Statistical Analysis Plan

A Randomized, Double-Blind Phase 3 Study of Ibrutinib in Combination With Corticosteroids versus Placebo in Combination With Corticosteroids in Subjects with New Onset Chronic Graft Versus Host Disease (cGVHD)

Protocol PCYC-1140-IM

Version: 1.0

Version Date: 28 January 2020

Confidentiality Statement

Statistical Analysis Plan Approval

Protocol Number:	PCYC-1140-IM
Protocol Title:	A Randomized, Double-Blind Phase 3 Study of Ibrutinib in Combination With Corticosteroids versus Placebo in Combination With Corticosteroids in Subjects with New Onset Chronic Graft Versus Host Disease (cGVHD)
SAP Version:	1.0
Date:	28 January 2020

By signing below, all parties accept that the analysis methods and data presentations are acceptable and that this document is final.

DocuSigned by:

[redacted]

[redacted]

Date (ddMMMyyyy)

DocuSigned by:

[redacted]

[redacted]

Date (ddMMMyyyy)

DocuSigned by:

[redacted]

[redacted]

Date (ddMMMyyyy)

DocuSigned by:

[redacted]

[redacted]

Date (ddMMMyyyy)

DocuSigned by:

[redacted]

[redacted]

Date (ddMMMyyyy)

DocuSigned by:

[redacted]

[redacted]

Date (ddMMMyyyy)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	adverse event
ATC	anatomical therapeutic chemical
cGVHD	chronic graft versus host disease
CI	confidence interval
CIF	Cumulative incidence function
CR	complete response
CTCAE v. 4.03	Common Terminology Criteria for Adverse Events version 4.03
CYP	cytochrome P450
DOR	duration of response
DMC	data monitoring committee
IST	Immunosuppressant treatment
ITT	Intent-to-treat
mITT	Modified intent-to-treat
IWRS	interactive web response system
GVHD	graft versus host disease
MedDRA	Medical Dictionary for Regulatory Activities
NIH	National Institutes of Health
PCYC	Pharmacyclics, LLC
OS	overall survival
PD	progressive disease
PK	Pharmacokinetics
PR	partial response
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SI	standard international
TEAE	treatment emergent adverse event
ULN	upper limit of normal
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) is to lay out key elements including definition and statistical methods for analysis of data in evaluation of efficacy and safety for the PCYC-1140-IM study. The SAP specifications including endpoint definitions, order of testing, and methods to be used to analyze the endpoints, supersedes information included in the protocol.

Analyses of biomarker and pharmacokinetics (PK) data are addressed in separate documents.

1.1. Study Design

This is a Phase 3, multicenter, international, randomized, double-blind study of oral ibrutinib in combination with prednisone versus placebo in combination with prednisone in subjects with new onset chronic graft versus host disease (cGVHD).

Approximately 186 subjects with newly diagnosed moderate or severe cGVHD, as defined by the 2014 NIH Consensus Development Project Criteria, were planned to be enrolled. The study enrolled 193 subjects who were randomized in a 1:1 ratio to receive either of the study treatments ibrutinib in combination with prednisone (Arm A) or placebo in combination with prednisone (Arm B).
The randomization between arms was stratified according to:

- Age group (12 to <22 years old vs. ≥22 years old)
- NIH Global Severity grade (moderate vs. severe)
- Ongoing use of systemic immunosuppressants that were initiated for either treatment of or prophylaxis for acute GVHD (yes vs. no)

Treatment with the study drug (ibrutinib/placebo) was administered continuously until unacceptable toxicity, progression of the underlying disease, death, or the start of a new systemic treatment for cGVHD. Corticosteroid therapy was tapered as per a standard taper regimen with the goal of reducing exposure to high-moderate dose corticosteroids as quickly as possible according to clinical severity of cGVHD.
Response was assessed according to the NIH Consensus Development Project Criteria (2014).

An independent data monitoring committee (DMC) was established to monitor and review safety and other relevant data on an ongoing basis to ensure the safety of the subjects enrolled in this study and to review results from an interim futility analysis based on data from the first 50 subjects enrolled in the study.

1.2. Endpoints

The primary and secondary endpoints that will be analyzed are listed in this section. Changes from what is outlined in the protocol are described in section 6.

1.2.1. Primary Endpoint

- Response rate at 48 weeks

1.2.2. Secondary Endpoints

- Time to withdrawal of all corticosteroids for treatment of cGVHD
- Time to withdrawal of all immunosuppressants, including corticosteroids for treatment of cGVHD (with the exception of ibrutinib/placebo)
- Response rate at 24 weeks
- Improvement in Lee cGVHD Symptom Scale
- Proportion of subjects who achieve reduction of corticosteroids dose level less than 0.15 mg/kg/d at 24 weeks sustained for at least 30 days
- Overall survival (OS)
- Duration of response (DOR)

Safety

- Safety and tolerability of ibrutinib in combination with prednisone compared to prednisone in combination with placebo.
- Differences in corticosteroids -related morbidities (e.g., hyperglycemia, hypertension)

1.2.3. Exploratory Endpoints

- Improvement in SF-36 patient reported outcome
- Improvement in Karnofsky performance scale score

1.3. Statistical Hypotheses

The primary hypothesis of this study is that the experimental treatment ibrutinib with prednisone compared with placebo with prednisone will significantly improve response rate π at 48 Weeks, as determined by NIH Consensus Development Project Criteria, in subjects with new onset moderate to severe cGVHD.

The statistical hypotheses are as follows:

H_0: The response rate of the experimental treatment group, $\pi_I(t)$, and the control group, $\pi_C(t)$, are equal at time t = 48 Weeks:

H_0: $\pi_I(t) = \pi_C(t)$, for t = 48 Weeks

versus

H_1: The response rate of experimental treatment group, $\pi_I(t)$, is different from that of the control group, $\pi_C(t)$, at time t =48 Weeks:

H_1: $\pi_I(t) \neq \pi_C(t)$, for t = 48 Weeks

These hypotheses will be tested using the chi-square test.

1.4. Sample Size Determination

The study was powered to test the primary endpoint of the study.

The sample size was computed assuming a response rate of 30% at 48 weeks for Arm B (placebo + prednisone). A sample size of 186 randomized subjects provides at least 80% power to detect a 20% difference in the response rates at 48 weeks between the 2 treatment arms at an alpha level of 5% (2-sided).

No replacement of subjects was implemented.

The sample size and power calculations were produced using the software package, East (Cytel Software Corp, Cambridge, MA).

1.5. Planned Analyses

1.5.1. Interim Analysis

No interim efficacy analysis was performed by the sponsor. An interim futility analysis was performed by the DMC after the first 50 subjects were enrolled and their response at week 24 was established.

1.5.2. Primary Analysis

The primary analysis will be performed approximately one year after enrollment of the last subject.

1.5.3. Final Analysis

After the primary analysis the sponsor will continue to follow the subjects for an additional year to further quantify response for subjects on the study treatment and to assess OS. In addition, withdrawal of all systemic therapies for cGVHD (including ibrutinib/placebo) will be assessed.

The final analysis will be performed approximately two years after enrollment of the last subject.

1.6. Testing Procedure and Level of Significance in the primary analysis

To preserve the study wise type I error rate of 0.05, the primary and secondary endpoints will be tested based on a serial gatekeeping testing procedure at the two-sided significance level of 0.05 according to the hierarchical order specified below:

1. Response rate at 48 weeks

2. Time to withdrawal of all corticosteroids for treatment of cGVHD

3. Time to withdrawal of all ISTs, including corticosteroids for treatment of cGVHD (with the exception of ibrutinib/placebo)

4. Response rate at 24 weeks

5. Improvement in Lee cGVHD Symptom Scale at 2 consecutive visits

6. Proportion of subjects who achieve reduction of corticosteroids dose level to less than 0.15 mg/kg/d at 24 weeks sustained for at least 30 days

Since the number of deaths at the primary analysis is expected to be low, formal analysis for OS will be performed at the final analysis. However, summary of OS will be provided based on available data at the time of the primary analysis. Duration of response will be analyzed to support the response endpoints.

1.7. Blinding and Randomization Methods

1.7.1. Blinding Method

This is a double-blind study; the subjects, the investigators, and the Sponsor are blinded to the treatment assignment.

1.7.2. Randomization Method

Central randomization was implemented in this study. Subjects were randomized based on the following stratification factors:

- Age group (12 to <22 years old vs ≥22 years old)
- NIH Global Severity grade (moderate vs. severe)
- Ongoing use of systemic immunosuppressants that were initiated for either treatment of or prophylaxis for acute GVHD (yes vs. no)

Subjects were randomized in a 1:1 ratio to either Treatment Arm A or Treatment Arm B within each randomization stratum. This randomization scheme was implemented within the Interactive Web Response System (IWRS).

2. GENERAL ANALYSIS CONSIDERATION

2.1. Analysis Sets

2.1.1. Intent-to-Treat Population (ITT)

The intent-to-treat (ITT) population includes all randomized subjects. Subjects in the ITT population will be analyzed according to the treatment to which they were randomized. The ITT population will be the primary population for the analyses of efficacy endpoints (except for duration of response), all baseline characteristics, and patient reported outcome (PRO) endpoints.

2.1.2. Modified Intent-to-Treat Population (mITT)

The mITT population will include all ITT subjects who do not have evidence of progression of underlying malignancy at or before randomization. The mITT population will be used for sensitivity analyses of the primary efficacy endpoints.

2.1.3. Safety Population

Subjects in the ITT population who received at least one dose of either ibrutinib or placebo will be included in the safety population. Subjects in the safety population will be analyzed according to the actual treatment received and will be used to summarize the safety (including dosing) data.

2.1.4. Pediatric Population

The pediatric population includes the subset of the ITT population who were <18 years of age at the time of randomization. Summary or listing of the primary endpoint and selected secondary endpoints will be produced for the pediatric population as appropriate.

2.1.5. Definition of Subgroups

Subgroup analyses will be performed for the selected variables to assess the internal consistency of the treatment benefit and/or safety. The subgroup variables and the cutoff values are subject to change if warranted to better represent the data.

Table 1. Subgroup Definition

Subgroup	Definition of Subgroup	Analysis Type
Age	<65 years, ≥65 years	D, E, S
Sex	Male, Female	D, E, S
Race	White, Non-White	D, E, S
Geographic region	US, Non-US	D, E, S
NIH Global cGVHD Severity grade	Moderate, Severe	D, E, S
Ongoing use of systemic immunosuppressants at enrollment that were initiated for either treatment of or prophylaxis for GVHD	Yes, No	D, E, S
Hepatic impairment (Based on NCI ODWG liver function classification)	Yes, No	D, S
Renal impairment	Yes, No	D, S
analysis type D= demographic and baseline disease characteristics analysis type E= efficacy analysis type S= safety		

3. SUBJECT INFORMATION

Subject information will be summarized descriptively with no formal testing performed. Summaries will be presented by arm (ibrutinib+ prednisone vs placebo + prednisone).

3.1. Subject Disposition

Subject randomization will be summarized by region, country, site and by stratification factors. Subject disposition for each study treatment (for ibrutinib/placebo and for prednisone) and for study participation will be tabulated.

Time on study is defined in the same way as overall survival with reversed censoring, i.e., subjects who died will be censored at death date. The Kaplan-Meier methodology will be used to estimate the median time on study.

3.2. Demographics and Baseline Characteristics

Demographic information and baseline characteristics will be summarized using descriptive statistics for ITT population by treatment arm. Demographic information includes age, gender, race, ethnicity, and geographic region. Baseline disease characteristics include time from initial diagnosis of cGVHD, cGVHD staging (moderate, severe), acute GVHD history, ongoing use of immunosuppressants, donor HLA matching, Karnofsky performance score, organs involved, primary disease (i.e., the disease that was indication for the transplant) and time from transplant to cGVHD diagnosis.

3.3. Prior and Concomitant Medications

Medications will be coded to preferred term and Anatomical Therapeutic Chemical (ATC) class according to World Health Organization (WHO) Drug dictionary where applicable.

Prior medications are defined as medications that started prior to the first dose date of study drug (ibrutinib/placebo). Concomitant medications are defined as medications that were taken at any time on treatment (i.e. from the date of the first dose of study drug through the date of the last dose of study drug).

Concomitant medications will be summarized by therapeutic class and preferred term and by treatment arm for all subjects in the safety population. Each subject will be counted once for each preferred term, and each therapeutic class.

The following medications will be summarized separately for the safety population:

- Prior acute GVHD treatment
- Prior GVHD prophylactic treatment
- Concomitant medications
- The following concomitant medications of special interest will be summarized separately:
  - Systemic immunosuppressants other than prednisone
  - Cyp3A Inhibitors (moderate, strong)
  - Cyp3A Inducers
  - Anticoagulants and antiplatelets
- Subsequent systemic cGVHD treatment

3.4. Extent of Exposure to Study Treatment

Exposure to study treatment will be summarized for ibrutinib/placebo and for prednisone by treatment arm for the safety population.

Prednisone daily dosing will be summarized over time. Duration of prednisone treatment and total dose received will be summarized

The following parameters will be summarized for the study drug ibrutinib/placebo:

- Treatment duration
- Relative dose intensity
- Number (%) of subjects with dose reduced due to AE

4. EFFICACY

Analysis of efficacy endpoints will be conducted on the ITT population, unless otherwise specified. Table 2 summarizes the efficacy endpoints and analysis methods to be performed. For subgroup and exploratory analyses, the analyses will be performed where appropriate.

An unstratified analysis will be performed for the primary endpoint, because only 8 subjects aged 12 to <22 years were enrolled, which may result in half of the strata with only a few or possibly no subjects randomized to a particular treatment arm. In addition, errors in the assignment of stratification factors into the IWRS for both cGVHD severity and ongoing use of immunosuppressants occurred at the site level in the study.

Table 2. Efficacy Endpoints definition and Analyses

Endpoint	Definition	Analysis Method	Population
Primary Endpoint			
Response Rate at 48 Weeks	Proportion of responders [complete response (CR) or partial response (PR)] as assessed by investigators at 48 weeks using the NIH Consensus Panel Chronic GVHD Activity Assessment. A subject will NOT be considered as a responder at 48 weeks if he/she meets any of following criteria: • Starts a second line systemic therapy for cGVHD at or prior to the response assessment at 48 weeks • Has evidence of progression of the underlying malignancy that was the indication for transplant at or prior to response assessment at 48 weeks • Subjects withdrawing from the study or from response assessment prior to 48 weeks.	Unstratified Chi-square test **Subgroups (from Table 1)** Comparison of the treatment arms within each subgroup using p-values from unstratified Chi square test and graph of Odds Ratio (95% CI) for each subgroup.	ITT For mITT and Pediatric population, summary or listing will be provided as appropriate
Secondary Endpoints			
Time to withdrawal of all corticosteroids for treatment of cGVHD	Time from randomization until the date of complete withdrawal (to 0mg) of corticosteroids used to treat GVHD and sustained for 30 days. Subjects will be considered having a competing risk at the time of death, relapse of the underlying disease, progression of cGVHD, or start of subsequent cGVHD therapy. Subjects with no competing event will be censored at last available study visit.	The two Cumulative Incidence Functions (CIFs) of corticosteroid withdrawal will be compared using Gray's Chi square test. Graphs of the CIFs and landmark estimates (95%) at 6 months and 12 months	ITT

Endpoint	Definition	Analysis Method	Population
		obtained from the CIFs will be provided.	
Time to withdrawal of all ISTs, including corticosteroids for treatment of cGVHD (with the exception of ibrutinib/placebo)	Time from randomization until the date of complete withdrawal (to 0mg) of all corticosteroids used to treat GVHD AND discontinuation of all ISTs sustained for 30 days. Subjects will be considered having a competing risk at the time of death, relapse of the underlying disease, progressive disease (PD) of cGVHD, or start of subsequent cGVHD therapy. Subjects with no competing event will be censored at last available study visit	The two CIFs of withdrawal of all immunosuppressants will be compared using Gray's Chi square test. Graphs of the CIFs and landmark estimates (95%) at 6 months and 12 months obtained from the CIFs will be provided	ITT
Response rate at 24 weeks	Proportion of responders [complete response (CR) or partial response (PR)] as assessed by investigators at 24 weeks using the NIH Consensus Panel Chronic GVHD Activity Assessment. A subject will NOT be considered as a responder at 24 weeks if he/she meets any of following criteria: • Starts a second line systemic therapy for cGVHD at or prior to the response assessment at 24 weeks • Has evidence of progression of the underlying malignancy that was the indication for transplant at or prior to response assessment at 24 weeks Subjects withdrawing from the study or from response assessment prior to 24 weeks.	Comparison of the proportions between the two arms will be performed using Chi-square test	ITT
Improvement in overall score on Lee cGVHD Symptom Scale	Proportion of subjects who had Lee total score decreasing at least 7 points at 2 consecutive visits with no PD, relapse of underlying disease or start of subsequent cGVHD treatment. Proportion of subjects who had Lee total score decreasing at least 7 points at any time during the study (not preceded by PD, relapse of underlying disease or start of subsequent cGVHD treatment)	Comparison of the proportions between the two arms will be performed using chi-square test. <u>Supportive:</u> Summary of the proportions of subjects with at least 7 points improvement at any	ITT

Endpoint	Definition	Analysis Method	Population
	Mean Lee scores overtime	time will be provided. Summary of mean Lee scores overtime and graphs by treatment arm will be provided	
Proportion of subjects who achieve reduction of corticosteroids dose level to less than 0.15 mg/kg/d at 24 weeks sustained for at least 30 days	Proportion of subjects with prednisone equivalent dose reduction to a level of less than 0.15 mg/kg/d for a duration of at least 30 days at 24 weeks Mean daily corticosteroid dosing overtime	Comparison of the proportions between the two arms will be performed using Chi-square test <u>Supportive:</u> Mean (95%CI) of daily corticosteroid dosing over time will be summarized and plotted by arm	ITT
Duration of response	The interval between the date of initial documentation of a response (PR or better), and the date of first documented evidence of progressive cGVHD, relapse of underlying disease, start of subsequent cGVHD treatment, or death for responders only.	DOR will be estimated by Kaplan-Meier methodology: median DOR and landmark estimates with 2-sided 95% CIs will be provided for each treatment arm. K-M plots will be provided	Responders include subjects with response at any time during study treatment
Overall Survival	Time from the date of randomization until the date of death due to any cause. All deaths observed at the time of the data extraction date will be considered as events. For subjects who were not known to have died at the time of the analysis, OS data will be censored at date last known alive.	Survival will be estimated by Kaplan-Meier methodology: median OS and landmark estimates with 2-sided 95% CIs will be provided for each treatment arm. Kaplan-Meier plots will be provided	ITT
Exploratory Endpoints			

Endpoint	Definition	Analysis Method	Population
Improvement in SF-36 patient reported outcome	SF-36 subscales over time for each treatment arm	Descriptive summary statistics mean and 95% CI for each timepoint	ITT
Improvement in Karnofsky performance scale score	Karnofsky scores over time for each treatment arm	Descriptive summary statistics mean and 95% CI for each timepoint	ITT

5. SAFETY

Safety data will be summarized by arm for the safety population and where appropriate for the pediatric subgroup. Table 3 summarizes the planned safety analyses. Adverse events (AEs) will be coded in accordance with the Medical Dictionary for Regulatory Activities (MedDRA) version 22.1. Severity of AEs will be graded by the investigator according to the NCI-CTCAE v4.03 for hematologic and non-hematological AEs.
The primary analysis will use the ibrutinib/placebo treatment-emergent period, which will include all reported AEs satisfying any of the following criteria:

- Those events that occurred or worsened after the first dose of ibrutinib/placebo, through the treatment phase, and within 30 days following the last dose of ibrutinib/placebo or initiation of subsequent cGVHD systemic therapy, whichever occurred first.
- Events with a missing onset date with a resolution date during the ibrutinib/placebo treatment phase; or
- Events that were considered study drug-related regardless of the start date of the event.

For completeness, an additional analysis will be performed to assess AEs that occurred while the patient was being treated with either component of the study treatment (ibrutinib/placebo and/or prednisone; hereinafter referred to as study treatment AEs) and which will include all reported AEs satisfying any of the following criteria:

- Those events that occurred or worsened after the first dose of ibrutinib/placebo or prednisone, through the treatment phase, and within 30 days following the last dose of ibrutinib/placebo or prednisone, or initiation of subsequent cGVHD systemic therapy, whichever occurred first.
- Events with a missing onset date with a resolution date during the ibrutinib/placebo or prednisone treatment phase; or
- Events that were considered related to either ibrutinib/placebo or prednisone regardless of the start date of the event.

Summaries of treatment emergent AEs (TEAEs) and study treatment AEs will be provided, unless otherwise specified.

All laboratory values will be converted to and reported as international standard (SI) units. In general, only data from the central laboratory will be summarized and analyzed. All gradable laboratory parameters will be graded using the NCI CTCAE v4.03. Unless otherwise specified, only baseline and post-baseline values collected during the ibrutinib/placebo treatment-emergent period will be included in the safety analysis. Additionally, summary of hematologic toxicity will be provided.

Table 3: Summary of Safety Analyses

Assessment Type	Definition	Methods
Study treatment (ibrutinib/placebo and/or prednisone) adverse events	Study treatment AEs, serious study treatment AEs, Grade 3 or higher study treatment AEs, study treatment AEs related to prednisone, study treatment AEs leading to prednisone treatment discontinuation. Study treatment AEs resulted in death Predefined corticosteroid-associated AEs will be summarized	Descriptive summary statistics and/or listings
Ibrutinib/placebo TEAEs	TEAEs, TE SAEs, Grade 3 or worse TEAEs, Related TEAEs TEAEs leading to treatment discontinuation, TEAEs leading to dose reduction. TEAEs related to corticosteroids, TEAEs leading to corticosteroid treatment discontinuation. TEAEs resulting in death Protocol-defined events of special interest (major hemorrhage) and other safety observations Predefined corticosteroid-associated AEs will be summarized	Descriptive summary statistics and/or listings
Central Laboratory results collected during the ibrutinib/placebo TEAE period	Worst post-baseline toxicity grade for selected CTCAE gradable hematology and chemistry laboratory measurements. Liver function abnormalities.	Descriptive summary statistics and/or listings
Vital Signs collected during the ibrutinib/placebo TEAE period	Blood pressure, heart rate	Descriptive summary statistics and/or listings

AE: adverse event; TEAE= treatment-emergent adverse events as defined in Section 5; SAE= serious adverse event; CTCAE= Common Terminology Criteria for Adverse Events.

6. MODIFICATIONS OF ANALYSIS TO THE PROTOCOL

- New secondary endpoint added: "Time to withdrawal of all corticosteroids for treatment of cGVHD sustained for at least 30 days". For all endpoints that involve corticosteroid or IST withdrawal, the period of withdrawal had to be sustained for 30 days.

- Karnofsky scores will be analyzed as an exploratory efficacy endpoint in this SAP rather than as a safety endpoint as stated in the study protocol.

7. REFERENCES

Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on
Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging
Working Group report. Biol Blood Marrow Transplant. 2015;21:389-401.

Lee SJ, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. Biology of Blood and Marrow Transplantation 2002;8:444-452.

Lee SJ, Wolff D, Kitko C, et al. Measuring Therapeutic Response in Chronic Graft-versus-Host Disease. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: IV. The 2014 Response Criteria Working Group Report. Biol Blood Marrow Transplant. 2015;21:984-999.

Gray, R. (1988), A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. The Annals of Statistics, 16, 1141–1154.