in the clinical database will be reconciled with the SAE database from the Gilead Drug Safety and Public Health Department before data finalization.

7.1.5. Treatment-Emergent Adverse Events

7.1.5.1. Definition of Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) for the double-blind phase are defined as 1 or both of the following:

- Any AEs with an onset date on or after study drug or placebo start date and no later than earlier of 30 days after permanent discontinuation of study drug or placebo, or first dose date of ENTO in OLE phase, if any.

- Any AEs leading to premature discontinuation of study drug or placebo in the double-blind phase.

Treatment-emergent adverse events (TEAEs) for OLE phase are defined as 1 or both of the following

- Any AEs with an onset date on or after the OLE ENTO first dose date and no later than 30 days after permanent discontinuation of ENTO

- Any AEs leading to premature discontinuation of study drug or placebo in OLE.

7.1.5.2. Incomplete Dates

If the onset date of the AE is incomplete and the AE stop date is not prior to the first dosing date of study drug, then the month and year (or year alone if month is not recorded) of onset determine whether an AE is treatment emergent. The event is considered double blind phase treatment emergent if both of the following 2 criteria are met:

- The AE onset is the same as or after the month and year (or year) of the first double blind dosing date of study drug

- The AE onset date is the same as or before the month and year (or year) of the date corresponding to 30 days after the date of the last dose of study drug or placebo, or date of first study drug dose of OLE if any

The event is considered OLE phase treatment emergent if both of the following 2 criteria are met:

- The AE onset is the same as or after the month and year (or year) of the first OLE dosing date of study drug

- The AE onset date is the same as or before the month and year (or year) of the date corresponding to 30 days after the date of the last dose of study drug

An AE with completely missing onset and stop dates, or with the onset date missing and a stop date later than the first dosing date of study drug, will be considered to be treatment emergent. In addition, an AE with the onset date missing and incomplete stop date with the same or later month and year (or year alone if month is not recorded) as the first dosing date of study drug will be considered treatment emergent.

If the last study drug dosing date is missing, the latest date among the study drug end date, clinical visit date, laboratory sample collection date, weight and vital signs assessment date that occurred during the on-treatment period will be used.

7.1.6. Summaries of Adverse Events and Deaths

Treatment-emergent AEs will be summarized based on the Safety Analysis Set.

A brief summary of AEs will show, by treatment group, the number and percentage of subjects who (1) had any treatment-emergent AE, (2) had any Grade 3 or 4 treatment-emergent AE, (3) had any treatment-emergent treatment-related AE, (4) had any Grade 3 or 4 treatment-emergent treatment-related AE, (5) had any treatment-emergent SAE, (6) had any treatment-emergent treatment-related SAE, (7) had any treatment-emergent AE leading to permanent study drug discontinuation, (8) had any treatment-emergent SAE leading to permanent study drug discontinuation, (9) died during study.

Summaries (n and percentage of subjects) of treatment-emergent AEs (by SOC and PT) will be provided by treatment group using the Safety Analysis Set as follows:

- Summary of AEs
- All treatment-emergent AEs
- All treatment-emergent AEs by preferred term
- All Grade 3 or higher treatment-emergent AEs
- All Grade 3 or higher treatment-emergent AEs by preferred term
- All treatment-emergent treatment-related AEs
- All treatment-emergent treatment-related AEs by preferred term
- All Grade 3 or higher treatment-emergent treatment-related AEs
- All Grade 3 or higher treatment-emergent treatment-related AEs by preferred term
- All treatment-emergent SAEs
- All treatment-emergent SAEs by preferred term
- All treatment-emergent AEs leading to permanent study drug discontinuation
- All treatment-emergent AEs leading to permanent study drug discontinuation by preferred term

- All deaths

Multiple events will be counted once only per subject in each summary. For data presentation, SOC will be ordered alphabetically with PT sorted by decreasing total frequency. For summaries by severity grade, the most severe event will be selected. In addition to the presentation by SOC and PT, summaries of treatment-emergent AEs may be presented by PT only, ordered by decreasing total frequency.

In addition to the by-treatment summaries, data listings will be provided for the following:

- All AEs (with a variable indicating whether the event is treatment-emergent, treatment-related, SAE)
- All Grade 3 and 4 AEs (with a variable indicating whether the event is treatment-emergent, treatment-related, SAE)
- All SAEs (with a variable indicating whether the event is treatment-emergent, treatment-related)
- All Deaths (with a variable indicating whether the event is treatment-emergent, treatment-related, SAE)

All AEs leading to study drug discontinuation (with a variable indicating whether the event is treatment-related, SAE)

Tables of AE will be based on double-blind data. OLE AE will be included in the listings with a column indicating whether the AE was reported during the OLE phase.

Any AE, lab result, or vital sign occurring on or after the date of the first dose of OLE ENTO and no later than 30 days after permanent discontinuation of ENTO is deemed belonging to the OLE phase.

7.1.7. Additional Analysis of Adverse Events

The following categories of AEs of special interest will be summarized (number and percentage) by treatment group.

- Relapse of primary disease/chimerism
- CMV reactivation
- PJP (Pneumocystis Jiroveci Pneumonia)
- EBV reactivation
- Interstitial lung disease

7.2. Laboratory Evaluations

Laboratory data collected during the study will be analyzed and summarized using both quantitative and qualitative methods. Summaries of laboratory data will be provided for the Safety Analysis Set. The analysis will be based on values reported in conventional units. When values are below the LLOQ, they will be listed as such, and the closest imputed value will be used for the purpose of calculating summary statistics. Hemolized test results will not be included in the analysis, but will be listed.

A listing for laboratory test results will be provided by subject ID number and visit in chronological order for hematology and serum chemistry separately. Values falling outside of the relevant reference range will be flagged in the data listings, as appropriate. No inferential statistics will be generated.

7.2.1. Summaries of Numeric Laboratory Results

Descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) will be provided by treatment group for each laboratory test specified in the study protocol as follows:

- Baseline values
- Values at each post-baseline assessment
- Change from baseline at each post-baseline assessment

A baseline laboratory value will be defined as the last measurement obtained on or prior to the date/time of first dose of study drug. Change from baseline to a post baseline visit will be defined as the visit value minus the baseline value. Laboratory test results collected at unscheduled visits will be included for the baseline. The mean, median, Q1, Q3, minimum, and maximum will be displayed to the reported number of digits; SD to the reported number of digits plus 1.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3.

Tables of lab results will be based on double-blind data. OLE lab results will be included in the listings with a column indicating whether the lab result was reported during the OLE phase.

7.2.2. Graded Laboratory Values

The criteria specified in the study protocol will be used to grade laboratory results as Grade 0, mild (Grade 1), moderate (Grade 2), severe (Grade 3), or potentially life threatening (Grade 4). Grade 0 includes all values that do not meet criteria for an abnormality of at least Grade 1. Some laboratory tests have criteria for both increased and decreased levels; analyses for each direction (ie, increased, decreased) will be presented separately.

7.2.2.1. Treatment-Emergent Laboratory Abnormalities

Double blind phase treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from baseline at any post-baseline time point, up to and including the earlier of the date of last dose of study drug plus 30 days, or the first dose of OLE study drug if any.

OLE phase treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from OLE day 1 (first OLE dose date) at any post-OLE day 1 time point, up to and including the date of last dose of OLE study drug plus 30 days.

If the relevant baseline laboratory value is missing, any abnormality of at least Grade 1 observed within the time frame specified above will be considered treatment emergent.

7.2.2.2. Treatment-Emergent Marked Laboratory Abnormalities

Double blind phase treatment-emergent marked laboratory abnormalities are defined as values that increase from baseline by at least 3 toxicity grades at any post-baseline time point, up to and including the earlier of the date of the last dose of study drug plus 30 days, or the first dose of OLE study drug if any.

OLE phase treatment-emergent marked laboratory abnormalities are defined as values that increase from OLE day 1 (first OLE dose date) by at least 3 toxicity grades at any post-OLE day 1 time point, up to and including the date of the last dose of study drug plus 30 days.

If the relevant baseline laboratory value is missing, any Grade 3 or 4 values observed within the timeframe specified above will be considered treatment-emergent marked abnormalities.

7.2.2.3. Summaries of Laboratory Abnormalities

The following summaries (number and percentage of subjects) for treatment-emergent laboratory abnormalities will be provided by lab test and treatment group; subjects will be categorized according to the most severe post-baseline abnormality grade for a given lab test:

- Treatment-emergent laboratory abnormalities
- Treatment-emergent Grade 3 and 4 laboratory abnormalities
- Treatment-emergent marked laboratory abnormalities

For all summaries of laboratory abnormalities, the denominator is the number of subjects in the Safety Analysis Set with non-missing post-baseline values in the given study period.

A listing of all Grade 3 or 4 laboratory abnormalities will be provided by subject ID number and visit in chronological order. A column will indicate if the laboratory abnormality is treatment-emergent. This listing will include all test results that were collected throughout the study for the lab test of interest, with all applicable severity grades displayed.

7.3. Body Weight and Vital Signs

Body weight and vital signs (heart rate, systolic and diastolic blood pressure, body temperature, and respiratory rate) at baseline and each post-baseline assessment, and change from baseline in body weight and vital signs will be summarized for the Safety Analysis Set using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) by treatment group. In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 2. No formal statistical testing is planned. Only double-blind phase data will be summarized in tables. Both double-blind and OLE phase data will be included in listings.

7.4. Prior Medications and Concomitant Medications

Medications collected at screening and during the study will be coded using the current version of the World Health Organization (WHO) Drug dictionary. The medications will be categorized as prior or concomitant using the following definitions:

- Prior medications: any medications that were stopped before the study drug start date
- Concomitant medications: any medications initially taken on or after the study drug start date and up to 30 days after permanent discontinuation of study drug.

7.4.1. Prior Medications

Prior medications will be summarized by Anatomical Therapeutic Chemical (ATC) drug class and preferred name (using number and percent of subjects) by treatment group and overall. A subject reporting the same medication more than once will be counted only once when calculating the number and percentage of subjects who received that medication. The summary will be ordered by descending overall frequency of ATC drug classes and then by preferred names within an ATC drug class. For drugs with the same frequency, sorting will be done alphabetically.

Summaries will be based on the Safety Analysis Set. No inferential statistics will be generated.

For purposes of analysis, any medication with a stop date that is prior to or on the study drug start date will be included in the summary of prior medications. If a partial stop date is entered, then any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the study drug start date will be included in the prior medications summary.

All prior medications (other than per-protocol study drugs) will be provided in a listing sorted by subject ID number and administration date in chronological order.

7.4.2. Concomitant Medications

Use of concomitant medications up to 30 days after permanent study drug discontinuation will be summarized by ATC drug class and preferred name (using number and percent of subjects) by treatment group. A subject reporting the same medication more than once will be counted only

once when calculating the number and percentage of subjects who received that medication. The summary will be ordered by descending active treatment group frequency of ATC medical classes and then by preferred names within an ATC drug class. For drugs with the same frequency, sorting will be done alphabetically.

Summaries will be based on the Safety Analysis Set. No inferential statistics will be generated.

For purposes of analysis, any medication with a stop date that is on or prior to the initial study drug dosing date or a start date that is after the last study drug dosing date plus 30 days will be excluded from a concomitant medication summary. If a partial stop date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the initial study drug dosing date will be excluded from the concomitant medication summary. If a partial start date is entered, then any medication with the month and year (if day is missing) or year (if day and month are missing) after the study drug stop date plus 30 days will be excluded from the concomitant medication summary. Medications with completely missing dates will be included in the concomitant medication summary.

All concomitant medications (other than per-protocol study drugs) will be provided in a listing sorted by subject ID number and administration date in chronological order.

7.5. Electrocardiogram Results

7.5.1. Investigator Electrocardiogram Assessment

Individual data for 12-lead ECG will be listed by subject and summarized by treatment group by incidence of events/abnormalities

7.6. Biomarker Analysis

No biomarker analyses are planned within the scope of this analysis plan. There will be a separate Biomarker Analysis Plan (BAP).

7.7. Changes From Protocol-Specified Safety Analyses

There are no deviations from the protocol-specified safety analyses.

8. PHARMACOKINETIC ANALYSES

The following tables will be provided:

- ENTO pharmacokinetic sampling details by subject including deviations in scheduled and actual draw times and procedures

9. REFERENCES

No citations

10. SOFTWARE

SAS® Software Version 9.3. SAS Institute Inc., Cary, NC, USA.

11. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
24 APR 2018	2.3	Week 14 Analysis deleted	Study discontinuation due to lack of efficacy
24 APR 2018	2.4	Week 48 Analysis deleted	Study discontinuation due to lack of efficacy
24 APR 2018	2.5	Section number changed to 2.3. Added description of final analysis.	Study discontinuation due to lack of efficacy
24 APR 2018	3.1.2	Deleted the definition of Safety Analysis Set for OLE phase	Study discontinuation due to lack of efficacy and very few OLE AE occurrences
24 APR 2018	3.1.3	Deleted the immunosuppressant PK Analysis Set	Study discontinuation due to lack of efficacy
24 APR 2018	3.4	Deleted the specified subgroup analysis	Study discontinuation due to lack of efficacy
24 APR 2018	3.5	Deleted the multiple hypothesis testing	Study discontinuation due to lack of efficacy
24 APR 2018	3.6.1	Deleted the MMRM analysis for LSS. Moved the language of Lab value missing from section 3.7	Study discontinuation due to lack of efficacy
24 APR 2018	3.8.2	Added “for All Except ECG” at the end of title of Table 1	To clarify the windows apply to all endpoints but ECG
24 APR 2018	3.8.2	Deleted Table 3	Study discontinuation due to lack of efficacy and very few OLE AE occurrences
24 APR 2018	3.8.2	Added “FU-30 Day” nominal visit	To summarize the data more precisely
24 APR 2018	4.1	Deleted “Number and percentage of randomized subjects in the PP Analysis Set” and “The reasons for exclusion from the PP Analysis Set will be summarized separately”	Study discontinuation due to lack of efficacy
24 APR 2018	4.2.2	Changed the formula of calculating adherence	The PPI usage is not recorded in the eCRF thus day based formula is used to calculate adherence
24 APR 2018	6.3	Added the missing imputation when the entire NIH questionnaire is missing	To clarify the missing imputation
24 APR 2018	6.4.2	Deleted the MMRM analysis	Study discontinuation due to lack of efficacy
24 APR 2018	6.5.1	Deleted two exploratory endpoints related to flares	Study discontinuation due to lack of efficacy and very few flares

24 APR 2018	6.5.2	Deleted the time to first flare analysis and methods handling multiple occurrences	Study discontinuation due to lack of efficacy and very few flares
24 APR 2018	6.6	Added all new changes to the section	Study discontinuation due to lack of efficacy
24 APR 2018	7.1.5.1	Clarify the definition of treatment emergent	To clarify the definition of treatment emergent
24 APR 2018	7.1.6	Updated the list of summaries of AEs	To provide TFLs for CSR
24 APR 2018	7.1.7	Deleted the language of using SMQ method	Some AEs of interest are not defined based on SMQ
24 APR 2018	7.2.1 7.3	Clarify only double blinded phase data will be present in table. Listing will include both double blinded phase and OLE data	To clarify the inclusion of data and there are very few data in OLE
24 APR 2018	7.5.1	Simplified the analysis for ECG	Study discontinuation due to lack of efficacy
24 APR 2018	7.6	Deleted the biomarker analysis	There will be separate analysis regarding biomarker
24 APR 2018	8	Deleted the table for PK analysis	Study discontinuation due to lack of efficacy
24 APR 2018	Appendix 2	Deleted the sub study schedule of assessment	Study discontinuation due to lack of efficacy
24 APR 2018	Appendix 6 Appendix 7	Deleted MMRM and negative binomial model	Study discontinuation due to lack of efficacy

12. APPENDIX

Appendix 1. Schedule of Assessments

Visit	Screening[a]	Baseline/ Day1[c]	END OF WEEK [b]																Every 12 Weeks	30 Day Follow-up
			2	4	8	12	16	20	24	30	36	42	48/ ESDD[d]	50	52	56	60	64	72	
Visit Window	**-30 days**	**0**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±5**	**±5**	**±5**	**±5**	**±3**	**±3**	**±3**	**±3**	**±3**	**±5**	**±5**
Written Informed Consent	X																			
Medical History	X																			
Smoking Status	X	X		X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	
Complete Physical Exam	X	X				X			X				X							
Symptom-Directed Physical Exam[e]			X	X	X		X	X		X	X	X		X	X	X	X	X	X	X
Serum Pregnancy Test[f]	X																			
Urine Pregnancy test[f]		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Serum FSH[f]	X																			
NIH cGVHD Diagnosis and Staging Criteria	X																			
NIH cGVHD Activity Assessment (NCAA) Form A with spirometry[g]		X		X	X	X	X	X	X	X	X	X	X			X	X	X	X	
NIH cGVHD Activity Assessment (NCAA) Form B[h]		X		X	X	X	X	X	X	X	X	X	X			X	X	X	X	
Lee Symptom Scale (LSS)[h]		X		X	X	X	X	X	X	X	X	X	X			X	X	X	X	

Visit	Screening[a]	Baseline/ Day1[c]	END OF WEEK [b]																Every 12 Weeks	30 Day Follow-up
			2	4	8	12	16	20	24	30	36	42	48/ ESDD[d]	50	52	56	60	64	72	
Visit Window	**-30 days**	**0**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±5**	**±5**	**±5**	**±5**	**±3**	**±3**	**±3**	**±3**	**±3**	**±5**	**±5**
Primary disease assessment per institutional guidelines	X								X				X							
Height	X																			
Vital Signs & Weight[i]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Hematology[j]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Chemistry[k]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Urine Analysis[l]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Metabolic Panel	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
HIV-1, HBV, & HCV Serology[m]	X																			
12 lead ECG-performed supine	X								X				X							
PK[n]			X	X	X	X	X	X	X	X	X	X	X				X		X	
Monitoring of calcineurin inhibitors and MPA [o]		X	X											X						
Biomarkers[p]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
PFT[s]	X								X				X							
Randomize		X																		
Drug Dispensing[q]		X		X	X	X	X	X	X		X		X		X	X	X		X	
Written Informed Consent for OLE													X							

Visit	Screening[a]	Baseline/ Day1[c]	END OF WEEK [b]															Every 12 Weeks	30 Day Follow-up
			2	4	8	12	16	20	24	30	36	42	48/ ESDD[d]	50	52	56	60	64	72	
Visit Window	**-30 days**	**0**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±5**	**±5**	**±5**	**±5**	**±3**	**±3**	**±3**	**±3**	**±3**	**±5**	**±5**
Adverse Events	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant Medications[r]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

1) Evaluations to be completed within 30 days prior to Day 1.
2) All study visits are to be scheduled relative to the Day 1 visit date.
3) Initiation of treatment with the study drug must take place within 24 hours after the Baseline/Day 1 Visit
4) Early Study Drug Discontinuation visit to occur within 72 hours of last dose of study drug
5) Symptom-directed physical exam as needed
6) FSH: For female subject post-menopausal for less than two years, if FSH < 40 mIU/ mL a serum pregnancy test will be required. Women determined to be of child bearing potential will have a serum pregnancy test performed at screening. All subsequent visits will have a urine test, and positive urine pregnancy tests will be confirmed with a serum test. During OLE, urine pregnancy test will be performed every 4 weeks starting with end of visit Week 64.
7) NCAA Form A (Appendix 3)
8)) with spirometry measures will be completed by the Investigator or Sub-I. In the OLE this will be assessed at Weeks 56, 60, 64, 72 and every 12 weeks thereafter. NCAA should also be performed at the time of initiation of second-line therapy.
9) LSS (Appendix 4) and the NCAA Form B (Appendix 3) is to be completed by the subject. The subject is to read the questionnaire by himself/herself and write/mark answers directly onto the questionnaire. In the OLE this will be assessed at Weeks 56, 60, 64, 72 and 96. The LSS and the NCAA Form B should also be administered at the time of initiation of second-line therapy.
10) Vital sign measurements include blood pressure, pulse, respiration rate, weight and temperature
11) Complete blood count (CBC) with differential platelet count
12) Chemistry profile: alkaline phosphatase, AST, ALT, GGT, total bilirubin, direct and indirect bilirubin, total protein, albumin, LDH, CPK, bicarbonate, BUN, calcium, chloride, creatinine, glucose, phosphorous, magnesium, potassium, sodium, uric acid, and amylase and (reflex lipase testing is performed in subjects with total amylase > 1.5 ULN)
13) Urine collection for urinalysis and urine pregnancy test (if applicable)
14) Subjects with positive HCV antibody and without detectable HCV RNA are permitted to enroll
15) Randomly timed PK sampling. Sites should enter the time of the last dose of study medication in the eCRF for sample reconciliation. PK samples will be drawn at Week 60 and every 12 weeks in the OLE.
16) For subjects receiving tacrolimus, cyclosporine or MMF the blood levels of tacrolimus, cyclosporine or the active metabolite of MMF (MPA) will be monitored at baseline/Day 1 (prior to starting ENTO or PTM), on day 2 or day 3 and day 6 or day 7 after initiating ENTO or PTM, and at the end of Week 2 and Week 50 study visit. Beyond these time points, concentration monitoring for tacrolimus, cyclosporine and MPA should follow the treating institutions protocol and therapeutic range for monitoring these medicines
17) Assessments may include but are not limited to: cytokines, cell subset analysis, pathway markers, and/or other B-cell activating factor, B-cell function, immunoglobulin levels, and/or antigen-specific B cell response. Timepoints may differ depending on the specific biomarker and will be outlined in the Central Lab Manual.
18) At visit 2, 30, 42 and (OLE visit 50) drug accountability only, study drug will not be dispensed at these visits
19) Will include the following: systemic corticosteroid dose, other immunosuppressive, antimicrobials, PPI

19) PFTs will include spirometry with bronchodilator, and diffusing capacity of the lungs for carbon monoxide (DLCO). PFTs will be done at ESDD, if ESDD visit greater than 12 weeks from Baseline. Any PFT done within 30 days prior to randomization can be used in lieu of Screening PFT.

Appendix 2. For Baseline + Subsequent Visits: Response Determination for the NIH cGVHD Activity Assessment (NCAA)

Organ	Complete Response	Partial Response	Progression
Skin	NIH Skin Score 0 after previous involvement	Decrease in NIH Skin Score by 1 or more points	Increase in NIH Skin Score by 1 or more points, except 0 to 1
Eyes	NIH Eye Score 0 after previous involvement	Decrease in NIH Eye Score by 1 or more points	Increase in NIH Eye Score by 1 or more points, except 0 to 1
Mouth	NIH Modified OMRS 0 after previous involvement	Decrease in NIH Modified OMRS of 2 or more points	Increase in NIH Modified OMRS of 2 or more points
Esophagus	NIH Esophagus Score 0 after previous involvement	Decrease in NIH Esophagus Score by 1 or more points	Increase in NIH Esophagus Score by 1 or more points, except 0 to 1
Upper GI	NIH Upper GI Score 0 after previous involvement	Decrease in NIH Upper GI Score by 1 or more points	Increase in NIH Upper GI Score by 1 or more points, except 0 to 1
Lower GI	NIH Lower GI Score 0 after previous involvement	Decrease in NIH Lower GI Score by 1 or more points	Increase in NIH Lower GI Score by 1 or more points, except from 0 to 1
Liver	Normal ALT, alkaline phosphatase, and Total bilirubin after previous elevation of 1 or more	Decrease by 50%	Increase by 2 × ULN
Lungs	- Normal %FEV1 after previous involvement - If PFTs not available, NIH Lung Symptom Score 0 after previous involvement	- Increase by 10% predicted absolute value of %FEV1 - If PFTs not available, decrease in NIH Lung Symptom Score by 1 or more points	- Decrease by 10% predicted absolute value of %FEV1 - If PFTs not available, increase in NIH Lung Symptom Score by 1 or more points, except 0 to 1
Joints and fascia	Both NIH Joint and Fascia Score 0 and P-ROM score 25 after previous involvement by at least 1 measure	Decrease in NIH Joint and Fascia Score by 1 or more points or increase in P-ROM score by 1 point for any site	Increase in NIH Joint and Fascia Score by 1 or more points or decrease in P-ROM score by 1 point for any site
Global	Clinician overall severity score 0	Clinician overall severity score decreases by 2 or more points on a 0-10 scale	Clinician overall severity score increases by 2 or more points on a 0-10 scale

ULN indicates upper limit of normal.

Appendix 3. For Baseline + Subsequent Visits: Chronic GVHD Activity Assessment- Clinician

FORM A

Current Patient Weight: _____ Today's Date: _____ MR#/Name: _____

CHRONIC GVHD ACTIVITY ASSESSMENT- CLINICIAN

Health Care Provider Global Ratings:

0=none
1= mild
2=moderate
3=severe

Where would you rate the severity of this patient's chronic GvHD symptoms on the following scale, where 0 is cGVHD symptoms that are not at all severe and 10 is the most severe cGVHD symptoms possible:

0 1 2 3 4 5 6 7 8 9 10
cGvHD symptoms not at all severe — **Most severe cGVHD symptoms possible**

Since the last assessment would you say that this patient's cGvHD is
+3= Very much better
+2= Moderately better
+1= A little better
0= About the same
-1=A little worse
-2=Moderately worse
-3=Very much worse

Mouth

Erythema	None	0	Mild erythema or moderate erythema (<25%)	1	Moderate (≥25%) or Severe erythema (<25%)	2	Severe erythema (≥25%)	3
Lichenoid	None	0	Lichen-like changes (<25%)	1	Lichen-like changes (25-50%)	2	Lichen-like changes (>50%)	3
Ulcers	None	0			Ulcers involving (≤20%)	3	Severe ulcerations (>20%)	6
			Total score for all mucosal changes					

Gastrointestinal-Esophageal
- Dysphagia OR Odynophagia

0= no esophageal symptoms
1=Occasional dysphagia or odynophagia with solid food or pills during the past week
2=Intermittent dysphagia or odynophagia with solid foods or pills, but not for liquids or soft foods, during the past week
3=Dysphagia or odynophagia for almost all oral intake, on almost every day of the past week

Gastrointestinal-Upper GI
- Early satiety OR
- Anorexia OR
- Nausea & Vomiting

0= no symptoms
1=mild, occasional symptoms, with little reduction in oral intake during the past week
2=moderate, intermittent symptoms, with some reduction in oral intake during the past week
3=more severe or persistent symptoms throughout the day, with marked reduction in oral intake, on almost every day of the past week

Gastrointestinal-Lower GI
- Diarrhea

0= no loose or liquid stools during the past week
1= occasional loose or liquid stools, on some days during the past week
*2=intermittent loose or liquid stools throughout the day, on almost every day of the past week, **without requiring** intervention to prevent or correct volume depletion*
*3=voluminous diarrhea on almost every day of the past week, **requiring** intervention to prevent or correct volume depletion*

Lungs (Liters and % predicted) • Bronchiolitis Obliterans	FEV1	FVC	Single Breath DLCO (adjusted for hemoglobin)		TLC	RV
Liver Values	Total serum bilirubin mg/dL	ULN mg/dL	ALT U/L	ULN U/L	Alkaline Phosphatase U/L	ULN U/L
Baseline Values	Total Distance Walked in 2 or 6 Mins: ❑ 2 min ❑ 6 min		Karnofsky or Lansky	Platelet Count K/uL	Total WBC K/uL	Eosinophils %
	☐ Abnormality present but explained entirely by non-GVHD documented cause (specify site/alternate cause):_____ ☐ Abnormality present but explained entirely by non-GVHD documented cause (specify site/alternate cause):_____ ☐ Abnormality present but explained entirely by non-GVHD documented cause (specify site/alternate cause):_____					

CHRONIC GVHD ACTIVITY ASSESSMENT- CLINICIAN (FORM A)

	SCORE 0	SCORE 1	SCORE 2	SCORE 3
SKIN *GVHD features to be scored by BSA:* **Check all that apply:** ☐ Maculopapular rash / erythema ☐ Lichen planus-like features ☐ Sclerotic features ☐ Papulosquamous lesions or ichthyosis ☐ Keratosis pilaris-like	☐ No BSA involved	☐ 1-18% BSA	☐ 19-50% BSA	☐ >50% BSA
☐ Abnormality present but explained entirely by non-GVHD documented cause (specify):_____				
SKIN FEATURES SCORE:	☐ No sclerotic features		☐ Superficial sclerotic features "not hidebound" (able to pinch)	**Check all that apply:** ☐ Deep sclerotic features ☐ "Hidebound" (unable to pinch) ☐ Impaired mobility ☐ Ulceration

If skin features score = 3, BSA% of non-moveable sclerosis/fasciitis _____

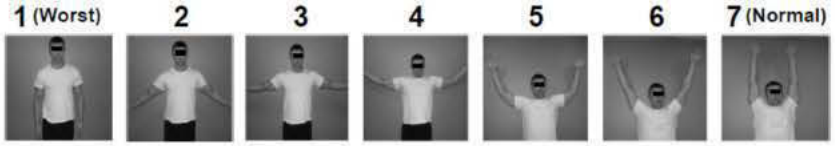
How would you rate the severity of this patient's skin and/or joint tightening on the following scale, where 0 is not at all severe and 10 is the most severe symptoms possible:

0 1 2 3 4 5 6 7 8 9 10

Symptoms not at all severe — Most severe symptoms possible

	SCORE 0	SCORE 1	SCORE 2	SCORE 3
EYES	☐ No symptoms symptoms	☐ Mild dry eye symptoms not affecting ADL (requirement of lubricant eye drops ≤ 3 x per day)	☐ Moderate dry eye symptoms partially affecting ADL (requiring lubricant eye drops > 3 x per day or punctal plugs), **WITHOUT** new vision impairment due to KCS	☐ Severe dry eye symptoms significantly affecting ADL (special eyeware to relieve pain) **OR** unable to work because of ocular symptoms **OR** loss of vision due to KCS
☐ Abnormality present but explained entirely by non-GVHD documented cause (specify):_____				
LUNGS	☐ No symptoms	☐ Mild symptoms (shortness of breath after climbing one flight of steps)	☐ Moderate symptoms (shortness of breath after walking on flat ground)	☐ Severe symptoms (shortness of breath at rest; requiring 0_2)
☐ Abnormality present but explained entirely by non-GVHD documented cause (specify):_____				
JOINTS AND FASCIA	☐ No symptoms	☐ Mild tightness of arms or legs, normal or mild decreased range of motion (ROM) AND not affecting ADL	☐ Tightness of arms or legs OR joint contractures, erythema thought due to fasciitis, moderate decrease ROM AND mild to moderate limitation of ADL	☐ Contractures WITH significant decrease of ROM AND significant limitation of ADL (unable to tie shoes, button shirts, dress self etc.)
☐ Abnormality present but explained entirely by non-GVHD documented cause (specify):_____				

Shoulder



❑ Not done

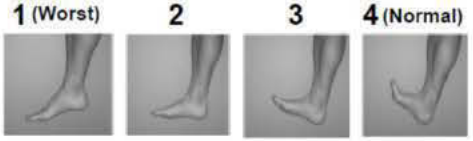
Elbow

❑ Not done

Wrist/finger

❑ Not done

Ankle



❑ Not done

FORM B

TODAY'S DATE: _____ MR#/NAME: _____

CHRONIC GVHD ACTIVITY ASSESSMENT-PATIENT SELF REPORT

Symptoms **Please rate how severe the following symptoms have been in the last seven days. Please fill in the circle below from 0 (symptom has not been present) to 10 (the symptom was as bad as you can imagine it could be) for each item.**	**Not Present** 0	1	2	3	4	5	6	7	8	9	10 **As Bad As You Can Imagine**
Your **skin itching** at its WORST?	O	O	O	O	O	O	O	O	O	O	O
Your **skin and/or joint tightening** at their WORST?	O	O	O	O	O	O	O	O	O	O	O
Your **mouth sensitivity** at its WORST?	O	O	O	O	O	O	O	O	O	O	O
Your **genital discomfort** at its WORST? (Women – vagina, vulva, or labia) (Men – penis)	O	O	O	O	O	O	O	O	O	O	O
Eyes	What is your main complaint with regard to your eyes?										
	Please rate how severe this symptom is, from 0 (not at all severe) to 10 (most severe):	0	1	2	3	4	5	6	7	8	9 10

Patient Global Ratings:

1. Overall, do you think that your chronic graft versus host disease is mild, moderate or severe?
1= mild
2=moderate
3=severe

2. Please circle the number indicating how severe your chronic graft versus host disease symptoms are, where 0 is cGvHD symptoms that are not at all severe and 10 is the most severe chronic GvHD symptoms possible.

0 1 2 3 4 5 6 7 8 9 10

cGvHD symptoms not at all severe — Most severe cGvHD symptoms possible

3. Compared to a month ago, overall would you say that your cGvHD symptoms are:

+3= Very much better
+2= Moderately better
+1=A little better
0= About the same
-1=A little worse
-2=Moderately worse
-3=Very much worse

Appendix 4. For Baseline + Subsequent Visits: Lee Symptom Scale

By circling one (1) number per line, please indicate how much you have been bothered by the following problems in the past 7 days:

SKIN:	Not at all	Slightly	Moderately	Quite a bit	Extremely
1. Abnormal skin color.....	0	1	2	3	4
2. Rashes.....	0	1	2	3	4
3. Thickened skin.....	0	1	2	3	4
4. Sores on skin.....	0	1	2	3	4
5. Itchy skin.....	0	1	2	3	4
EYES AND MOUTH:	**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
6. Dry eyes.....	0	1	2	3	4
7. Need to use eye drops frequently..	0	1	2	3	4
8. Difficulty seeing clearly.....	0	1	2	3	4
9. Need to avoid certain foods due to mouth pain.....	0	1	2	3	4
10. Ulcers in mouth.....	0	1	2	3	4
11. Receiving nutrition from an intravenous line or feeding tube....	0	1	2	3	4
BREATHING:	**Not at all**	**Slightly**	**Moderately**	**Quite a bit**	**Extremely**
12. Frequent cough.....	0	1	2	3	4
13. Colored sputum.....	0	1	2	3	4
14. Shortness of breath with exercise..	0	1	2	3	4
15. Shortness of breath at rest.....	0	1	2	3	4
16. Need to use oxygen.....	0	1	2	3	4

EATING AND DIGESTION:	Not at all	Slightly	Moderately	Quite a bit	Extremely
17. Difficulty swallowing solid foods....	0	1	2	3	4
18. Difficulty swallowing liquids.....	0	1	2	3	4
19. Vomiting.....	0	1	2	3	4
20. Weight loss.....	0	1	2	3	4

MUSCLES AND JOINTS:	Not at all	Slightly	Moderately	Quite a bit	Extremely
21. Joint and muscle aches.....	0	1	2	3	4
22. Limited joint movement.....	0	1	2	3	4
23. Muscle cramps.....	0	1	2	3	4
24. Weak muscles.....	0	1	2	3	4

ENERGY:	Not at all	Slightly	Moderately	Quite a bit	Extremely
25. Loss of energy.....	0	1	2	3	4
26. Need to sleep more/take naps.....	0	1	2	3	4
27. Fevers.....	0	1	2	3	4

MENTAL AND EMOTIONAL:	Not at all	Slightly	Moderately	Quite a bit	Extremely
28. Depression.....	0	1	2	3	4
29. Anxiety.....	0	1	2	3	4
30. Difficulty sleeping.....	0	1	2	3	4

Summary of the Lee Symptom Scale

Which 2 of the following organs/systems have you found to affect you the most in the last seven days:

1) Skin

2) Eyes

3) Mouth

4) Breathing

5) Eating and Digestion

6) Muscles and Joints

7) Energy

8) Mental and Emotional

Appendix 5. Predictive Probability of Success Calculation

Notation

Let

- n_1=number of patients in treatment group
- n_0=number of patients in control group
- n_{i1}=number of patients who stayed in trial at least 24 weeks or who stayed less than 24 weeks but responded already at interim in treatment group
- n_{i0}=number of patients who stayed in trial at least 24 weeks or who stayed less than 24 weeks but responded already at interim in control group
- n_{i11}=number of responders at interim in treatment group
- n_{i01}=number of responders at interim in control group
- x_1=number of new responders after interim and at week 24 analysis in treatment group
- x_0=number of new responders after interim and at week 24 analysis in control group

Predictive probability of success

Predictive probability of success is defined as the probability of obtaining a statistically significant 30% difference favorable result at the week 24 analysis (at 2-sided alpha level α_D), accounting for uncertainty about the true treatment effect given the data observed at the interim, and assuming that future data is subject to sampling variation given any possible true effect.

Predictive probability of success can be calculated as:

$$\sum_{x_1=0}^{n_1-n_{i1}} \sum_{x_0=0}^{n_0-n_{i0}} \left(\frac{n_{i11}+x_1}{n_1} > \left(\frac{n_{i01}+x_0}{n_0} + 0.3\right)\right) \{pvalue(x_1, x_0) < 0.05\} f(n_1 - n_{i1}, x_1, n_{i11}+1, n_{i1}-n_{i11}+1)\, f(n_1 - n_{i1}, x_1, n_{i11}+1, n_{i1}-n_{i11}+1)$$

where function *f* is the probability density function of Beta-Binomial Distribution

$f(n, x, a, b) = \binom{n}{x} \frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ where Γ is gamma function.

For Programming

```
/*************************************************************************/
/* Predictive probability of success Macro:
*/
/* Calculates the probability of statistical significance at final      */
/* analysis given data observed at interim                              */
/*                                                                      */
/* Input Parameters:                                                    */
/* n1 = sample size in treatment group                                  */
/* n0 = sample size in control group                                    */
/* ni1 = sample size stayed at least 24 weeks or less than 24 weeks but */
/*       responded at interim in treatment group                        */
/* ni0 = sample size stayed at least 24 weeks or less than 24 weeks but */
/*       responded at interim in control group                          */
/* ni11 = responders at interim in treatment group                      */
/* ni01 = responders at interim in control group                        */
/* alpha = 2-sided alpha level at week 24 analysis                       */
/*************************************************************************/



%macro PP_calc(n1,n0,ni1,ni0,ni11,ni01,alpha);

proc iml;
n1=&n1; /*sample size in treatment group*/
n0=&n0; /*sample size in control group*/
ni1=&ni1; /*sample size stayed at least 24 weeks or less than 24 weeks but
respsonded at interim in treatment group*/
ni0=&ni0; /*sample size stayed at least 24 weeks or less than 24 weeks but
respsonded at interim in control group*/
ni11=&ni11; /*responders at interim in treatment group*/
ni01=&ni01; /*responders at interim in control group*/

c1=(n1-ni1); /*rest patients who are potential to have week 24 responses in
treatment group*/
a1=ni11+1;
b1=ni1-ni11+1;
c0=(n0-ni0); /*rest patients who are potential to have week 24 responses in
treatment group*/
a0=ni01+1;
b0=ni0-ni01+1;

pm=j(c1+1,c0+1,0);
fd=j(4*(c1+1)*(c0+1),4,0); /*data output for fisher's exact test*/
do x1=0 to c1;
   p1=comb(c1,x1)*beta(x1+a1,c1-x1+b1)/beta(a1,b1); /*the probability to
observe x1 more responders in treatment group*/
   do x0=0 to c0;
      pass=(x1*(c0+1)+x0)*4;

      p0=comb(c0,x0)*beta(x0+a0,c0-x0+b0)/beta(a0,b0); /*the probability to
observe x0 more responders in control group*/
```

```
        pm=p1*p0; /*the probability to observe x1 more responders in
treatment group and x0 more responders in control group*/
        fd[(pass+1):(pass+4),2]=repeat(pm,4,1);

      fd[(pass+1),1]=ni11+x1; /*count of treatment responses at final*/
        fd[(pass+2),1]=n1-ni11-x1; /*count of treatment non-responses at
final*/
        fd[(pass+3),1]=ni01+x0; /*count of control responses at final*/
        fd[(pass+4),1]=n0-ni01-x0; /*count of control non-responses at
final*/

        su=((ni11+x1)/n1>(ni01+x0)/n0+0.3); /*marker if treatment is superial
30%*/
      fd[(pass+1):(pass+4),3]=repeat(su,4,1);

        rn=x1*(c0+1)+x0+1; /*replicate number*/
        fd[(pass+1):(pass+4),4]=repeat(rn,4,1);
   end;
end;

indicator=repeat(t({1 1 0 0, 1 0 1 0}), (c1+1)*(c0+1), 1);

fm=fd||indicator;
varNames={"count" "pm" "su" "rep" "group" "response"};
create mydata from fm [colname=varNames];
append from fm;
close mydata;

quit;

proc sort data=mydata;
by rep;
run;
proc freq data=mydata;
by rep;
tables group*response/exact cmh;
weight count;
ods output FishersExact=fx; /*output fisher's exact test result*/
run;
data fx1;
set fx;
if Name1='XP2_FISH';
keep rep Nvalue1;
run;

proc sort data=fx1;
by rep;
run;

data mydata1;
merge mydata fx1;
by rep;
if first.rep;
fp=(Nvalue1<&alpha)*pm*su; /*if the fisher's exact test is significant and
treatment is superial*/
```

```
drop count group response;
run;

proc means data=mydata1 sum; /*add the probability together*/
var fp;
run;

%mend;

/* Sample macro call */
%PP_calc(n1=50,n0=50,ni1=25,ni0=25,ni11=20,ni01=10,alpha=0.05);
```

Appendix 6. Lee Score Symptom SAS Code and Note

For Programming

```
data newdata;
  set olddata;
*************************************************************************
**************************
*** Lee chronic GVHD symptom scales
*************************************************************************
**************************;
  array oldsx{30} sx1-sx30;
  array newsx{30}
          b3a b3b b3c b3d b3e b3f b3g b3h b3i b3j b3k b3l b3m b3n b3o
          b3p b3q b3r b3s b3t b3u b3v b3w b3x b3y b3z b3aa b3bb b3cc
b3dd;
  do i = 1 to 30;
     newsx{i}=oldsx{i};
     if newsx{i} not in (0,1,2,3,4) then newsx{i}=.;
  end;
  if nmiss(b3a,b3b,b3c,b3d,b3e) le 2 then
sx_skin=mean(b3a,b3b,b3c,b3d,b3e)*25;
  if nmiss(b3n,b3u,b3v,b3w,b3x,b3y,b3z) le 3 then
sx_energy=mean(b3n,b3u,b3v,b3w,b3x,b3y,b3z)*25;
  if nmiss(b3l,b3m,b3o,b3p,b3aa) le 2 then
sx lung=mean(b3l,b3m,b3o,b3p,b3aa)*25;
  if nmiss(b3f,b3g,b3h) le 1 then sx_eye=mean(b3f,b3g,b3h)*25;
  if nmiss(b3k,b3q,b3r,b3s,b3t) le 2 then
sx_nutrition=mean(b3k,b3q,b3r,b3s,b3t)*25;
  if nmiss(b3i,b3j) le 1 then sx mouth=mean(b3i,b3j)*25;
  if nmiss(b3bb,b3cc,b3dd) le 1 then sx_psych=mean(b3bb,b3cc,b3dd)*25;
  if
nmiss(sx_skin,sx_energy,sx_lung,sx_eye,sx_nutrition,sx_mouth,sx_psych)
le 3
     then
sx_sum=mean(sx_skin,sx_energy,sx_lung,sx_eye,sx_nutrition,sx_mouth,sx_
psych);
  label
    sx_skin = "Lee symptom skin scale (0-100)"
    sx_energy = "Lee symptom energy scale (0-100)"
    sx_lung = "Lee symptom lung scale (0-100)"
    sx_eye = "Lee symptom eye scale (0-100)"
    sx_nutrition = "Lee symptom nutrition scale (0-100)"
    sx_psych = "Lee symptom psychological scale (0-100)"
    sx_mouth = "Lee symptom mouth scale (0-100)"
    sx_sum = "Lee symptom overall summary scale (0-100)";
  drop i
       b3a b3b b3c b3d b3e b3f b3g b3h b3i b3j b3k b3l b3m b3n b3o
       b3p b3q b3r b3s b3t b3u b3v b3w b3x b3y b3z b3aa b3bb b3cc
b3dd;
```

```
run;

proc print data=newdata noobs;
run;
```

Code Translation:

Define q1 as the value for question 1, so does q2, q3,..., q30.

If value of q1-q30 is not 0,1,2,3,4 then the value is set to be missing.

If number of missingness in q1,q2,q3,q4,q5 <=2, then skin scale=mean(q1,q2,q3,q4,q5)*25. Otherwise missing.
Here mean(q1,q2,q3,q4,q5) is defined as sum of non-missing values divided by the number of non-missing values.
Specifically, if nothing is missing, then mean(q1,q2,q3,q4,q5)=(q1+q2+q3+q4+q5)/5
if q1 is missing, q2-q5 are non-missing, then mean(q1,q2,q3,q4,q5)=(q2+q3+q4+q5)/4
if q1 and q2 are missing, then mean(q1,q2,q3,q4,q5)=(q3+q4+q5)/3

If number of missingness in q14,q21,q22,q23,q24,q25,q26 <=3 then energy scale=mean(q14,q21,q22,q23,q24,q25,q26)*25. Otherwise missing.
If number of missingness in q12,q13,q15,q16,q27 <=2 then lung scale=mean(q12,q13,q15,q16,q27)*25. Otherwise missing.
If number of missingness in q6,q7,q8 <=1 then eye scale=mean(q6,q7,q8)*25. Otherwise eye scale is missing.
If number of missingness in q11,q17,q18,q19,q20 <=2 then nutrition scale=mean(q11,q17,q18,q19,q20)*25. Otherwise missing.
If number of missingness in q9, q10 <=1 then mouth scale=mean(q9,q10)*25. Otherwise missing.
If number of missingness in q28,q29,q30 <=1 then psychological scale=mean(q28,q29,q30)*25. Otherwise missing.
If number of missingness in skin, energy, lung, eye, nutrition, mouth and psychological <=3 then overall summary scale=mean(skin, energy, lung, eye, nutrition, mouth, psychological scale), otherwise missing.

The range of skin, energy, lung, eye, nutrition, mouth, psychological and overall scale is 0-100.