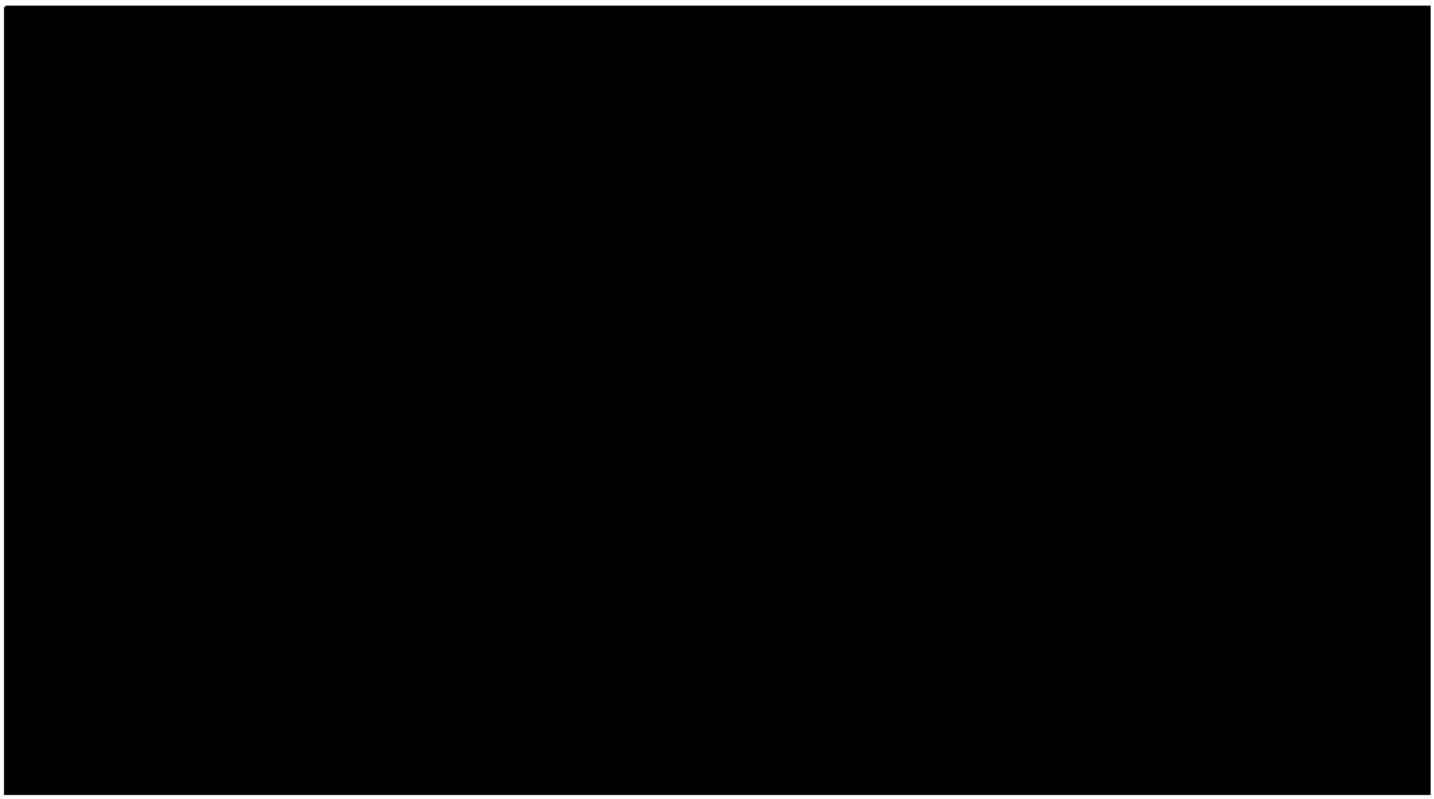
STATISTICAL ANALYSIS PLAN

A Phase 2a, Dose-Escalation, Open-Label Study to Evaluate the Safety, Tolerability, and Activity of KD025 in Subjects with Chronic Graft Versus Host Disease

Protocol Number: KD025-208

Statistical Analysis Plan Version:	Date of Statistical Analysis Plan:
Version 1	30-Nov-2017
Version 2	06-Jun-2019
Version 2.1	12-Aug-2019
Version 2.2	23-Sep-2019
Version 2.3	29-Oct-2019

SIGNATURE PAGE



DOCUMENT HISTORY

Version	Author	Description
1.0	[redacted]	Nov 30, 2017 New Document
2.0		Revision The primary reasons for this new version include: • Change of study design: expansion cohort removed. In its place, there will be another new randomized, open-label Phase 2 study, KD025-213: "A Phase 2, Randomized, Multicenter Study to Evaluate the Efficacy and Safety of KD025 in Subjects with Chronic Graft Versus Host Disease After At Least 2 Prior Lines of Systemic Therapy" • Add more definitions of duration of response in accordance with FDA type B meeting held on 01 May 2019 • Rearrange and add some secondary efficacy endpoints • Correct the calculation of symptom score for the Lee cGVHD Symptom Scale
2.1		Revision The primary reasons for this new version include: • Add subsection of "Missing event dates" • Change/clarify censoring rule for duration of response (DOR), time to new therapy (TTNT) and failure free survival (FFS) • Clarify the definition of discontinuation of Corticosteroid
2.2		Revision The primary reasons for this new version include: • Add overall survival analysis as a secondary efficacy endpoint • Add refractory to prior line status into subgroup analysis
2.3		Revision The primary reason for this new version is to move Lee symptom scale score from exploratory to secondary endpoint.

Table of Contents

1 INTRODUCTION 9

2 STUDY SUMMARY 9

2.1 Study Objectives 9

2.1.1 Primary Objectives 9

2.1.2 Secondary Objectives 10

2.1.3 Exploratory Objectives 10

2.2 Study Design 10

2.3 Visit Schedule and Study Assessment 11

3 STATISTICAL METHODS 11

3.1 General Methods 11

3.1.1 Computing Environment 11

3.1.2 General Considerations 12

3.1.3 Study Day 12

3.1.4 Baseline 13

3.1.5 Handling of Incomplete or Missing Data 13

3.2 Analysis Populations 14

3.3 Subject Disposition and Evaluability 15

3.4 Protocol Deviations 15

3.5 Demographics and Baseline Characteristics 16

3.5.1 Demographics 16

3.5.2 Medical History 16

3.5.3 cGVHD History 16
3.5.4 Other Baseline Characteristics 17
3.6 Prior and Concomitant Medications 17
3.7 Treatment Compliance and Exposure 18
3.8 Endpoints for Activity Analysis 19
3.8.1 Primary Endpoint 19
3.8.2 Secondary Endpoints 19
3.8.3 Exploratory Endpoint 25
3.8.4 Subgroup Analyses 25
3.9 Safety Analysis 26
3.9.1 Adverse Events 26
3.9.2 Clinical Laboratory Evaluation 27
3.9.3 Vital Signs 27
3.9.4 ECG 27
3.10 Pharmacokinetic Analysis 28
3.11 Pharmacodynamic Analysis 28
3.12 Sample Size Justification 28
3.13 Interim Analysis 28
3.14 Primary CSR Analysis 28
4 LIST OF TABLES, FIGURES, AND LISTINGS 29
5 REFERENCES 36
6 APPENDICES 37

List of Tables

Table 1. Cohorts in the Study 11
Table 2. Censoring Rules for Duration of Response 20
Table 3. Demographics and Baseline Characteristics TLFs 29
Table 4. Treatment Exposure and Concomitant Medication TLFs..... 30
Table 5. Efficacy TLFs 31
Table 6. Safety TLFs..... 34

LIST OF ABBREVIATIONS

Abbreviation	Full Term
7-PtR	7-point reduction
ADI	Actual dose intensity
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine transaminase
ATC	American Therapeutic Chemical (Classification)
BID	Twice daily
BMI	Body mass index
C1D1	Cycle 1 Day 1
CI	Confidence interval
cGVHD	Chronic graft versus host disease
CM	Concomitant medication
CNI	Calcineurin inhibitor
CR	Complete response
CRF	Case report form
CSR	Clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DL_{CO}	Diffusing capacity of carbon monoxide
DO7-PtR	Duration of ≥ 7-PtR
DOR	Duration of response
ECG	Electrocardiogram
EOT	End of treatment
FEV1	Forced Expiratory Volume in One Second
FFS	Failure-free survival
HCT	Hematopoietic cell transplantation
FVC	Forced vital capacity
ICH	International Conference on Harmonisation
IL	Interleukin
LR	Lack of response
LSS	Lee cGVHD Symptom Scale
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-treat
KPS	Karnofsky Performance Scale
NE	Not evaluable
ORR	Overall response rate
OS	Overall survival
PFT	Pulmonary function test

Abbreviation	Full Term
PD	Pharmacodynamics
PDI	Planned dose intensity
PK	Pharmacokinetics
PO	Oral(ly)
PPI	Proton pump inhibitor
PR	Partial response
PT	Preferred Term
RDI	Relative dose intensity
RV	Residual volume
QD	Once daily
ORR	Overall response rate
QTcF	Corrected QT interval using Fridericia's formula
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SOC	System Organ Class
STB	Stable
TEAE	Treatment-emergent adverse event
TLC	Total lung capacity
TTNT	Time to next therapy
TTR	Time to response

1 INTRODUCTION

Chronic graft versus host disease (cGVHD) remains a major complication of allogeneic hematopoietic cell transplantation (HCT) occurring in approximately 50% of transplant recipients and involving multiple organs. Subjects with cGVHD require prolonged immunosuppressive treatment for an average of 2 to 3 years from the initial diagnosis. Ten percent of those surviving for at least 7 years still require immunosuppressive treatment. Glucocorticoids, with or without calcineurin inhibitors (CNI), remain the standard initial treatment, but significant side effects and unsatisfactory outcomes, particularly for subjects with high-risk features of cGVHD, support the need for more effective and less toxic therapies.

This Statistical Analysis Plan (SAP) describes detailed statistical procedures to be used for study KD025-208 as specified in the protocol (Amendment No. 6, 15-Oct-2018): A Phase 2a, Dose-Escalation, Open-Label Study to Evaluate the Safety, Tolerability, and Activity of KD025 in Subjects with Chronic Graft Versus Host Disease. A primary clinical study report (CSR) analysis will be conducted 1 year after the last patient is enrolled. The pharmacokinetic (PK) and pharmacodynamic (PD) data will be analyzed and reported separately.

The SAP was written in accordance with the recommendations outlined in the International Conference on Harmonisation (ICH) E9 Guideline entitled “Guidance for Industry: Statistical Principles for Clinical Trials” and the most recent ICH-E3 Guideline, entitled “Guidance for Industry: Structure and Content of Clinical Study Reports.”

2 STUDY SUMMARY

2.1 Study Objectives

2.1.1 Primary Objectives

The primary objectives of this study are to evaluate:

- The activity of KD025 in subjects with steroid-dependent cGVHD and active disease in terms of partial response (PR) and complete response (CR), as defined

by the 2014 National Institute of Health (NIH) Consensus Development Project on clinical trials in cGVHD (see Appendix A)

- The safety and tolerability of KD025 in subjects with cGVHD

2.1.2 Secondary Objectives

The secondary objectives of this study are to evaluate:

- Duration of response (DOR)
- Time to response (TTR)
- Response by organ system
- Time to next therapy (TTNT)
- Changes in corticosteroid and CNI doses
- Failure-free survival (FFS)
- Overall survival (OS)
- Change in cGVHD severity as based on the Clinician-reported Global cGVHD Activity Assessment
- Changes in symptom activity using the cGVHD Activity Assessment Patient Self-Report (Appendix C)
- Plasma PK of KD025 in subjects with cGVHD

2.1.3 Exploratory Objectives

The exploratory objectives of this study are to evaluate:

- Changes in the expression of several cytokines (including interleukin [IL]-17A, IL-21, and IL-2) in plasma after KD025 administration
- Changes in percentage of immune-cell subtypes (including Th17 and Treg cells) in whole blood after KD025 administration
- Changes in symptom burden/bother using the Lee cGVHD Symptom Scale (LSS, Appendix D)

2.2 Study Design

This is a Phase 2a, open-label, dose-escalation, safety, tolerability, and activity study.

Subjects who have signed an Institutional Review Board/Independent Ethics Committee-approved informed consent form and who have met all of the inclusion/exclusion criteria will be enrolled. After approximately 4 weeks for screening, subjects will receive KD025 orally (PO) in 28-day cycles until disease progression or unacceptable toxicity occurs followed by 4 weeks of follow-up. In addition, subjects will be contacted every 8 weeks until study close-out to confirm survival status and the initiation of any anticancer treatment or any changes in treatment for cGVHD.

Sequential cohorts of approximately 16 subjects will be dosed at KD025 200 mg once daily (QD) PO, KD025 200 mg twice daily (BID) PO, and KD025 400 mg QD (Table 1). Prior to enrolling subsequent cohorts, the safety data in each previous cohort will be evaluated after 8 subjects have reached 2 months of treatment.

Table 1. Cohorts in the Study

Cohort	Number of Subjects	Treatment
Cohort 1	~16	200 mg KD025 QD
Cohort 2	~16	200 mg KD025 BID
Cohort 3	~16	400 mg KD025 QD

BID = twice daily; QD = once daily

2.3 Visit Schedule and Study Assessment

The flow chart of visit schedule and study assessments is given in the KD025-208 Protocol.

3 STATISTICAL METHODS

3.1 General Methods

3.1.1 Computing Environment

All statistical programming and data analyses will be performed using SAS® Version 9.4 on Windows platform.

3.1.2 General Considerations

General considerations for descriptive statistics and presentation for continuous and categorical data are given below.

Continuous data will be described using descriptive statistics: number of observations (n), mean, standard deviation, median, minimum, and maximum. Frequencies and percentages will be used for summarizing discrete (categorical) data. When categorical data are presented, the percent will be suppressed when the count is zero in order to draw attention to the nonzero counts. The denominator for all percentages, unless otherwise specified, will be the number of subjects in the cohort or in the specified analysis population.

Means, medians, standard deviations, and confidence intervals (CIs) will be reported to one decimal place more than the data reported on the case report form (CRF) or by the vendor. Minimum and maximum will be reported to the same number of decimal places displayed on the CRF or by the vendor. P-values will be reported to 4 decimal places.

Unless otherwise indicated, all summary tables will be presented by the following groups: Cohort 1 (KD025 200 mg QD PO), Cohort 2 (KD025 200 mg BID PO), Cohort 3 (KD025 400 mg QD PO), and Overall.

Because of the exploratory nature of the study, no adjustments will be made for the multiplicity of endpoints.

3.1.3 Study Day

The Study Day for all assessments prior to the first study drug administration is calculated as the difference between the date of the event or measurement (e.g., adverse event [AE] onset date, assessment date, sample collection date) and the start date of study treatment. The day before the start of study treatment is Study Day -1.

The Study Day for all post-assessments after the first study drug administration is calculated as the difference between the date of the event or measurement (e.g., AE

onset date, assessment date, sample collection date) and the start date of study treatment plus one day. The first day of study treatment is Study Day 1.

3.1.4 Baseline

Baseline value is defined as the valid and last non-missing value obtained within 28 days prior to subject receiving the first study medication, unless otherwise stated under the related assessment section. Baseline can be the day before the first study medication or on the same day as the first study medication if a pre-dose assessment is available. Subjects without data on a parameter before the first study medication will have a missing baseline for this parameter.

3.1.5 Handling of Incomplete or Missing Data

Missing data will not be imputed in general and will be reported as missing in all listings. For categorical variables, patients with missing data are not included in calculations of percentages unless otherwise specified.

Missing start and end dates for AE and concomitant medication (CM)

The assumption will be the worst or most conservative judgment when imputing AE and CM start and end dates. The purpose of imputing a start date is to help define whether the AE/CM started while taking study drug.

For a partial or missing start date:

- If the day is missing, the first day of the month will be imputed. If the missing day is the same as the month of first dose of study drug, then the first dose date will be imputed.
- If the day and month are missing, the first day of January will be imputed. If the year is the same as the first dose date, then the first dose date will be imputed.
- If the day is completely missing, the first dose date will be imputed. If the end date suggests it could have started prior to this, the first day of January of the same year as the end date will be imputed.
- When imputing a start date, the start date will ensure that the new imputed date is sensible, i.e., is prior to the end date of the AE or CM.

For a partial or missing end date:

- If the day is missing, the last day of the month or the last assessment date, whichever is earlier, will be imputed.
- If the day and month are missing, the 31st of December or the last assessment date, whichever is earlier, will be imputed
- If the date is completely missing, there will be a need to look at whether the AE/CM is still ongoing before imputing a date. If the ongoing flag is missing, then it will be assumed that AE is still present, or CM is still being taken (i.e., do not impute a date). If the AE/CM has stopped, then the last assessment date will be imputed.

These data imputations are for categorization purpose only and will not be used in the listings.

If the assessment of the relationship of the AE to KD025 is missing, then it will be assumed that the AE is related to KD025 and the AE considered as such in the frequency tables of possibly related AEs. No imputation should be done at the data level.

Missing event dates

Event date will be imputed only when day is missing, and the purpose of imputing an event date is to most conservatively calculate time to event.

If the day is missing, the first (mid, last) day of the month will be imputed for undesired (neutral, desired) event. If the missing day is the same as the month of first dose of study drug, then the first dose date will be imputed for undesired event.

These data imputations are for time to event calculation only and will not be used in the listings.

3.2 Analysis Populations

The following populations will be analyzed:

Modified Intent-to-treat (mITT) Population: The primary population for efficacy analyses will be a Modified Intent-to-treat (mITT) Population defined as all subjects who receive at least 1 dose of study medication. The mITT Population will be used for tables of demography, baseline characteristics, and efficacy.

Responder Population: The Responder Population is defined as subjects in the mITT population that achieved a PR or CR response at any post-baseline response assessment.

Non-responder Population: The Non-responder Population is defined as any subject in the mITT Population who is not a responder.

The responder and non-responder populations will be used for some subgroup analyses.

Safety Population: The Safety Population is defined as all subjects who receive at least 1 dose of study medication. In this study, the Safety Population is equivalent to the mITT Population.

3.3 Subject Disposition and Evaluability

Subjects who failed screening (i.e., subjects who signed the informed consent were screened but never started the study treatment, and their basic demographics and any AE after signing the informed consent may have been collected in the CRF) will be excluded from any populations defined in Section 3.2. Therefore, these subjects will be excluded from any summary tables or listings.

The number of subjects discontinuing from the study and the primary reason for discontinuation will be summarized.

3.4 Protocol Deviations

All protocol deviations will be identified and classified as major or minor before clinical database lock and will be presented in a listing.

Major Deviation: Protocol deviation that may impact the accuracy, and/or reliability of the study data or that may impact subject rights, safety or well-being.

Minor Deviation: Protocol deviation that does not impact the accuracy, and/or reliability of the study data or subject rights, safety or well-being.

Serious Non-compliance: Serious non-compliance presents a significant risk to the safety of study patients or significantly affects the scientific value of the reported results of the study. This classification may include fraud, scientific misconduct and serious breaches of ethical conduct. Persistent clinical investigator site noncompliance, even if not serious, is also considered within this definition.

3.5 Demographics and Baseline Characteristics

3.5.1 Demographics

Subject demographics and baseline characteristics will be summarized for the mITT Population. Descriptive statistics will be provided for age, height, weight, and body mass index (BMI). Frequencies and percentages will be tabulated for sex, race, and ethnicity.

BMI will be calculated as follows:

$$\text{BMI } (kg/m^2) = \text{weight } (kg)/\text{height}^2\ (m^2),$$

where the weight and height are obtained at baseline.

3.5.2 Medical History

Medical history will be summarized by primary System Organ Class (SOC) and Preferred Term (PT). Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 20.1 (or higher) terminology.

3.5.3 cGVHD History

Transplant history will be summarized for:

- Indication for transplant
- Stem cell type (bone marrow, peripheral blood, cord blood)
- Donor source (related, unrelated)
- Conditioning regimen (myeloablative, nonmyeloablative)
- Donor age
- Donor gender
- Donor-recipient gender mismatch

- Donor-recipient cytomegalovirus (CMV) serostatus

The baseline characteristics of cGVHD will be summarized for:

- Time from most recent transplant to cGVHD diagnosis
- Time from cGVHD diagnosis to study enrollment
- Time from most recent transplant to enrollment
- Prior therapies for cGVHD will be summarized for:
    - Number of lines of prior systemic therapies
    - Prior systemic therapies
- Organ involvement will be summarized for
    - Number of organs involved at baseline
    - Organ involvement at baseline including:
        - NIH Skin Score and Skin Features Score
        - NIH Eye Score
        - NIH-modified Oral Mucosa Rating Scale (OMRS)
        - NIH Esophagus Score
        - NIH Upper Gastrointestinal (GI) Score
        - NIH Lower GI Score
        - Alanine transaminase (ALT), Alkaline phosphatase (ALP), Total Bilirubin
        - NIH Lung Symptom Score and Forced Expiratory Volume in One Second (FEV1) (% predicted)
        - NIH Joint and Fascia Score and P-ROM score
        - Global Severity Rating

3.5.4 Other Baseline Characteristics

Other baseline values including Karnofsky Performance Scale (Appendix E; Karnofsky et al., 1949) and childbearing potential will also be summarized.

3.6 Prior and Concomitant Medications

Concomitant medications will be coded using the World Health Organization Drug Dictionary. Prior medications are defined as those medications that began and

stopped before the start of study treatment. Concomitant medications are defined as medications taken after the start of study treatment and during the study period, including those began before but ongoing at the start of study treatment. If a medication start date is partially or fully missing and it is unclear as to whether the medication is prior or concomitant, it will be assumed that the medication is concomitant.

Number and percentage of incidence of prior and CM will be summarized according to Anatomical Therapeutic Class (ATC) and preferred drug name.

3.7 Treatment Compliance and Exposure

The relative dose intensity (RDI) will be used to assess the treatment compliance. The RDI is defined as:

$$\text{RDI (\%)} = 100 \times \text{ADI (mg/day)}/\text{PDI (mg/day)},$$

where ADI and PDI are the actual dose intensity and planned dose intensity, respectively:

$$\text{PDI (mg/day)} = \text{planned cumulative dose (mg)} / \text{duration of exposure (days)},$$
$$\text{ADI (mg/day)} = \text{actual cumulative dose (mg)}/ \text{duration of exposure (days)}.$$

The planned cumulative dose is the planned daily dose amount multiplied by the duration of exposure, while the actual cumulative dose is the sum of actual total daily dose amount over the duration of exposure. The actual total daily dose will need the information of dose change and dose held (interruption) captured in CRF (Missed Dose and Dose Modification). If a subject did not take any study drug, the actual RDI by definition is zero.

The duration of exposure is defined as:

$$\text{Duration of exposure (days)} = (\text{Date of last dose} - \text{Date of Study Day 1}) + 1.$$

The descriptive summary statistics of the treatment duration, actual cumulative dose, ADI, and RDI will be presented.

3.8 Endpoints for Activity Analysis

3.8.1 Primary Endpoint

The overall response rate (ORR) is the primary endpoint of the study.

The ORR is defined as the proportion of subjects meeting the overall response criteria assessment of CR or PR at any post-baseline response assessment.

The overall response determination of cGVHD is based on the cGVHD response assessment performed by clinicians as per the 2014 NIH Consensus Development Project for Clinical Trials in cGVHD criteria. The overall response is defined using the scores from 9 organs (skin, eyes, mouth, esophagus, upper GI, lower GI, liver, lungs, and joints and fascia; see Appendix B) plus global severity rating. The results of overall response at each assessment time point are evaluated by investigator and categorized as CR, PR, or Lack of Response (LR), where LR includes the response status of unchanged, mixed, or progression.

If a treated subject does not have any response assessment, this subject will be counted as a non-responder.

The ORR and its two-sided 95% CI using Clopper-Pearson method will be presented.

3.8.2 Secondary Endpoints

3.8.2.1 Duration of Response (DOR)

The primary definition of DOR is the time from first documentation of response to the time of first documentation of deterioration from best response (e.g., CR to PR, or PR to LR).

The secondary definition of DOR is the time from first documentation of response to the time of first documentation of lack of response.

The tertiary definition of DOR is the time from first documentation of response to the time of initiation of new systemic cGVHD therapy (which will be defined by clinical team review).

The quaternary definition of DOR is the time from first documentation of response to the time of first documentation of lack of response (as the secondary definition) but with durations summed for multiple response/lack of response episodes.

The DOR will be reported only for responders and statistics will include:

- Kaplan-Meier plots and descriptive statistics of DOR. The censoring rules in Table 2 will be applied
- Landmark analyses: Number and percentage of subjects with response sustained for ≥ 12, ≥ 20, ≥ 24, ≥ 32, ≥ 36, and ≥ 48 weeks

Table 2. Censoring Rules for Duration of Response

DOR	Events	Censoring
Primary	• Deterioration from best response • Initiation of new systemic therapy for cGVHD • Death	• Last documented response assessment • If LR or initiation of new systemic therapy happens immediately after two or more missed response assessments, the event date should be set as four weeks (one cycle) after last documented response assessment prior this event
Secondary	• Documented LR • Initiation of new systemic therapy for cGVHD • Death	
Tertiary	• Initiation of new systemic therapy for cGVHD • Death	Last response assessment or long term follow up assessment, whichever is the latest and available
Quaternary	• Documented LR • Initiation of new systemic therapy for cGVHD • Death With summation of DOR from multiple episodes	Same with censoring rule for primary and secondary

cGVHD = chronic graft versus host disease; DOR = duration of response; LR = lack of response

3.8.2.2 Time-to-Response (TTR)

Time-to-response will be measured as the time from first treatment to the time of first documentation of response. Descriptive statistics and plots of cumulative number and percent of responders over time (at 4, 8, 12, 16, 24, 32, 40, and ≥ 48 weeks) will be provided. The TTR analyses will only be conducted for the Responder Population.

3.8.2.3 Response by Organ System

The database did not specifically capture baseline organ involvement. Organ involvement was determined as follows:

- All organs, except lungs:
  - If the baseline score is > 0, the organ will be defined as Involved, unless all response assessments are Not Evaluable (NE)
  - If the baseline score is 0, the organ will be defined as Not Involved, unless any response assessment of CR or PR which triggers clinical review to confirm baseline involvement.
- For the lungs:
  - If FEV1 ≥ 75%, the organ will be defined as Not Involved
  - If FEV1 < 75%, the organ will be defined as Involved, unless all responses are NE

The response assessment for the 9 individual organs (skin, eyes, mouth, esophagus, upper GI, lower GI, liver, lungs, and joints and fascia) and global severity rating is categorized into 5 response status: CR, PR, Progression (P), NE, Stable (STB).

Number and percentage of subjects with best response of CR or PR and total (CR+PR) by individual organ system and global severity rating will be summarized and 2 series of percentages presented:

1. With the total number of subjects in the mITT Population with involvement of the given organ at baseline as denominator
2. With the number of subjects in the Responder Population with involvement of the given organ at baseline as denominator

3.8.2.4 Time to Next Therapy (TTNT)

The TTNT will be measured as the time from first treatment to the time of new systemic cGVHD treatment (which will be defined by clinical team review), censored by last response assessment or long term follow up assessment, whichever is the latest and available. TTNT will be analyzed by the Kaplan-Meier survival method as well as the landmark analysis at 6, 12, 18, and 24 months.

3.8.2.5 Corticosteroid Dose

Corticosteroid doses on Day 1 of each treatment cycle will be presented as mg/kg/day prednisone equivalent dose. Descriptive statistics for the mITT Population, Responder and Non-responder Populations and subgroups defined by baseline corticosteroid dose level (upper and lower 50th percentiles) will be provided for:

- Systemic corticosteroid dose over time
- Change and percent change from baseline (Cycle 1 Day 1 [C1D1]) to the greatest corticosteroid dose reduction during KD025 treatment period
- Number and percentage of subjects who reduced systemic corticosteroid dose during KD025 treatment period
- Number and percentage of subjects who ever discontinued systemic corticosteroid usage during KD025 treatment period

If subjects are not using prednisone as the systemic corticosteroid, then the equivalent dose of 1 mg of prednisone will be determined according to following conversion ratios:

- 4 mg Hydrocortisone
- 0.8 mg Methylprednisolone
- 0.15 mg Dexamethasone
- 1 mg Prednisolone
- 0.8 mg Triamcinolone

3.8.2.6 Calcineurin Inhibitor (CNI) Doses

Calcineurin inhibitors include systemic tacrolimus and cyclosporine. Descriptive statistics will be provided for:

- Number and percentage of subjects who reduced CNI dose
- Number and percentage of subjects who ever discontinued CNI during KD025 treatment period

3.8.2.7 Failure-free Survival (FFS)

Failure-free survival will be defined as the time from first dose of study drug to either the start of another new systemic treatment for cGVHD, relapse of the underlying

disease (both will be defined by clinical team review), or death. If no such events happen, FFS will be censored by last response assessment or long term follow up assessment, whichever is the latest and available. Failure-free survival will be analyzed by the Kaplan-Meier survival method as well as the landmark analyses at 6, 12, 18, and 24 months. In addition, analyses for the three components of FFS will also be provided.

3.8.2.8 Overall Survival (OS)

Overall survival will be defined as the time from first dose of study drug to death due to any reason. If there is no death, OS will be censored by last visit, last long term follow up, or study cutoff date, whichever is the latest. OS will be analyzed by the Kaplan-Meier survival method as well as the landmark analyses at 6, 12, 18, and 24 months.

3.8.2.9 Change in cGVHD Severity as Based on the Clinician-reported Global cGVHD Activity Assessment

The Clinician-reported global cGVHD Activity Assessment is a 0-10 point numeric rating scale with a score of 0 indicating "cGVHD symptoms not at all severe" and a score of 10 being "Most severe cGVHD symptoms possible". The activities are assessed on Day 1 of each cycle from Cycle 1 Day 1 to EOT.

Change from baseline in cGVHD severity based on the Clinician-reported global cGVHD Activity Assessment will be summarized as a categorical endpoint at all scheduled assessment visits.

3.8.2.10 Change in Symptom Activity Based on cGVHD Activity Assessment Patient Self Report

The cGVHD Activity Assessment-Patient Self Report is in Appendix C. Activities are assessed on Day 1 of each cycle from C1D1 to end of treatment (EOT). The symptom activity item is a 0-10-point numeric rating scale with a score of 0 indicating "cGVHD symptoms not at all severe" and a score of 10 being "most severe cGVHD symptoms possible". The status reported by subjects are categorized as none, mild, moderate, and severe. The comparison of cGVHD symptoms to a month ago will also be reported by subjects, ranging from -3 (very much worse) to +3 (very much better).

Changes in cGVHD symptoms based on global cGVHD Activity Assessment by the Patient Self Report will be summarized as a continuous endpoint at all scheduled assessment visits. Both scores as well as the change-from-baseline values will be presented.

The summary of the change-from-baseline of Global Severity Rating on categorical status and the summary of comparison of cGVHD to a month ago will also be presented.

The number and percentage of subjects reporting none, mild, moderate, and severe cGVHD will be summarized by visit.

3.8.2.11 Lee Symptom Scale Score (LSS)

The Lee cGVHD Symptom Scale Score (LSS) is the exploratory endpoint in this study.

The LSS (see Appendix D) will be assessed on Day 1 of Cycles 2-5, then on Day 1 of every other cycle thereafter and EOT. The questionnaire consists of 30 items over 7 domains: skin, eyes and mouth, breathing, eating and digestion, muscles and joints, energy, and mental and emotional. Each question is scored 0, 1, 2, 3, or 4.

A domain score will be calculated for each domain by taking the mean of all items completed if more than 50% were answered and normalizing to a 0 to 100 scale. A summary score will be calculated as average of all non-missing domain scores if more than 50% of them are non-missing. A higher score indicates more bothersome symptoms. A 7-point difference on the summary score of cGVHD symptom scale has been found to be clinically meaningful (Lee et al., 2002).

$$Raw \quad Score = \frac{I_1 + \cdots + I_m}{m}, \qquad Domain \quad Score = (Raw \quad Score / 4) * 100$$

$$Summary \quad Score = \frac{Domain \quad Score_1 + \cdots + Domain \quad Score_n}{n}$$

In the above formulae, ‘m’ is the number of items that have non-missing values in a domain, and ‘n’ is the number of domains have non-missing domain scores.

The analyses will include:

- Both score and change-from-baseline values (summary score and domain scores) will be summarized as continuous variables by visit
- Number and percentage of subjects with a ≥ 7-point reduction (7-PtR) from baseline (C1D1)
- Number and percentage of subjects with a 7-PtR from baseline on 2 consecutive assessments
- Duration of a 7-PtR (DO7-PtR) (defined as time from documentation of the first ≥ 7-point reduction to the first documentation of less than 7-point reduction). If there are multiple episodes, then DO7-PtR will be measured as the sum of DO7-PtR from all episodes.

These analyses will be performed on mITT, Responder, and Non-responder Populations.

Descriptive statistics of domain, overall scores, and their change from baseline will be provided by cohort and visit.

3.8.2.12 Pulmonary Function Tests (PFT)

Five test variables of PFTs, Forced Vital Capacity (FVC), Residual Volume (RV), Diffusing Capacity of Carbon Monoxide (DLco), FEV1, and Total Lung Capacity (TLC) are collected from the CRF. Change from baseline values of FVC, RV, DLco, FEV1, and TLC will be summarized at each scheduled time point.

3.8.3 Exploratory Endpoint

Not applicable.

3.8.4 Subgroup Analyses

Subgroup analyses for the endpoints of ORR and DOR will be conducted for the following subgroups:

- Number of prior line therapies (1 vs. ≥ 2)
- Number of organs involved at baseline (< 4 vs. ≥ 4)
- Baseline severity (severe vs. not severe, baseline activity assessment used as surrogate)

- Take concomitant medication proton pump inhibitor (PPI) on C1D1 (Yes/No)
- Refractory to most recent line of therapy prior to enrollment (Yes / No)

3.9 Safety Analysis

Safety assessments will include AEs, serious adverse events (SAEs), vital sign measurements, clinical laboratory evaluations (hematology and chemistry), and electrocardiograms (ECGs). Unscheduled visits for safety assessments will not be presented in summary tables but will be in listings. All safety analyses will be performed using the safety population.

3.9.1 Adverse Events

AEs will be coded using the MedDRA dictionary (Version 20.1 or higher). Treatment-emergent AEs (TEAEs) are any AE occurring or worsening in severity after the first administration of study medication. All AEs (including SAEs) will be graded using the 5-point Common Terminology Criteria for Adverse Events (CTCAE) V5.0 scale (mild, moderate, severe, life-threatening, or death). Causality with KD025 will be classified as: definitely related; probably related; possibly related; unlikely related; or not related.

The investigator will further assess the relationship of AEs to cGVHD or the underlying disease.

The number and percentage of patients who experienced at least one TEAE as well as the number and percentage of patients who experienced AEs of each specific SOC and PT will be presented. For the presentation of AE incidences, the SOCs and the PTs within each SOC will be presented by decreasing total frequency. Tabulation by maximum severity and relationship to KD025 will also be included by treatment group.

The TEAEs, Grade ≥ 3 TEAEs, SAEs, and TEAEs leading to dose modification/discontinuation will be summarized by treatment arm, SOC, and PT. These analyses will be repeated for events considered related (definitely related/probably related/possibly related) to KD025.

Subject listings will be provided for SAEs, AEs resulting in study drug discontinuation and deaths.

Adverse events will also be presented in listings. Time to onset and duration of AEs will be included in listings, along with action taken and outcome.

3.9.2 Clinical Laboratory Evaluation

The summary statistics (including number, mean, standard deviation, median, minimum and maximum) of all laboratory variables and changes from baseline will be calculated for each visit or study assessment by treatment group. For parameters of white blood cell counts, neutrophils (absolute count), lymphocytes (absolute count), monocytes (absolute count), hemoglobin, platelets, ALP, ALT, aspartate aminotransaminase, gamma glutamyl transferase, total bilirubin, glomerular filtration rate, plots of mean/mean changes from baseline with the corresponding standard error will be displayed.

For shift tables, laboratory results will be classified using the CTCAE Version 5.0. All graded laboratory parameters will be summarized separately for hematology and biochemistry. Corresponding shift tables comparing baseline to the worst post-baseline grade within the treatment period will be provided.

3.9.3 Vital Signs

Descriptive statistics for vital signs (weight, temperature, blood pressure, pulse rate, and respiratory rate) values and the change from baseline will be presented for each scheduled assessment time point.

3.9.4 ECG

Descriptive statistics for ECG parameters (i.e., PR interval, QRS interval, and QTcF interval) at each time point with triplicate ECGs will be presented for the values and change from baseline scores. (QTcF is the QT interval using Fridericia's correction which is calculated by $QTcF = QT/RR^{1/3}$.)

The number and percentage of subjects with observed QTcF values that satisfy the following conditions will be presented by treatment group and study visit and categorized as: ≤ 450 ms; > 450 to 480 ms; > 480 to 500 ms; and > 500 ms.

The number and percentage of subjects having change from baseline QTcF values that satisfy the following conditions will be presented by treatment group and study visit and categorized as: ≤ 0 ms; > 0 to ≤ 30 ms; > 30 to ≤ 60 ms; and > 60 ms.

3.10 Pharmacokinetic Analysis

Further details will be described in the Pharmacokinetic Analysis Plan.

3.11 Pharmacodynamic Analysis

Further details will be described in the Pharmacodynamics Analysis Plan.

3.12 Sample Size Justification

The total sample size of 16 subjects per KD025 dose group will provide > 90% probability of ≥ 1 subject experiencing an AE that has an underlying rate of ≥ 14% and > 80% probability of ≥ 1 subject in the study experiencing an AE that has an underlying rate of ≥ 10%.

3.13 Interim Analysis

This is a dose-escalation, open-label study. Interim analyses may occur periodically during the course of this trial. There is no early stopping for efficacy.

3.14 Primary CSR Analysis

A primary CSR analysis will be conducted 1 year after the last subject is enrolled.

4 LIST OF TABLES, FIGURES, AND LISTINGS

In all tables, figures and listing, results will be presented by treatment cohorts and overall, unless be otherwise specified.

Table 3 lists demographic and baseline characteristics TLFs; Table 4 lists treatment exposure and concomitant medication TLFs; Table 5 lists efficacy TLFs; and Table 6 lists safety TLFs.

Table 3. Demographics and Baseline Characteristics TLFs

T/F/L	Title	Population
T	Analysis populations by cohort (mITT Population)	mITT
T	Subject disposition	mITT
L	Subject disposition	mITT
T	Summary of follow-up duration	mITT
L	Protocol deviations	mITT
T	Demographics and baseline characteristics	mITT
L	Demographics and baseline characteristics	mITT
T	Medical histories by primary system organ class and preferred terms	mITT
L	Medical history	mITT
T	cGVHD and transplant history	mITT
L	cGVHD and transplant history	mITT
T	Summary of organ involvement at baseline and baseline score	mITT
T	Summary of prior systemic cGVHD therapy	mITT
T	Summary of prior systemic cGVHD therapy by ATC class	
L	Prior systemic cGVHD therapy	mITT

Table 4. Treatment Exposure and Concomitant Medication TLFs

T/F/L	Title	Population
T	Concomitant corticosteroid medications for cGVHD by ATC class, preferred term and cohort	mITT
L	Concomitant corticosteroid medications for cGVHD by cohort and subject	mITT
T	Concomitant calcineurin inhibitors for cGVHD by ATC class, preferred term and cohort	mITT
L	Concomitant calcineurin inhibitors for cGVHD by cohort and subject	mITT
T	Concomitant medications prior to the start of study drug by ATC class, preferred term and cohort	mITT
L	Concomitant medications prior to the start of study drug by cohort and subject	mITT
T	Concomitant medications after the start of study drug by ATC class, preferred term and cohort	mITT
L	Concomitant medications after the start of study drug by cohort and subject	mITT
T	KD025 treatment exposure and compliance	mITT
T	KD025 dose reductions and interruptions	mITT
L	KD025 dose reductions and interruptions	mITT

Table 5. Efficacy TLFs

T/F/L	Title	Population
Primary Endpoint: ORR		
T	Overall Response Rate	mITT
T	Best Overall Response	mITT
L	Listing of Overall Response by subject by visit	mITT
Secondary Endpoint: DOR		
T	Descriptive Kaplan-Meier and landmark statistics for primary DOR	Responder
F	Kaplan-Meier plot of primary DOR	Responder
L	Listing of all primary DOR episodes for each subject	Responder
T	Descriptive Kaplan-Meier and landmark statistics for secondary DOR	Responder
F	Kaplan-Meier plot of secondary DOR	Responder
L	Listing of all secondary DOR episodes for each subject	Responder
T	Descriptive Kaplan-Meier and landmark statistics for tertiary DOR	Responder
F	Kaplan-Meier plot of tertiary DOR	Responder
L	Listing of all tertiary DOR episodes for each subject	Responder
T	Descriptive Kaplan-Meier and landmark statistics for quaternary DOR	Responder
F	Kaplan-Meier plot of quaternary DOR	Responder
L	Listing of all quaternary DOR episodes for each subject	Responder
Secondary Endpoint: TTR		
T	Descriptive statistics for time to response (TTR) and cumulative response rate over time	Responder
L	Listing of TTR for each subject	Responder
Secondary Endpoint: Response by Organ System		
T	Best response by individual organ	mITT
T	Best response by individual organ	Responder
L	Listing of organ score and response assessment by subject and visit	mITT
T	Descriptive statistics for time to response (TTR) and cumulative response rate over time by organ	mITT
T	Descriptive statistics for time to response (TTR) and cumulative response rate over time by organ	Responder
Secondary Endpoint: TTNT		
T	Descriptive Kaplan-Meier and landmark statistics for TTNT	mITT
F	Kaplan-Meier plot for TTNT	mITT
Secondary Endpoint: Corticosteroid Dose		
T	Descriptive statistics for prednisone equivalent dose of corticosteroids	mITT Responder

T/F/L	Title	Population
		Non-responder BL corticosteroid dose level (upper and lower 50th percentiles)
T	Summary of raw and change from baseline on prednisone equivalent dose of corticosteroids by visit	mITT Responder Non-responder BL corticosteroid dose level (upper and lower 50th percentiles)
L	Listing of all the steroid raw and standardized dosing by subject and visit	mITT
Secondary Endpoint: CNI Dose		
T	Summary of descriptive statistics for CNI	mITT
L	Listing of all the CNI dosing by subject and visit	mITT
Secondary Endpoint: FFS		
T	Descriptive Kaplan-Meier and landmark statistics for FFS (with cumulative failure rates for each of the 3 components FFS)	mITT
F	Kaplan-Meier plot for FFS (with cumulative failure rates for 3 components of failures)	mITT
L	Listing of failure events	mITT
Secondary Endpoint: Change in cGVHD severity as based on the Clinician-reported Global cGVHD Activity Assessment		
T	Maximal change from baseline in global severity rating based on Clinician-reported cGVHD Assessment	mITT
Secondary Endpoint: Symptom activity using the cGVHD Activity Assessment Patient Self-Report		
T	Maximal change from baseline in symptom activity (scored 0-10) based on cGVHD Activity Assessment Patient Self Report	mITT
T	Summary of severity rating on categorical status (none, mild, moderate, severe) based on cGVHD Activity Assessment Patient Self Report by visit	mITT
T	Descriptive statistics of cGVHD symptoms based on cGVHD Activity Assessment Patient Self Report (change of cGVHD symptom to a month ago) by visit	mITT
Secondary Endpoint: LSS score		

T/F/L	Title	Population
T	Descriptive statistics for LSS score	mITT Responder Non-responder
T	Summary of LSS Score (overall and each of seven domains) raw and change from baseline by visit	mITT Responder Non-responder
L	Listing of Lee Symptom Scale Score by subject, domain and visit	mITT
Subgroup analyses (subgroups are listed in section 3.8.4 Subgroup Analyses)		
T	Overall Response Rate by Subgroup	mITT
T	Descriptive Kaplan-Meier and landmark statistics for primary DOR by subgroup	mITT
F	Kaplan-Meier plot of all DOR by subgroup	mITT
L	Listing of PPI patients by cohort	mITT

Table 6. Safety TLFs

T/F/L	Title	Population
T	Overview of subjects for safety analysis by cohort	Safety
T	Treatment emergent adverse events regardless of the relationship of study drug by primary system organ class, preferred term, maximum grade	Safety
T	Treatment-emergent adverse events related to the study drug by primary system organ class, preferred term, maximum grade	Safety
T	Treatment emergent serious adverse events regardless of the relationship of study drug, by primary system organ class, preferred term, maximum grade	Safety
T	Treatment emergent serious adverse events related to the study drug, by primary system organ class, preferred term, maximum grade	Safety
T	Treatment emergent adverse events leading to treatment discontinuation, regardless of the relationship of study drug, by primary system organ class and preferred term	Safety
T	Serious adverse events leading to treatment discontinuation, regardless of the relationship of study drug, by primary system organ class and preferred term	Safety
T	Adverse events with grade ≥ 3, regardless of the relationship of study drug, by primary system organ class and preferred term	Safety
T	Treatment emergent adverse events leading to treatment discontinuation, regardless of the relationship of study drug, by preferred term	Safety
T	Serious adverse events leading to treatment discontinuation, regardless of the relationship of study drug, by preferred term	Safety
T	Adverse events with grade ≥ 3, regardless of the relationship of study drug, by preferred term	Safety
T	Adverse events with any grade, regardless of the relationship of study drug, by preferred term	Safety
L	Listings of deaths, other serious and significant AEs	Safety
T	Summary of the worst on-treatment grade for hematology by cohort	Safety
T	Summary of the worst on-treatment grade for biochemistry by cohort	Safety
T	Hematology shift table based on CTC grade by cohort	Safety
T	Biochemistry shift table based on CTC grade by cohort	Safety
T	Change from baseline of PFT by visit and cohort	Safety
T	Change from baseline of hematology by visit and cohort	Safety
T	Change from baseline of biochemistry by visit and cohort	Safety
T	Change from baseline of urinalysis by visit and cohort	Safety

T/F/L	Title	Population
F	Liver function test: Mean ALT over time by cohort	Safety
F	Liver function test: Mean change from baseline of ALT over time by cohort	Safety
F	Liver function test: Distribution of ALT over time by cohort	Safety
F	Liver function test: Mean AST over time by cohort	Safety
F	Liver function test: Mean change from baseline of AST over time by cohort	Safety
F	Liver function test: Distribution of AST over time by cohort	Safety
F	Liver function test: Mean alkaline phosphatase over time by cohort	Safety
F	Liver function test: Mean change from baseline of alkaline phosphatase over time by cohort	Safety
F	Liver function test: Distribution of alkaline phosphatase over time by cohort	Safety
F	Liver function test: Mean total bilirubin over time by cohort	Safety
F	Liver function test: Mean change from baseline of total bilirubin over time by cohort	Safety
F	Liver function test: Distribution of total bilirubin over time by cohort	Safety
T	Vital signs shift table of vital signs by visit and cohort	Safety
T	Change from baseline of in ECG by visit and cohort	Safety
T	Number and percentage of subjects with QTcF by visit and cohort	Safety
T	Number and percentage of subjects with QTcF change from baseline by visit and cohort	Safety
L	Deaths during on-treatment period by cohort	Safety
L	Adverse events by cohort	Safety

5 REFERENCES

Karnofsky, D.A., and Burchenal, J.H. (1949). The Clinical Evaluation of Chemotherapeutic Agents in Cancer. In Evaluation of Chemotherapeutic Agents: Symposium Held at the New York Academy of Medicine, C.M. MacLeod, ed. (New York, Columbia University Press), pp. 191-205.

Lee S, Cook EF, Soiffer R, et al. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 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.

6 APPENDICES

Appendix A: Clinician-Reported Global cGVHD Activity Assessment Form A

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**	**Over the <<time>> 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):_____

Figure 1. Chronic GVHD Activity Assessment- Clinician Report.

Appendix A: Clinician-Reported Global cGVHD Activity Assessment Form A (cont.)

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				
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):_____				

Figure 1. (continued).

Abstracted from: Lee SJ, et al. *Biol Blood Marrow Transplant*. 2015; 21:984-999.

Appendix B: cGVHD Response Assessment

Response Determination for Chronic GVHD Clinical Trials based on Clinician Assessments

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.

Abstracted from: Lee SJ, Wolff D, Kitko C, et al. *Biol Blood Marrow Transplant* 2015; 21:984 – 999.

Appendix C: cGVHD Activity Assessment-Patient Self Report

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	**As Bad As You Can Imagine** 10
Your **skin itching** at its WORST?	O	O	O	O	O	O	O	O	O	O	O
Your **mouth dryness** at its WORST?	O	O	O	O	O	O	O	O	O	O	O
Your **mouth pain** at its 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

Eyes	What is your main complaint with regard to your eyes?	
	Please rate how severe is this eye symptom, between 0 (not at all severe) and 10 (most severe):	0 1 2 3 4 5 6 7 8 9 10
Vulvovaginal Symptoms (females only)	Do you have any burning, pain or discomfort in the area of your vagina, vulva or labia? **OR** Do you have any discomfort or pain with sexual intercourse?	o Yes o No o Not applicable

Patient Global Ratings:

1. Overall, do you think that your chronic graft versus host disease is mild, moderate or severe?
0= none
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

Attach copies of:

Adults (persons 18 years or older):
-Lee cGvHD Symptom Scale
-Human Activity Profile
-SF-36 v.2
-FACT-BMT

Children/Adolescents (persons 17 years or younger):
-Lee cGVHD Symptom Scale (persons 8-12 years old may complete with help of the health care professional)
-ASK - Activities Scale for Kids
-CHRIs-Generic and Disease Specific Inventory

Abstracted from: 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.

Appendix D: Lee cGVHD Symptom Scale

	Not at all	Slightly	Moderately	Quite a bit	Extremely
SKIN:					
a. Abnormal skin color	0	1	2	3	4
b. Rashes	0	1	2	3	4
c. Thickened skin	0	1	2	3	4
d. Sores on skin	0	1	2	3	4
e. Itchy skin	0	1	2	3	4
EYES AND MOUTH:					
f. Dry eyes	0	1	2	3	4
g. Need to use eyedrops frequently	0	1	2	3	4
h. Difficulty seeing clearly	0	1	2	3	4
i. Need to avoid certain foods due to mouth pain	0	1	2	3	4
j. Ulcers in mouth	0	1	2	3	4
k. Receiving nutrition from an intravenous line or feeding tube	0	1	2	3	4
BREATHING:					
l. Frequent cough	0	1	2	3	4
m. Colored sputum	0	1	2	3	4
n. Shortness of breath with exercise	0	1	2	3	4
o. Shortness of breath at rest	0	1	2	3	4
p. Need to use oxygen	0	1	2	3	4
EATING AND DIGESTION:					
q. Difficulty swallowing solid foods	0	1	2	3	4
r. Difficulty swallowing liquids	0	1	2	3	4
s. Vomiting	0	1	2	3	4
t. Weight loss	0	1	2	3	4
MUSCLES AND JOINTS:					
u. Joint and muscle aches	0	1	2	3	4
v. Limited joint movement	0	1	2	3	4
w. Muscle cramps	0	1	2	3	4
x. Weak muscles	0	1	2	3	4
ENERGY:					
y. Loss of energy	0	1	2	3	4
z. Need to sleep more/take naps	0	1	2	3	4
aa. Fevers	0	1	2	3	4
MENTAL AND EMOTIONAL:					
bb. Depression	0	1	2	3	4
cc. Anxiety	0	1	2	3	4
dd. Difficulty sleeping	0	1	2	3	4

Abstracted from: Lee S, Cook EF, Soiffer R, et al. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant.* 2002; 8:444-452.

Appendix E: Karnofsky Performance Scale

Condition	Percent	Description
Able to carry on normal activity and to work. No special care needed.	100	Normal, no complaints, no evidence of disease.
	90	Able to carry on normal activity; minor signs or symptoms of disease.
	80	Normal activity with effort; some signs or symptoms of disease.
Unable to work. Able to live at home and care for most personal needs. A varying degree of assistance is needed.	70	Cares for self, unable to carry on normal activity or to do active work.
	60	Requires occasional assistance but is able to care for most of his/her needs.
	50	Requires considerable assistance and frequent medical care.
Unable to care for self. Requires equivalent of institutional or hospice care. Disease may be progressing rapidly.	40	Disabled, requires special care and assistance.
	30	Severely disabled, hospitalization indicated. Death not imminent.
	20	Very sick, hospitalization indicated. Death not imminent.
	10	Moribund, fatal processes progressing rapidly.
	0	Dead

Reference: Karnofsky, D.A., and Burchenal, J.H. (1949). The Clinical Evaluation of Chemotherapeutic Agents in Cancer. In Evaluation of Chemotherapeutic Agents: Symposium Held at the New York Academy of Medicine, C.M. MacLeod, ed. (New York, Columbia University Press), pp. 191-205.