STATISTICAL ANALYSIS PLAN

A Phase 2, Randomized, Multicenter Study to Evaluate the Efficacy and Safety of KD025 in Subjects with Chronic Graft Versus Host Disease (cGVHD) After At Least 2 Prior Lines of Systemic Therapy
(The ROCKstar Study)

NCT Number:	NCT03640481
Protocol Number:	KD025-213
Study Drug:	KD025
IND Number:	IND 125890
Phase:	2
Sponsor:	Kadmon Corporation, LLC 450 East 29th Street New York, NY 10016
Version 2.1:	26 June 2020

SIGNATURE PAGE

DOCUMENT HISTORY

Version	Author	Description
1.0	[redacted]	Original Document
2.0	[redacted]	Revision according to "KD025-213_Protocol Amendment 1_26Jun19_FINAL" The primary reasons for this new version include: • Change of study design: add a formal interim analysis with alpha spending .0025 • Change definitions of duration of response in accordance with FDA type B meeting held on 01 May 2019 • Rearrange and add some secondary efficacy endpoints • Add subgroup analysis for proton pump inhibitor
2.1	[redacted]	The primary reasons for this new version include: • Add censoring rule of new systemic cGVHD treatment for organ and overall response, corticosteroids usage and Lee symptom score • Change term of "Sponsor Assessments" into "Kadmon Algorithmic Response Assessments (KARA)" • Add some additional analyses (section 3.8.2.12) • Add analyses of PROMIS for responder and nonresponder populations • In subgroup analyses: add baseline GFR, change age into <65 / ≥65 from by 50 percentile • Clarify response assessment in appendix B • Minor typo and wording changes

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ABBREVIATIONS

7-PtR	≥7 point reduction
ADI	Actual dose intensity
AE	adverse event
ANCOVA	analysis of covariance
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BID	twice daily
BMI	body mass index
CI	confidence interval
CNI	Calcineurin inhibitor
CR	Complete response
CRF	case report form
CSR	clinical study report
DL_{CO}	diffusing capacity of carbon monoxide
DO7-PtR	Duration of a ≥7-PtR
DOR	duration of response
FFS	Failure free survival
FVC	forced vital capacity
GFR	Glomerular filtration rate
GSR	Global Severity Rating
Hb	Hemoglobin
HLGT	high level group term
IA	interim analysis
ICH	International Conference on Harmonisation
ITT	Intent-to-treat
KARA	Kadmon Algorithmic Response Assessments
KPS	Karnofsky Performance Scale
LR	Lack of Response
LR-M	Lack of Response: Mixed
LP-U	Lack of Response: Unchanged
LP-P	Lack of Response: Progression
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent-to-treat
mL	Milliliter

Ms	Millisecond
OS	Overall survival
ORR	Overall response rate
PA	primary analysis
PDI	Planned dose intensity
PE	physical exam
PFTs	pulmonary function tests
PPI	proton pump inhibitor
PR	Partial response
PROMIS	Patient-Reported Outcomes Measurement Information System
PT	preferred term
RDI	Relative dose intensity
QD	once-daily
OMRS	oral mucosa rating scale
QTcF	corrected QT interval using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SOC	system organ class
TEAE	treatment emergent adverse event
TTNT	Time to next treatment
TTR	Time to response
WHO	World Health Organization

1 INTRODUCTION

The objective of the KD025-213 study is to evaluate the efficacy and safety of KD025, at dose levels of 200mg QD and 200mg BID, in subjects with cGVHD who have previously been treated with two to five prior lines of systemic therapy.

This Statistical Analysis Plan (SAP) describes data-handling and statistical procedures to be used for the analysis and reporting of efficacy and safety data collected under Study KD025-213 and presented in the clinical study report (CSR). This SAP has been developed and finalized prior to locking the clinical database. Pharmacokinetic (PK) and Pharmacodynamic (PD) analyses will be described in separate documents.

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

Primary objective:

- To evaluate the efficacy (overall response rate (ORR) as assessed by investigator) and safety of KD025, at dose levels of 200mg QD and 200mg BID, in subjects with cGVHD who have previously been treated with two to five prior lines of systemic therapy.

Secondary objectives:

- Duration of response (DOR)
- Changes in the Lee Symptom Scale Score
- Response by organ system
- Time to response
- Time to next treatment (TTNT)
- Percentage of subjects who have a best response of partial response (PR) and percentage of subjects who have a best response of complete response (CR)
- Change in corticosteroid dose
- Change in calcineurin inhibitor dose
- Failure-free-survival (FFS)

- Overall survival (OS)
- Change in cGVHD severity using the Clinician-Reported Global cGVHD Activity Assessment
- Change in symptom activity using the cGVHD Activity Assessment Patient Self-Report
- PK of KD025 in subjects with cGVHD

Exploratory Objectives:

- To evaluate changes in the PROMIS Global Health sub-scores for physical and mental functioning
- To evaluate ORR using sponsor assessments of overall response
- To evaluate changes in relevant biomarkers after KD025 administration

2.2 Study Design

KD025-213 is a Phase 2, open label, randomized, multicenter study in subjects with cGVHD who have previously been treated with two to five prior lines of systemic therapy. The study schematic is shown below in Figure 1. Approximately 126 subjects with active cGVHD will be randomized (1:1) to receive treatment with one of two KD025 regimens:

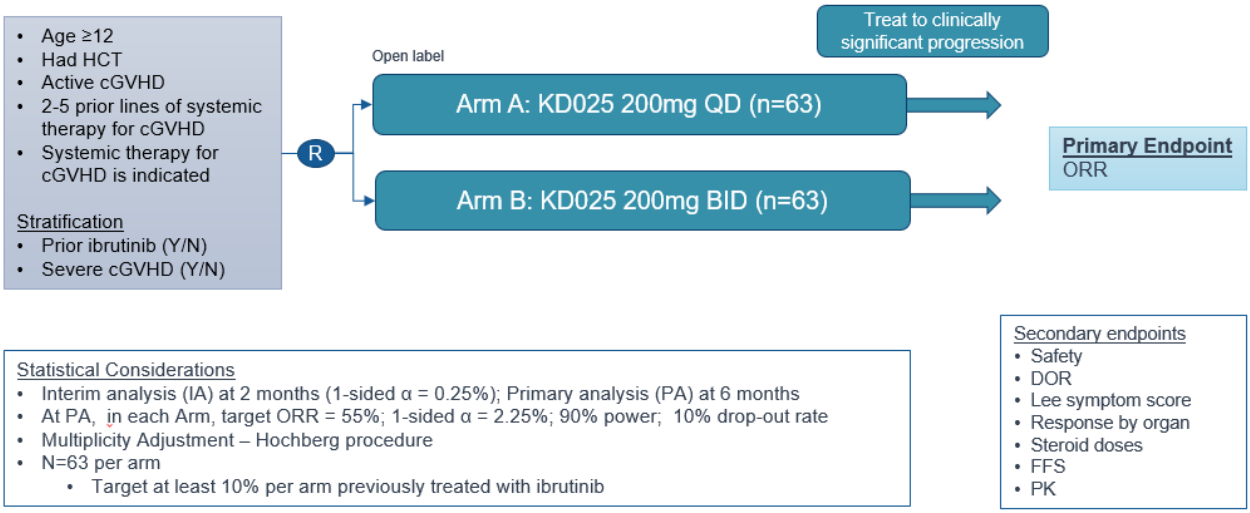
- Arm A: KD025 200mg QD
- Arm B: KD025 200mg BID

Subjects may receive treatment in 28-day treatment cycles until clinically significant progression of cGVHD (defined as progression requiring addition of new systemic therapy for cGVHD), histologic recurrence of underlying malignancy, unacceptable toxicity, Investigator decision, subject preference / withdrawal of consent, loss to follow-up, sponsor decision, or death (whichever occurs first) Subjects who have not achieved a response after 12 cycles of KD025 should be withdrawn if in the Investigator's judgment there is no evidence of clinical benefit.

Subjects will undergo evaluations as outlined in the Schedule of Assessments (see Table 12 in KD025-213 Study Protocol).

The primary endpoint is the ORR with responses as defined by the 2014 National Institute of Health (NIH) Consensus Development Project on clinical trials in cGVHD.

Figure 1 KD025-213 Study Schema



2.3 Number of Subjects

Approximately 126 subjects will be enrolled.

2.4 Randomization

Subjects will be randomized to either Arm A (200mg QD) or Arm B (200mg BID) with 1:1 ratio and block size 4. Randomization will be stratified according to:

- Prior ibrutinib therapy (Yes / No)
- Severe cGVHD (yes/no) where severe cGVHD is defined as at least one organ with a score of 3, or a lung score of 2 or 3

2.5 Replacements

Randomized subjects withdrawn from the study before receiving any study drug will be replaced.

3 STATISTICAL METHODS

Three analyses are planned:

- An interim analysis (IA) will be conducted approximately 2 months after 126 subjects have been enrolled into the mITT population. A nominal 1-sided alpha of 0.0025 will be spent, but there will be no early study termination for efficacy.
- The primary analysis (PA) will be conducted approximately 6 months after 126 subjects have been enrolled into the mITT population, with 1-sided alpha 0.0225 (or 0.025 if the ORRs of both arms are significant at interim).
- A follow-up analysis will be conducted approximately 12 months after 126 subjects have been enrolled into the mITT population.

Alpha will only be allocated to the primary endpoint, ORR.

3.1 Sample Size Justification

The sample size for KD025-213 is based upon the following considerations:

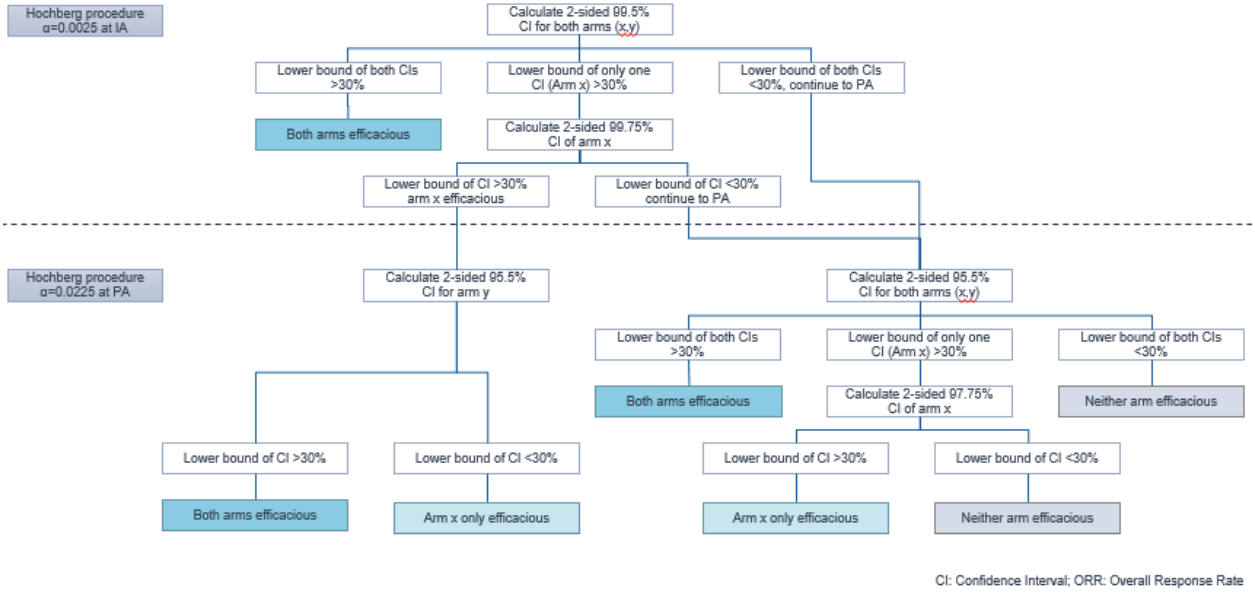
- The primary efficacy objective is to demonstrate an ORR > 30%, i.e. that the lower bound of the CI of ORR is greater than 30%
- There is an interim analysis with .0025 1-sided alpha spending
- A true ORR of 55% is assumed based upon data from an ongoing study of KD025 (KD025-208) in patients with cGVHD after 1-3 prior lines of systemic therapy. In the KD025-208 study, the ORR was 65% in Cohort 1 (200mg QD), 63% in Cohort 2 (200mg BID) and 52% in Cohort 3 (400mg QD) as of 13-September, 2018.
- A 10% dropout rate is assumed based upon data from the ongoing KD025-208 study (estimated ~6% dropout rate in KD025 208). Dropouts are defined as a subjects discontinuing KD025 treatment prior to any response assessment, due to reasons other than an AE related to study drug or cGVHD progression

For a single arm, assuming a true ORR of 55%, dropout rate of 10%, power 90%, and 2-side alpha of .045 to demonstrate ORR > 30%, the sample size is calculated to be 63.

3.2 Multiplicity Adjustment

Since the study has two KD025 treatment arms, the Hochberg procedure will be used for multiplicity adjustment for the primary endpoint as shown in Figure 2.

Figure 2 ORR Multiplicity Adjustment



Simulations were conducted with various hypothetic ORR scenarios for two arms of 63 subjects, by using the Hochberg procedure for multiplicity adjustment; results (10000 runs per setting) are presented in Table 1.

Only descriptive statistics will be provided for all secondary and exploratory endpoints, without multiplicity adjustment.

3.3 Sample Size Adjustment

If during the course of the current study the dropout rate is observed to exceed 10%, the sample size may be adjusted to maintain overall power e.g., to approximately 85 subjects per arm if the dropout rate reaches 15%.

3.4 Analysis Populations

mITT: The primary population for efficacy analyses will be a modified Intent to treat (mITT) population defined as all randomized subjects who receive at least one dose of study medication.

Table 1: Power for various hypothetic ORRs

Table 1a: IA Power (1-sided alpha 0.0025)*						
For both arms to be statistically significant						
		Arm 1 ORR (%)				
		40	**45**	**50**	**55**	**60**
Arm 2 ORR (%)	**40**	< 1	< 1	< 1	< 1	< 1
	45	-	< 1	1	2	3
	50	-	-	2	6	9
	55	-	-	-	12	22
	60	-	-	-	-	37
For at least one arm to be statistically significant						
		Arm 1 ORR (%)				
		40	**45**	**50**	**55**	**60**
Arm 2 ORR (%)	**40**	1	3	10	28	51
	45	-	5	12	29	53
	50	-	-	20	35	57
	55	-	-	-	65	65
	60	-	-	-	-	77
Table 1b: PA Power (1-sided alpha 0.0225)*						
		Arm 1 ORR (%)				
		40	**45**	**50**	**55**	**60**
Arm 2 ORR (%)	**40**	20	39	64	84	95
	45	-	52	72	89	97
	50	-	-	83	93	98
	55	-	-	-	97	99
	60	-	-	-	-	100
IA = Interim Analysis; PA = Primary Analysis * Power estimated based upon 10,000 simulations for each cell						

The mITT population will be used for tables of demography, baseline characteristics and efficacy.

Responder: The responder population is defined as subjects in the mITT population that achieved a partial or complete response at any post-baseline response assessment.

Nonresponder: The nonresponder population is defined as anyone in mITT population that is not a responder.

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

Safety: The safety population is defined as all randomized subjects who receive at least one dose of study medication. In this study, safety population is equivalent to mITT population.

3.5 Protocol Deviations

All protocol deviations will be identified and classified as major or minor (as defined below) before the 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 noncompliance 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.

3.6 Subject Disposition

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

3.7 Demographics and Baseline Characteristics

3.7.1 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.

3.7.2 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.

3.7.3 GVHD and Transplant History

The GVHD history will be summarized including transplant history and prior therapies.

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 CMV serostatus

The baseline characteristics of GVHD will be summarized for:

- GVHD prophylaxis
- Prior aGVHD
- Time from transplant to cGVHD diagnosis
- Time from cGVHD diagnosis to enrollment
- Time from transplant to enrollment
- cGVHD severity: mild / moderate / severe

Prior Therapies

- Number of prior lines of systemic cGVHD therapy
- Ongoing systemic cGVHD therapies at baseline

- Listing of all non-systemic cGVHD therapies

Organ Involvement

- Number of organs involved at baseline
- Organ involvement at baseline, Including
    - baseline skin NIH score and skin feature score
    - eye NIH score
    - NIH modified oral mucosa rating scale (OMRS) score
    - esophagus NIH score
    - upper GI NIH score
    - lower GI NIH score
    - alanine transaminase (ALT), alkaline phosphatase (ALP), total bilirubin
    - lung NIH score and FEV1 (% predicted)
    - joint/fascia NIH score and P-ROM score
    - global severity rating

3.7.4 Other Baseline Characteristics

Other baseline values including Karnofsky Performance Scale (KPS), child bearing potential, will also be summarized.

3.8 Efficacy Endpoints and Analyses

3.8.1 Primary Efficacy Endpoint: ORR

The primary efficacy endpoint is the overall response rate, with response status as assessed by investigators based upon the 2014 NIH Consensus Development Project on Clinical Trials in cGVHD (see Protocol Appendix F: cGVHD Response Assessment).

Response is assessed using individual scores from ten systems (Skin, Eyes, Mouth, Esophagus, Upper GI, Lower GI, Liver, Lungs, Joints and Fascia and Global Severity Rating (GSR)). Response is assessed with respect to the baseline (Cycle 1, Day 1 (C1D1)) cGVHD assessment. The overall response at each assessment time point will be categorized as Complete Response (CR), Partial Response (PR), or Lack of Response (LR), where LR includes the response status of unchanged, mixed, or progression as defined in Table 2.

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

If a treated subject is lost to follow up without response assessment, this subject will be counted as a non-responder. Any response on and after new systemic cGVHD treatment will be censored as non-response.

Point estimates, various confidence intervals (Clopper-Pearson (exact) method) listed in Figure 2, and raw and Hochberg adjusted p-values corresponding to the null hypothesis of ORR ≤ 30% vs alternative hypothesis of ORR > 30% by treatment arms will be reported.

The number and percentage of subjects who have a best response of PR and number and percentage of subjects who have a best response of CR will also be reported.

Table 2: cGVHD Response Definitions

Response	**Definition**
Complete Response (CR)	Resolution of all manifestations of cGVHD in each organ or site
Partial Response (PR)	Improvement in at least one organ or site without progression in any other organ or site
Lack of Response (LR)	
Mixed (LR-M) *	Complete or partial response in at least one organ accompanied by progression in another organ
Unchanged (LR-U)	Outcomes that do not meet the criteria for complete response, partial response, progression or mixed response
Progression (LR-P)	Progression in at least one organ or site without a response in any other organ or site

* Considered progression for purposes of analysis

3.8.2 Secondary Efficacy 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.

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 3 will be applied
- Landmark analyses: Number and percentage of subjects with response sustained for ≥ 12, ≥ 20, ≥ 24, ≥ 32, ≥ 36, and ≥ 48 weeks

Table 3. 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 Lee Symptom Scale Score (LSS)

Lee cGVHD Symptom Scale[1] (see questionnaire in Protocol 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 30 items of 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 nonmissing domain scores if more than 50% of them are nonmissing. A higher score indicated more bothersome symptoms. 7 points difference on the summary score of cGVHD symptom scale was found to be clinically meaningful.

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

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

In the above formulas, m is the number of items have nonmissing value in a domain, and n is the number of domains have nonmissing domain scores.

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 percent of subjects with a ≥7-point reduction (7-PtR) from baseline (C1D1)
- Number and percent 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 nonresponder populations. 7-PtR and DO7-PtR will be censored after new systemic cGVHD treatment.

3.8.2.3 Response by organ system

The best response (CR, PR) for the nine individual organs (Skin, Eyes, Mouth, Esophagus, Upper GI, Lower GI, Liver, Lungs, and Joints and fascia) plus global severity rating will be summarized.

Time to response at the organ level will also be evaluated. Descriptive statistics and plots of cumulative number and % of responders over time (4, 8, 12, 16, 24, 32, 40, >=48 weeks) will be provided. Two series of percentages will be presented:

1. with total number of subjects in mITT with involvement of the given organ at baseline as denominator
2. with number of subjects in the responder population with involvement of the given organ at baseline as denominator.

3.8.2.4 Time to Response (TTR)

TTR 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 (4, 8, 12, 16, 24, 32, 40, ≥48 weeks) will be provided. TTR analyses will only be conducted for the responder population.

3.8.2.5 Time to Next Therapy (TTNT)

The TTNT will be measured as the time from first treatment to the time of new systemic cGVHD treatment, 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.

3.8.2.6 Corticosteroid dose

Corticosteroid doses will be presented as mg/kg/day prednisone equivalent dose. Descriptive statistics for the mITT population, responder /nonresponder populations and subgroups defined by baseline corticosteroid dose level (upper and lower 50 percentiles) will be provided for:

- Systemic corticosteroid dose over time
- Change and % change from baseline (C1D1) to the greatest corticosteroid dose reduction during KD025 treatment period
- Number and % of subjects who reduced systemic corticosteroid dose during KD025 treatment period

- Number and % of subjects who ever discontinued systemic corticosteroid usage for over 28 days during KD025 treatment period

The above analyses will be censored on or after new systemic cGVHD treatment.

If subjects are not using prednisone as the systemic corticosteroid, then the prednisone dose equivalent will be determined according to following conversion ratios[2]:

1 mg prednisone is equivalent to:

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

Transient increases in corticosteroid dosing (not exceeding 1mg/kg/day prednisone equivalent, are permitted for the treatment of cGVHD flare, but dose must be reduced back to the pre-randomization dose within 6 weeks. If the dose remains elevated for more than 6 weeks, this will be considered a KD025 treatment failure. More than 2 episodes of cGVHD flare requiring increased corticosteroid therapy in the first 6 months of KD025 treatment will also be considered a KD025 treatment failure.

3.8.2.7 Calcineurin inhibitor (CNI) doses

CNI includes systemic tacrolimus and cyclosporine. Descriptive statistics will be provided for:

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

3.8.2.8 Failure-free-survival (FFS)

FFS is defined as the absence of new cGVHD systemic therapy, non-relapse mortality and recurrent malignancy (i.e. underlying disease) and is censored by by last response assessment or long term follow up assessment, whichever is the latest and available. Kaplan-Meier plots, descriptive statistics of FFS and landmark FFS at 6, 12, 18 and 24 months will be provided. In addition, analyses for the three components of FFS will also be provided.

3.8.2.9 Overall survival (OS)

OS is defined as time from first dose of KD025 to the date of death due to any cause. Kaplan-Meier plots, descriptive statistics of OS and landmark OS at 6, 12, 18 and 24 months will be provided.

3.8.2.10 Change in cGVHD severity using 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.

3.8.2.11 Change in symptom activity using the cGVHD Activity Assessment Patient Self-Report

The cGVHD Activity Assessment-Patient Self Report is in Protocol Appendix F. 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.

In addition to the specific analyses described above, descriptive statistics by visit and arm will be provided for all applicable secondary endpoints.

3.8.2.12 Additional Analyses

Additional analyses will be presented including:

- ORR and DOR for responses occurring within the first 6 and 12 months of treatment with KD025
- Organ responses according to baseline NIH / severity score
- Landmark FFS with CR/PR at 6 months and at 12 months
- Time to corticosteroid discontinuation

3.8.3 Exploratory Endpoints

3.8.3.1 Changes in the PROMIS Global Health subscores for physical and mental functioning (see definition in Protocol Appendix E)

Patient-Reported Outcomes Measurement Information System (PROMIS)[3], a state-of-the-science PRO measurement system, was developed using mixed qualitative and quantitative methods and uses item response theory–calibrated item banks for numerous patient-reported symptoms and functional domains. These PRO measures can be used across chronic diseases and in the general population. A feature of PROMIS measures is that an individual's or group's score is represented as a T score, normalized and calibrated against the US population (United States population average score, 50; 10 points = 1 standard deviation [SD]). The PROMIS Global Health 10 measure comprises ten items with two summary scores for physical and mental functioning with higher scores indicating better functioning. Those two summary scores were calculated using the Assessment Center Scoring Service (https://www.assessmentcenter.net/ac_scoringservice)from raw data.

Analyses will include:

- Both raw scores and change-from-baseline values (physical and mental domains) will be summarized as continuous variables by visit.
- Number of subjects with a ≥4.7-point reduction from baseline (C1D1)

These analyses will be performed on mITT, responder and nonresponder populations.

3.8.3.2 ORR using Kadmon Algorithmic Response Assessment (KARA)

In addition to ORR assessed by investigators, the sponsor will conduct response assessments according to criteria listed in Appendix B for each individual organ, then assign overall response according to Table 2 (cGVHD response definitions). Analyses of ORR and DOR will be repeated with this sponsor assessment of overall response.

Concordance with investigator determined responses will be analyzed.

3.8.4 Subgroup Analyses

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

- Prior ibrutinib (Yes / No)
- Severe cGVHD at screening (Yes / No)
- Number of organs involved at baseline (<4 / ≥4)
- Number of prior lines of therapy (≤3 / >3)
- Duration of cGVHD before enrollment (by 50th percentile)
- Baseline corticosteroid dose level (by 50th percentile)
- Lung involvement at baseline (Yes / No)
- Take concomitant medication proton pump inhibitor (PPI) on C1D1 (Yes/No)
- Gender
- Age (<65, ≥65 years))
- Race
- Prior ruxolitinib (Yes / No)
- Best response to the last prior treatment (SD/PD)
- Take concomitant medication proton pump inhibitor (PPI) on C1D1 (Yes/No)
- Baseline GFR (<60, 60-90 and ≥90 mL/min)

All subgroup analyses will be exploratory and with no multiplicity adjustment.

3.8.5 Multivariate Modeling

Logistic regression will be carried out as exploratory/supportive analyses. This will provide a means to assess the effects of those potential prognostic indicators listed in subgroup analyses section. The following modeling procedure will be used:

- **Univariate modeling**

A logistic regression model (response ~ an individual factor) will be used for a factor of arm, potential prognostic factors listed in subgroup analyses section and other potential prognostic indicators. The resulting odds ratio, its confidence interval (CI) and p value will be summarized in a univariate analysis table.

- **Multivariate modeling**

Based on the univariate modeling results and clinical judgement, a multivariate logistic regression model might be chosen as a predictive model.

3.9 Concomitant Medications

Concomitant medications will be coded using WHO Drug Dictionary Enhanced. Concomitant medications are all medications taken after the start of study consent, during the treatment period (day 1 until 28 days after last dose), including those started before but ongoing at the start of study treatment.

Concomitant medications will be summarized according to ATC class and preferred drug name.

3.10 Treatment Compliance and Exposure

The extent of KD025 exposure will be assessed using the duration of KD025 treatment. Pill dispensation and return counts will be the primary source of data.

Duration of KD025 treatment is defined as last dose date – first dose date + 1 day regardless of unplanned intermittent discontinuations. Duration of treatment will be summarized as a continuous variable and also categorically by numbers and percentages for each of the following categories and cumulatively according to these categories: 0 to 6, 6 to 12, and ≥12 months.

The cumulative duration of treatment will be provided, defined as the sum of the duration of treatment for all patients, and will be expressed in patient years.

The relative dose intensity (RDI) will be presented. The RDI is defined as:

RDI (%) = 100*ADI (mg/day)/PDI (mg/day)

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

PDI (mg/day) = planned cumulative dose (mg) / duration of treatment (days),

ADI (mg/day) = actual cumulative dose (mg) / duration of treatment (days).

The planned cumulative dose is the planned daily dose amount * duration of treatment, while the actual cumulative dose is the sum of actual doses received over the duration of treatment. The actual total dose will incorporate information of dose reductions and dose interruptions captured in the CRF.

Descriptive summary statistics of the treatment duration, actual cumulative dose, ADI, RDI, dose reduction and dose interruption will be presented.

3.11 Safety Analysis

Safety assessments include AEs, serious adverse events (SAEs), vital sign measurements, clinical laboratory evaluations (hematology, chemistry) and ECGs. Clinically significant PE findings will be captured as AEs.

Safety analyses will further include assessments of the defined stopping rules (protocol section 10), namely rates of:

1. Secondary graft failure
2. Histological recurrence of underlying malignancy within 6 months of randomization
3. Withdrawal due to related AEs within 3 months of randomization

3.11.1 Adverse Events

AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) dictionary (Version 20.0 or higher). TEAEs are any AEs occurring or worsening in severity after the first administration of study medication. All AEs (including SAEs) will be graded using the 5-point CTCAE V5.0 scale (mild, moderate, severe, life threatening, or death); their causality with KD025 will be classified into following 5 classes: definitely related, probably related, possibly related, unlikely related, not related. The Investigator will further assess the relationship of adverse events 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 adverse events of each specific system organ class (SOC) and preferred term (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.

TEAEs, Grade ≥3 TEAEs, SAEs and TEAEs leading to dose modification / discontinuation will be summarized according to treatment arm, system-organ-class, and preferred terms. 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.

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

3.11.2 Adverse events of Note

AEs of note include:

- Hepatic AEs by CTCAE V5.0 criteria; grouped terms will also be presented.
- Bacterial infection, Viral infection and Fungal infection
- Cytopenias

Those AEs of note will be summarized according to treatment group and preferred terms. Also see Section 3.11.3 for LFT analyses.

3.11.3 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 WCC, neutrophils (absolute count), lymphocytes (absolute count), monocytes (absolute count), Hb, platelets, ALP, ALT, AST, GGT, total bilirubin, GFR, 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 V5.0. All graded laboratory parameters will be summarized separately for hematology and biochemistry. Corresponding shift tables to compare baseline to the worst post-baseline grade within the treatment period will be provided.

3.11.4 Vital Signs and Other Physical Findings

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

3.11.5 ECG

Descriptive statistics for ECG parameters (i.e., heart rate (HR), PR interval, RR interval, QRS interval, QT interval, and QTcF interval) at each time point with triplicate ECGs will be presented for the values and change from baseline scores (note: Fridericia's correction: 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:

- ≤ 450 ms
- >450 to ≤ 480 ms
- >480 to ≤ 500 ms
- > 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:

- ≤ 0 ms
- >0 to ≤ 30 ms
- >30 to ≤ 60 ms
- > 60 ms

3.11.6 Physical Examination

Clinically significant physical examination findings will be captured as AEs.

3.12 Pharmacokinetic Analysis

Further details will be described in the Pharmacokinetic Analysis Plan.

3.13 Pharmacodynamics Analysis

Further details will be described in the Pharmacodynamics Analysis Plan.

3.14 General Methods

Computing Environment

All statistical analyses will be performed using SAS® Version 9.3 or higher for Windows.

Reporting of Numerical Values

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 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 specified analysis population or group.

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

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

Baseline

Baseline value is defined as the valid and last non-missing value obtained within 14 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.

Handling of Incomplete or Missing Data

Missing data will not be imputed in general and it 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 listings.

If the assessment of the relationship of the AE to KD025 is missing, then the relationship to KD025 has to be assumed and the AE considered as such in the frequency tables of possibly related AEs, but 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.

4 REFERENCES

1. Lee SJ, Cook EF, Soiffer R, et al. 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

2. Liu D, Ahmet A, Ward L, et al. A practical guide to the monitoring and management of the complications of systemic corticosteroid therapy. Allergy Asthma Clin Immunol. 2013;9(1):30. Published 2013 Aug 15. doi:10.1186/1710-1492-9-30

3. Lee SJ, Onstad L, Chow EJ, et al. Patient-reported outcomes and health status associated with chronic graft-versus-host disease. Haematologica. 2018; 103(9), 1535-1541

5 APPENDIX A: Tables, Listings and Figures (TLFs)[1]

Table 3: Demographics and Baseline Characteristics TLFs

T/F/L	Title	Population
T	Patient disposition	mITT
L	Patient disposition	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
L	Prior systemic cGVHD therapy	mITT

Table 4: Efficacy TLFs

T/L/F	Title	Population
Primary Endpoint: ORR		
T	Overall Response Rate*	mITT
T	Overall response rate (ORR) for responses occurring within 6 months of treatment	mITT
T	Overall response rate (ORR) for responses occurring within 12 months of treatment	mITT
T	Landmark ORR at 6 months	mITT
T	Landmark ORR at 12 months	mITT
F	Forest plot of ORRs (1. all subjects in an arm, subgroups within arm; 2. Two arm combined if ORRs of those two arms are similar)	mITT

[1] In all tables, figures and listing, results will be presented by two treatment arms and overall, unless be otherwise specified.

T/L/F	Title	Population
L	Listing of overall response by subject by visit	mITT
Secondary Endpoint: DOR		
T	Descriptive Kaplan-Meier and landmark (12, 20, 24, 32, 36, 48 weeks) statistics for DOR*	Responder
F	Kaplan-Meier plot of DOR*	Responder
L	Listing of all DOR episodes for each subject	Responder
Secondary Endpoint: TTR		
T	Descriptive statistics for cumulative response rate over time (TTR)	Responder
F	Cumulative response rate over time (TTR)	Responder
L	Listing of TTR for each subject	Responder
Secondary Endpoint: LSS score		
T	Descriptive statistics for LSS score for: a) number of subjects with a 7-PtR from baseline; b) number of subjects with 7-PtR on 2 consecutive assessments; c) DO7-PtR (<8, 8-16, 16-24, 24-32, 32-40, 40-48, >48 weeks)	mITT, Responder, Nonresponder
T	Summary of LSS Score (overall and each of seven domains) raw and change from baseline by visit	mITT, Responder, Nonresponder
F	Summary of Lee Symptom Scale Score (overall and each of seven domains) raw and change from baseline by visit	mITT, Responder, Nonresponder
L	Listing of Lee Symptom Scale Score by subject, domain and visit	mITT
Secondary Endpoint: Response by Organ System		
T	Best response by individual organ	mITT
T	Organ responses according to baseline NIH / severity score	mITT
L	Listing of organ score and response assessment by subject and visit	mITT
T	Descriptive statistics for cumulative response rate over time (TTR) by organ	mITT, Responder
F	Cumulative response rate over time (TTR) by organ	mITT, Responder
Secondary Endpoint: Corticosteroid Dose		
T	Descriptive statistics for prednisone equivalent dose of corticosteroids for a) change and % change from baseline to greatest reduction; b) the number and % subjects who reduced dose; c) number and % subjects who discontinued corticosteroid usage	mITT, Responder, Nonresponder, BL corticosteroid dose level (upper and lower 50th percentiles)

T/L/F	Title	Population
T	Descriptive Kaplan-Meier statistics for time to corticosteroid discontinuation	mITT
T	Summary of raw and change from baseline on prednisone equivalent dose of corticosteroids by visit	mITT, Responder, Nonresponder, BL corticosteroid dose level (upper and lower 50th percentiles)
F	Box plot of prednisone equivalent dose of corticosteroids over time	mITT, Responder, Nonresponder, 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 for a) the number and % subjects who reduced dose from baseline; b) number and % subjects who discontinued CNI	mITT
L	Listing of all the CNI dosing by subject and visit	mITT
Secondary Endpoint: FFS		
T	Descriptive Kaplan-Meier and landmark (6, 12, 18 and 24 months) 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: OS		
T	Descriptive Kaplan-Meier and landmark (6, 12, 18 and 24 months) statistics for OS	mITT
F	Kaplan-Meier plot of OS	mITT
Secondary Endpoint: Symptom activity using the cGVHD Activity Assessment Patient Self-Report		
T	Descriptive statistics of change from baseline in symptom activity based on cGVHD Activity Assessment Patient Self Report (global rating) by visit	mITT

T/L/F	Title	Population
T	Descriptive statistics of symptom activity based on cGVHD Activity Assessment Patient Self Report (change of cGVHD symptom to a month ago) by visit	mITT
L	Listing of symptom activity based on cGVHD Activity Assessment Patient Self Report	mITT
Exploratory Endpoints		
T	Overall Response Rate (KARA)	mITT
T	Concordance of overall response between investigator determined responses and KARA	mITT
T	Descriptive Kaplan-Meier and landmark (12, 20, 24, 32, 36, 48 weeks) statistics for DOR (KARA)	Responder
F	Kaplan-Meier plot of DOR (KARA)	Responder
T	Best response by individual organ (KARA)	mITT
T	Descriptive statistics of change from baseline in PROMIS scale by visit	mITT, Responder, Nonresponder
T	Summary table of univariate modeling	mITT

* Repeat TLF for subgroups of:

- Prior ibrutinib (Yes / No)
- Severe cGVHD at baseline (Yes / No)
- Number of organs involved at baseline (<4 / ≥4)
- Number of prior lines of therapy (≤3 / >3)
- Duration of cGVHD before enrollment (by 50th percentile)
- Baseline corticosteroid dose level (by 50th percentile)
- Lung involvement at baseline (Yes / No)
- Proton pump inhibitor (PPI) on C1D1 (Yes/No)
- Gender
- Age (<65 / ≥65 years)
- Race
- Prior ruxolitinib (Yes / No)
- Best response to the last prior treatment (SD/PD)
- Take concomitant medication proton pump inhibitor (PPI) on C1D1 (Yes/No)
- Baseline GFR (<60, 60-90 and ≥90 mL/min)

Table 5: Safety TLFs

T/F/L	Title	Population

T	Concomitant medications	mITT
L	Concomitant medications	mITT
T	Treatment exposure and compliance	mITT
L	Treatment exposure and compliance	mITT
T	Dose modifications and interruption	mITT
L	Dose modifications and interruption	mITT
T	Overall summary of TEAEs	Safety
T	TEAEs by Soc/PT	Safety
T	TEAEs by PT	Safety
T	TEAEs by severity and PT	Safety
T	Treatment related AEs by PT	Safety
T	Serious TEAEs by Soc/PT	Safety
T	Serious TEAEs by PT	Safety
T	Treatment related SAEs by PT	Safety
T	Grade 3-5 TEAEs by Soc/PT	Safety
T	Grade 3-5 TEAEs by PT	Safety
T	Treatment related grade 3-5 AEs by Soc/PT	Safety
T	TEAEs leading to Study treatment discontinuation by Preferred Term	Safety
T	Related TEAEs leading to Study treatment discontinuation by Preferred Term	Safety
T	TEAEs leading to dose interruption by Preferred Term	Safety
T	Related TEAEs leading to dose interruption by Preferred Term	Safety
T	TEAEs leading to dose modification by Preferred Term	Safety
T	Related TEAEs leading to dose modification by Preferred Term	Safety
T	TEAEs leading to dose reduction by Preferred Term	Safety
T	Related TEAEs leading to dose reduction by Preferred Term	Safety
L	AEs	Safety
L	AE Leading to treatment discontinuation	Safety
L	Death	Safety
T	Lab values and their change from baseline	Safety
F	Lab values by visit	Safety
T	Shifts in CTC grade from baseline to highest grade post-baseline of laboratory	Safety
L	Lab values with CTC grade ≥ 3	Safety
T	Vital signs and their change from baseline	Safety
L	Vital signs	Safety
T	ECG parameters and their change from baseline	Safety

F	ECG parameters by visit	Safety
T	Number and percentage of subjects with QTcF by visit a) absolute ranges and b) change from baseline	Safety
T	QTcF abnormalities	Safety
L	ECG values	Safety

6 APPENDIX B: Kadmon Algorithmic Response Assessments

For determination of responses and DOR per sponsor assessment, criteria for baseline organ involvement and response assessments are described below.

Baseline Organ Involvement

Organ	Organ Involvement
Esophagus	NIH Esophagus Score ≥ 1, not entirely explained by non-cGVHD cause
Upper GI	NIH Upper GI Score ≥ 1, not entirely explained by non-cGVHD cause
Lower GI	NIH Lower GI Score ≥ 1, not entirely explained by non-cGVHD cause
Lungs	FEV1 <75% predicted, not entirely explained by non-cGVHD cause If no PFTs available, use NIH lung score ≥ 1, not entirely explained by non-cGVHD cause
Eyes	NIH Eye Score ≥ 1, not entirely explained by non-cGVHD cause
Joints and Fascia	P-ROM Score <25 (< maximal score) or NIH Joints/Fascia Score ≥ 1, not entirely explained by non-cGVHD cause
Skin	NIH Skin Score ≥ 1, not entirely explained by non-cGVHD cause
Mouth	NIH Modified Oral Mucosal Rating Score ≥ 1, not entirely explained by non-cGVHD cause
Liver	Elevation > 2x ULN of one or more (ALT, ALP, Total Bilirubin), not entirely explained by non-cGVHD cause

Response Assessments

Organ	Complete Response	Partial Response	Unchanged	Progression	Non Evaluable
Esophagus	NIH Esophagus Score 0 after previous involvement	Decrease in NIH Esophagus Score by 1 or more points	No change or increase from 0 to 1 in NIH Esophagus Score	Increase in NIH Esophagus Score by 1 or more points, except 0 to 1	NIH esophagus score is Non evaluable
Upper GI	NIH Upper GI Score 0 after previous involvement	Decrease in NIH Upper GI Score by 1 or more points	No change or increase from 0 to 1 in NIH Upper GI Score	Increase in Upper GI Score by 1 or more points, except 0 to 1	NIH upper GI score is Non evaluable
Lower GI	NIH Lower GI Score 0 after previous involvement	Decrease in NIH Lower GI Score by 1 or more points	No change or increase from 0 to 1 in NIH Lower GI Score	Increase in Lower GI Score by 1 or more points, except 0 to 1	If NIH Lower GI score is Non evaluable
Lungs	%FEV1 ≥ 80% after previous involvement; or NIH Lung Symptom Score 0 after previous involvement if PFTs not available	Increase by 10% absolute value of %FEV1; or decrease in NIH Lung Symptom Score by 1 or more points if PTSs not available	Not in the other categories	Decrease by 10% absolute value of %FEV1 and %FEV1 < 75%; or increase in NIH Lung Symptom Score by 1 or more points, except 0 to 1 if PFTs not available	NIH Lung score is Non evaluable and %FEV1 is missing
Eyes	NIH Eye Score 0 after previous involvement	Decrease in NIH Eye Score by 1 or more points	No change or increase from 0 to 1 in NIH Eye Score	Increase in NIH Eye Score by 1 or more points, except 0 to 1	NIH eye score is Non evaluable
Joints and Fascia	NIH Joint and Fascia Score 0 and P-ROM score 25 after previous	Decrease in NIH Joint and Fascia Score by 1 or more points or increase in P-	Not in the other categories	Increase in NIH Joint and Fascia Score by 1 or more points or decrease in P-	Both in NIH Joint and Fascia and P-ROM scores are Non

Organ	Complete Response	Partial Response	Unchanged	Progression	Non Evaluable
	involvement in at least one measure	ROM score by 1 or more points for any site		ROM score by 1 or more points for any site	evaluable
Skin	NIH Skin Score 0 after previous involvement	Decrease in NIH Skin Score from baseline by 1 or more points	No change or increase from 0 to 1 in NIH Skin Score	Increase in NIH Skin Score from baseline by 1 or more points, except 0 to 1	NIH skin score is Non evaluable
Mouth	NIH Modified OMRS 0 after previous involvement	Decrease in NIH Modified Oral Mucosa Rating Score of 2 or more points	No change in NIH Modified Oral Mucosa Rating Score	Increase in NIH Modified Oral Mucosa Rating Score of 2 or more points	NIH mouth score is Non evaluable
Liver	Normal ALT, ALP and total bilirubin after elevation of one or more at baseline	50% improvement in ALT, ALP, or total bilirubin without increase by 2x ULN for any other parameters of ALT, ALP or total bilirubin	Not in the other categories	Increase by 2x ULN for any one of ALT, ALP or total bilirubin	If ALT, ALP and total bilirubin are missing
Global	GSR 0	GSR decreases by 2 or more points	Not in the other categories	GSR increases by 2 or more points	Not applicable

Organs not involved at baseline can progress according to the definitions in this Response Assessments table, but cannot achieve a response (PR or CR).