STATISTICAL ANALYSIS PLAN

Protocol Title:	AGAVE-201, Phase 2, Open-label, Randomized, Multicenter Study to Evaluate the Efficacy, Safety and Tolerability of Axatilimab at 3 Different Doses in Patients with Recurrent or Refractory Active Chronic Graft Versus Host Disease who have Received at least 2 Lines of Systemic Therapy
Protocol Number:	SNDX-6352-0504
Protocol Version/Date:	Version 5.0, 05 April 2023 GLOBAL
Investigational Product:	Axatilimab (SNDX-6352)
Sponsor:	Syndax Pharmaceuticals, Inc. 35 Gatehouse Drive Building D, Floor 3 Waltham, MA 02451
SAP Version/Date:	SAP V4.0/27 June 2023

SIGNATURE PAGE

Protocol Title:	AGAVE-201, A Phase 2, Open-label, Randomized, Multicenter Study to Evaluate the Efficacy, Safety and Tolerability of Axatilimab at 3 Different Doses in Patients with Recurrent or Refractory Active Chronic Graft Versus Host Disease who have Received at least 2 Lines of Systemic Therapy
Protocol Number:	SNDX-6352-0504
SAP Version/Date:	SAP V4.0/27 June 2023

We, the undersigned, have reviewed and approved this Statistical Analysis Plan:

Signature	**Date**
[redacted]	
Syndax Pharmaceuticals, Inc.	
[redacted]	
Syndax Pharmaceuticals, Inc.	

Incyte Corporation

VERSION HISTORY

Version	Version Date	Description
1.0	8JUL2022	Original signed version
1.1	17OCT2022	
2.0	11APR2023	Updated per Global Protocol Amendment #4, Version 5.0 (05APR2023)
3.0	13JUN2023	Updated to add the estimands for primary endpoint and key secondary endpoints, and to clarify the threshold that is used in the analysis of mLSS, and to clarify the BOR and DOR derivations.
4.0	27JUN2023	Updated to clarify DOR derivations.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
[redacted]	[redacted]
AE	Adverse Event
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
[redacted]	[redacted]
BOR	Best Overall Response
BUN	Blood Urea Nitrogen
cGVHD	Chronic Graft Versus Host Disease
CPK	Creatine Phosphokinase
CNI	Calcineurin Inhibitor
CR	Complete Response
CrCl	Creatinine Clearance
CRF	Case Report Form
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
[redacted]	[redacted]
CSF-1	Colony stimulating factor 1
CSF-1R	Colony stimulating factor 1 receptor
DLT	Dose Limiting Toxicity
DOR	Duration of Response
ECG	Electrocardiogram
EOT	End of Treatment
FAS	Full Analysis Set
FDA	Food and Drug Administration
FFS	Failure Free Survival
GGT	Gamma Glutamyl Transferase
GLDH	Glutamate Dehydrogenase
hCG	Human Chorionic Gonadotropin
HCV	Hepatitis C Virus
HLA	Human Leukocyte Antigen
HSCT	Hematopoietic Stem Cell Transplantation
IA	Interim Analysis
ICF	Informed Consent Form
IDMC	Independent Data Monitoring Committee
IHC	Immunohistochemistry
IL	Interleukin
IL-34	Interleukin-34

Abbreviation	Definition
INR	International Normalized Ratio
IRT	Interactive Response Technology
ITT	Intent-to-Treat
IV	Intravenous
kg	Kilograms
LFT	Liver Function Test
mAb	Monoclonal Antibody
MedDRA	Medical Dictionary for Regulatory Activities
mLSS	modified Lee Symptom Scale score
NIH	National Institutes of Health
NK	Natural Killer
OR	Objective Response
ORR	Overall Response Rate
OS	Overall Survival
PD	Progressive Disease
PK	Pharmacokinetics
PR	Partial Response
Q2W	Every 2 Weeks
Q4W	Every 4 Weeks
RBC	Red Blood Cell
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SOA	Schedule of Activities
SRC	Safety Review Committee
SRR	Sustained Response Rate
TEAE	Treatment Emergent Adverse Event
TTR	Time to Response
ULN	Upper Limit of Normal
WBC	White Blood Cell Count

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide a description of the statistical methods to be implemented for the analysis of data from the study with Protocol number SNDX-6352-0504 (Amendment 4.0 GLOBAL, Version 5.0, 05 April 2023). The SAP will be finalized prior to any database snapshot or database lock for any interim or final analyses. Any deviations from the SAP after database lock will be documented in the final Clinical Study Report (CSR).

Summary of Major Changes

Section	Change	Rationale
Section 3.4.3.1, Section 3.4.3.2	Definition of DOR is updated.	To clarify the primary DOR definition and definitions used for DOR sensitivity analyses.
Appendix 3	Derivation of DOR is updated	To clarify the primary derivation of DOR.

2 STUDY OVERVIEW

2.1 Study Objectives and Endpoints

The table below outlines the primary, secondary, and exploratory objectives of the study and their associated endpoints:

Objectives	**Endpoints**
Primary	
• To evaluate the overall response rate (ORR) of axatilimab at 0.3 mg/kg IV Q2W, 1 mg/kg IV Q2W, and 3 mg/kg IV Q4W in patients with cGVHD after failure of least 2 prior lines of therapy.	• ORR in the first 6 cycles as defined by the 2014 NIH Consensus Development Project on Criteria for Clinical Trials in cGVHD.
Key Secondary - Efficacy	
• To evaluate key secondary measures of clinical benefit.	• Proportion of patients with a clinically significant improvement in modified Lee Symptom Scale score (mLSS).
Secondary - Efficacy	
• To evaluate secondary measures of clinical benefit.	• ORR on study as defined by the 2014 NIH Consensus Development Project on Criteria for Clinical Trials in cGVHD. • DOR defined as the time from initial response of PR or CR until documented progression of cGVHD, start of new therapy, or death for any reason. • Sustained response rate (SRR) • Organ-specific response rate based on 2014 NIH Consensus Development Project on Criteria for Clinical Trials in cGVHD • Joints and fascia response rate based on refined NIH response algorithm for cGVHD (Inamoto et al 2020) • Percent reduction in average daily dose (or equivalent) of corticosteroids • Proportion of patients who discontinue corticosteroid use after study entry. • Percent reduction in average daily dose (or equivalent) of calcineurin inhibitors • Proportion of patients who discontinue calcineurin inhibitors use after study entry

Objectives	Endpoints
Secondary - Safety	
• To evaluate the safety and tolerability of axatilimab in patients with cGVHD	• Frequency and severity of adverse events (AEs) and serious adverse events (SAEs) • Change from baseline in values for vital signs, safety laboratory parameters, physical and neurological examination, ECG and Karnofsky/Lansky performance scale
[redacted]	
Secondary - PK/Pharmacodynamic	
• To assess the plasma population PK (pop PK) profile of axatilimab in patients with cGVHD	• Axatilimab PK parameters and patient factors that may explain variability in drug exposure
• To assess pharmacodynamic profile of axatilimab	• Change from baseline in colony stimulating factor 1 (CSF-1), interleukin 34 (IL-34) levels and its association with cGVHD response
• To determine or assess the changes in monocyte level with response	• Change from baseline in circulating monocyte number and phenotype (CD14/16)
• To determine or assess the baseline in monocyte level with response	• Baseline circulating monocyte number and phenotype (CD14/16)
Secondary – Immunogenicity	
[redacted]	

Objectives	Endpoints
[redacted]	
Exploratory – Efficacy	
• To explore possible additional evidence of clinical benefit.	[redacted] • Proportion of patients with FFS at Cycle 7 Day 1 and 1 year. FFS is defined as the time from randomization to addition of another systemic immune suppressive therapy for cGVHD, relapse of underlying malignancy, or death whichever is earlier. • Overall survival (OS) • Time to response [redacted]

2.2 Study Design

2.2.1 Overview

AGAVE-201 is a Phase 2, open-label, randomized, multicenter study to evaluate the efficacy, safety and tolerability of axatilimab at 3 different dose levels, in patients with recurrent or refractory active cGVHD who have received at least 2 prior lines of systemic therapy due to progression of disease, intolerability or toxicity. Disease progression is defined 1) by the NIH 2014 consensus criteria, either in terms of organ specific algorithm or global assessment or, 2) as active, symptomatic cGVHD for whom the physician believes that a new line of systemic therapy is required.

The study will consist of 3 periods: Screening, Treatment, and Follow-up. Throughout the study, patients will be evaluated as specified in the Schedule of Activities (SOA) (Study Protocol Section 1.2).

After signing informed consent, potential candidates will undergo screening procedures to determine eligibility. At enrollment, eligible patients will be randomized to one of 3 dose cohorts (axatilimab 0.3 mg/kg IV every 2 weeks [Q2W], 1 mg/kg IV Q2W, or 3 mg/kg IV Q4W). Patients must begin treatment (Cycle 1 Day 1) within 3 days of randomization/enrollment and will receive axatilimab from Cycle 1 Day 1, in 4-week (28-day) treatment cycles, until disease progression (as defined by the NIH 2014 consensus criteria), lack of efficacy by 9 months (Study Protocol Section 6.7.2), withdrawal of consent, or unacceptable toxicity. Following treatment discontinuation, patients will receive an End of Treatment (EOT) visit 30 days after the last dose of study drug.

Simon's optimal 2-stage design will be implemented within each dose cohort. In the first stage 27 patients will be randomized to each of the 3 dose cohorts. To limit the potential exposure of patients to an inefficacious dose and obviate the need for a pause in accrual, the initial futility analysis will be based on an early endpoint (ie, overall response in the first 3 cycles). Each dose will be evaluated for futility and unacceptable toxicity as follows:

- Interim Analysis (IA) #1: Futility assessment based on responses in the first 3 cycles will be conducted. This assessment will occur when the first 27 patients in each cohort have had the opportunity to complete 3 cycles of therapy. If ≤6 patients achieve a response after the first 3 cycles of axatilimab (up to and including Cycle 4, Day 1 assessment), the randomization to this dose level may be stopped for futility. Safety assessment will occur at this IA. For IA # 1, the boundary for unacceptable toxicity is ≥8 out of the first 27 patients having a toxic event defined as any serious or severe (≥Grade 3) TEAE that is attributed to study drug. Grade 2 events that are considered treatment-related and result in medical intervention or hospitalization will be counted as a toxic event.
- IA #2: IA #2 will consist of an overall evaluation of available clinical data. It will be conducted when the first 27 patients in each cohort have had the opportunity to complete 6 cycles of therapy. If ≤9 patients achieve a response to axatilimab, the randomization to this dose level will be stopped for futility. For IA#2, given the full enrollment of the study as of the Protocol Amendment version 5.0, the rate of toxicity events will be provided for all treated patients and will be reviewed as part of the overall benefit/risk assessment. For the overall benefit/risk assessment, a comprehensive safety and efficacy analysis will be provided to the IDMC:
    - Safety for all participants.
    - Efficacy for all participants who have had the opportunity to receive 3 cycles of axatilimab.

2.2.2 Randomization and Blinding

All patients will be centrally assigned to axatilimab dose in a 1:1:1 randomization ratio using an Interactive Response Technology (IRT). Patient assignments will be stratified for severity of cGVHD (mild/moderate vs. severe) by the 4-point scale and prior use of at least one of the following therapies: ibrutinib, ruxolitinib and belumosudil (prior therapy vs. no prior therapy). Before the study is initiated, the log-in information and directions for the IRT will be provided to each site. Although this study is an open-label study, the goal of implementing blinding is to minimize bias in the study conduct, oversight and outcome decisions by restricting the visibility of efficacy, safety and/or treatment data. Details are documented in the Study Blinding Plan.

2.2.3 Study Intervention

This table outlines the study intervention as well as detailed information in regard to the study intervention.

Intervention Name	Axatilimab (SNDX-6352)
Type	Biologic
Dose Formulation	Solution for infusion
Drug Product Strength	[redacted] mg/ml
Unit Dose Strength(s)	mg/kg
Dosage Level(s)	0.3 mg/kg IV Q2W, 1 mg/kg IV Q2W (Day 1 and 15 of each 4-week cycle) or 3 mg/kg IV Q4W (Day 1 of each 4-week cycle)
Route of Administration	IV infusion via an infusion pump over [redacted]; time windows of [redacted] are permitted (ie, infusion time is [redacted]). The exact duration of infusion should be recorded in both source documents and eCRFs.

2.2.4 Sample Size Determination

The sample size was calculated to provide adequate power for testing the statistical hypotheses regarding the primary endpoint of ORR in the first 6 cycles. Based on Simon's optimal 2-stage design, a total of 70 patients in each arm will provide approximately 90% power to detect a true ORR in the first 6 cycles of 50% with a 1-sided significance level of 0.025 in the arm.

The study will enroll patients in each arm in two stages, with 27 patients enrolled in the first stage. An interim analysis (IA) in each arm will be conducted when the first 27 patients have had the opportunity to complete 3 cycles of therapy, and the dose will be evaluated for futility or unacceptable toxicity. In addition, a futility assessment may also be conducted at the time when each patient has had the opportunity to complete 6 cycles of therapy. Details of the IA are described in Section 4.1.

At the final analysis in each arm, if there are ≥29 patients with overall response in the first 6 cycles observed out of 70 patients enrolled in each arm, it will be concluded that the null hypothesis is rejected, and the study has demonstrated that the true ORR in the first 6 cycles exceeds 30% in the arm. The final boundary for efficacy will be re-calculated if there is over enrollment in the second stage of the design. For example, if the total number of randomized patients in a dose level is 80, this dose level will be considered successful if ≥32 patients have had a response to axatilimab. The calculation of the efficacy boundary is presented in Appendix 4.

At the final analysis, if 19 or more patients out of the total 70 patients experience a toxic event (the boundary will be recalculated if there is over enrollment), this will indicate a 90% posterior probability of the toxicity rate exceeding 20% and the final determination of acceptability of a given dose will be based on the overall benefit/risk indicated by review of the totality of evidence.

3 STATISTICAL METHODOLOGY

3.1 General Statistical Considerations

Unless otherwise stated, summary tabulations will be presented that display the number of observations, mean, standard deviation, median, minimum, and maximum for continuous variables, and the number and percentage (of non-missing values) per category for categorical data. Time-to-event data will be analyzed using the Kaplan-Meier method and results will be summarized by 25th, 50th (median) and 75th percentiles with associated 2-sided 95% CIs, as well as the percentage of censored observations.

3.1.1 Study Day

Study day will be calculated in reference to the date of the first dose of axatilimab. Each date will be assigned a study day calculated as follows:

- if the date < first dose date then study day = (Assessment date/Event date – first dose date)
- if the date ≥ first dose date then study day = (Assessment date/Event date – first dose date) +1.

There will be no Day 0. Study Day 1 represents the date of the first dose of study drug.

3.1.2 Analysis Visits

Efficacy analysis of primary efficacy endpoint (ORR in the first 6 cycles) will be based on the time from randomization up to Day 169 or the beginning of Cycle 7, whichever is later.

For all other study variables, the by-visit summaries will be based on the nominal scheduled visit. Unscheduled measurements will not be included in by-visit table summaries but will contribute to the best or worst-case values table summaries. The results of all scheduled and unscheduled measurements will be included in the data listings.

3.1.3 Definition of Baseline

Where applicable, baseline measurements will be defined as the last non-missing observation before the first administration of axatilimab. Change from baseline is defined as any post-baseline measurements minus baseline measurements.

For analyses based on ITT population, baseline measures will be defined as the last non-missing observation before the first administration of axatilimab for patients who have received at least one dose of axatilimab, and will be defined as the last non-missing observation up to Cycle 1 Day 1 visit for patients who were randomized but didn't receive any axatilimab.

3.1.4 Hypothesis Testing

There are three randomized arms in the study with three different dose levels. The primary objective in each arm is to determine the efficacy of axatilimab with respect to ORR in the first 6 cycles. In each arm, the study will test the null hypothesis that the ORR in the first 6 cycles is ≤30% versus the alternative hypothesis that the ORR in the first 6 cycles is >30%. No multiplicity adjustment will be made for this study.

3.1.5 Multicenter Studies

The center effect will not be considered for this study.

3.1.6 Handling of Dropouts and Missing Data

In general, missing data values will be recorded as missing and will not be imputed for the statistical analysis unless specified otherwise. Incomplete dates will be assumed as the most conservative values. Handling incomplete dates (e.g., AE and concomitant medications) is described in Appendix 1. Details on how to handle missing data for mLSS are in Section 3.4.2.

3.2 Analysis Populations

The following populations are defined:

Population	Description
Intent-to-treat Analysis Set (ITT)	ITT analysis set for each dose level consists of all patients randomized to the dose level. Grouping in ITT analysis set will be based on treatment the subject is randomized to.
Safety Analysis Set	Safety Analysis Set consists of all enrolled patients who received at least one dose of study drug during the study. For primary analyses, grouping in safety analysis set will be based on actual initial treatment received on C1D1.
PK Analysis Set	All patients who receive at least one dose of axatilimab and have at least one valid plasma concentration of axatilimab determined.
ADA Analysis Set	All patients who receive at least one dose of axatilimab and have at least one anti-drug antibody analysis captured.
Pharmacodynamic Analysis Set	Patients who are exposed to axatilimab and have sufficient post-baseline samples collected to permit pharmacodynamic analyses.

3.3 Subject Data and Study Conduct

3.3.1 Subject Disposition

Patient disposition will be summarized for all patients in terms of counts and percentages. The following patient disposition categories will be summarized by dose cohort and in total for all screened patients:

- Patients who were screened,
- Patients who were randomized,
- Patients who were screen failed,
- Patients who were treated
- Patients who discontinued from treatment
- Patients who discontinued from study

For patients who were screen failed, who discontinued from treatment, and who discontinued from study, or discontinuation from treatment, or discontinuation from study will be summarized.

All disposition data will be listed by patient.

3.3.2 Protocol Deviations

Protocol deviations will be identified and reported by the process described in the current version of the Study Protocol Deviation Plan.

The number and percentage of patients with major (referred to as 'CSR reportable', where CSR is defined as Clinical Study Report, in the Protocol Deviation Plan) Protocol deviations will be tabulated by category and by dose cohort and in total for the ITT Analysis Set.

The number and percentage of patients with COVID-19 related CSR reportable deviations may also be listed.

All Protocol deviations will be listed by patient.

3.3.3 Analysis Populations

Counts and percentages of subjects in each analysis population will be summarized by dose cohort and in total based on ITT Analysis Set.

3.3.4 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized descriptively by dose cohort and in total for the ITT Analysis Set.

Race, sex, ethnicity, severity of cGVHD (mild/moderate, severe) from IRT and/or CRF, prior use of at least one of the following therapies for cGVHD: ibrutinib, ruxolitinib, and belumosudil (Yes, No) from IRT and/or CRF, prior use of ibrutinib (Yes, No), prior use of ruxolitinib (Yes, No), prior use of belumosudil (Yes, No), Karnofsky or Lansky (for patients under 16 years of age) Performance Status score, and patients with at least 4 organs involved for cGVHD at Cycle 1 Day 1 will be summarized using frequencies and percentages.

Age at informed consent, number of organs involved for cGVHD at Cycle 1 Day 1, platelet count, and total WBC will be summarized by descriptive statistics.

All demographic and baseline characteristics data will be listed by patient.

3.3.5 Disease History

Transplantation related information will be summarized descriptively by dose cohort and in total for the ITT Analysis Set. Number of transplants, type of transplant, type of allogeneic transplant, degree of HLA match, stem cell source and allogeneic transplant type will be summarized using frequencies and percentages. Time from most recent allogeneic transplant to randomization, time from initial cGVHD diagnosis to randomization and time from most recent transplant to cGVHD will be summarized descriptively.

All disease history data will be listed by patient.

3.3.6 Medical History

Medical history will be coded to system organ class and preferred term using the Medical Dictionary for Regulatory Activities (MedDRA) version 24.0. Frequencies and percentages of subjects with medical history by system organ class and preferred term will be summarized by dose cohort and in total based on ITT Analysis Set.

3.3.7 Prior and Concomitant Medical Procedures

Prior and concomitant medical procedures will be coded to system organ class and preferred term using the Medical Dictionary for Regulatory Activities (MedDRA) version 24.0. Prior medical procedures are medical procedures that were performed before the first dose date of study intervention. Concomitant medical procedures are procedures that were performed on or after the first dose date of study intervention.

Both prior and concomitant medical procedures will be summarized descriptively for each dose cohort and in total based on ITT Analysis Set, system organ class and preferred term. All medical procedures data will be listed by patient.

3.3.8 Prior and Concomitant Medications

Prior and concomitant medications will be coded to anatomical therapeutic chemical (ATC) class and preferred term using the WHO Drug Dictionary version B3 Global, September 2020, or higher. Prior medications are medications used and stopped prior to the first dose of study intervention and concomitant medications are medications that were taken at any time on or after the first dose of study intervention (including those started prior to the first dose of study intervention and were ongoing after the first dose of study intervention).

Frequencies and percentages of subjects taking prior and concomitant medications by ATC class and preferred term will be summarized by dose cohort and in total based on the ITT Analysis Set. All medications will be listed by patient.

3.3.9 Prior Therapy for Underlying Malignancy

Prior therapy for underlying malignancy will be coded to anatomical therapeutic chemical (ATC) class and preferred term using the WHO Drug Dictionary version B3 Global, September 2020, or higher.

All prior therapies for underlying malignancy will be listed based on ITT Analysis Set.

3.3.10 Prior GvHD Therapy

Prior GvHD therapy will be coded to anatomical therapeutic chemical (ATC) class and preferred term using the WHO Drug Dictionary version B3 Global, September 2020, or higher.

Prior GvHD therapy will be summarized by ATC class and preferred term for each dose cohort and in total for ITT Analysis Set. Number of regimens for prior cGvHD will also be tabulated by dose cohort and in total. All prior GvHD therapies will be listed by patient.

3.3.11 Study Drug Exposure and Compliance

The following summaries of the study intervention will be presented for Safety Analysis Set:

- Number of cycles,
- Treatment duration, derived as the last dose date of study intervention – first dose date of study intervention + 14/28 day if the last dose occurred in a cycle with Q2W or Q4W dosing, respectively,
- Number of doses,
- Frequency and percentage of patients with at least one dose modification (dose reduction, dose escalation, dosing schedule change),
- Actual dose intensity of axatilimab is defined as the actual cumulative dose (mg) divided by the total number of treatment cycles,
- Relative dose intensity is defined as the actual cumulative dose divided by the planned cumulative dose then multiplied by 100%. Actual cumulative dose is the actual cumulative dose of axatilimab administered until the actual last day of dosing considering dose modification. Planned cumulative dose is the planned cumulative dose of axatilimab until the planned last day of dosing. Planned dose and actual dose recorded on the CRF at each visit will be used for the calculation. Actual dose is only recorded when it's different from the planned dose.

For patients who were with at least one dose modification, the reason for dose reduction, dose escalation, dosing schedule change will be summarized. Swimmer plots will be produced which will include the length of study drug (and will also depict any gaps in dosing), timing of clinical response and disease progression (if applicable), and treatment discontinuation, and/or dose modification.

All study intervention data will be listed by patient.

3.4 Efficacy Assessment

Efficacy data will be summarized by randomized dose cohort based on the ITT Analysis Set.

3.4.1 Primary Efficacy Endpoint

The primary efficacy endpoint is overall response rate (ORR) in the first 6 cycles, defined as the proportion of patients with objective response (OR) at any timepoint during the first 6 cycles, where the first 6 cycles is defined as the time from randomization up to Day 169 or the beginning of Cycle 7, whichever is later; and OR is defined as CR or PR as defined by the NIH Consensus Development Project on Clinical trials in cGVHD (Lee et al 2015). The primary analysis of ORR will be performed after the last patient in the FAS enrolled in the study has had the ability to complete Cycle 7 Day 1 assessment. Table 9 in Study Protocol Section 8.1.1 contains the Working Group proposed consensus definitions of CR, PR and progression for assessment of organ-specific responses as well as a global response determination. Overall response at each visit will be derived using organ specific response. Organ specific responses are evaluated by the physicians based on Table 9 in Section 8.1.1 of the Protocol, using baseline cGVHD severity as a reference for response documentation. Patients with missing response assessments will be considered non-responders. GVHD assessments collected after

new anti-GVHD systemic therapy (including increased dose on axatilimab) would not be considered in ORR assessment.

For ORR, the point estimates and 95% exact binomial confidence intervals will be provided for each dose cohort. If there are ≥29 patients with overall response in the first 6 cycles observed out of 70 patients enrolled in each dose level, it will be concluded that the null hypothesis of that dose level is rejected, and the study has demonstrated that the true ORR in the first 6 cycles exceeds 30% in the dose level. The boundary may be re-evaluated based on Simon-2-Stage design in statistical analysis if different number of patients were randomized. For example, if the total number of randomized patients in a dose level is 80, this dose level will be considered successful if ≥32 patients have had a response to axatilimab.

The BOR is defined as the best response assessment assigned to a patient at any timepoint in the first 6 cycle. The response categories for BOR will be as follows: Complete response, Partial response, No change, Progression, and Other.

The estimand for the primary endpoint is the treatment effect of axatilimab at 0.3 mg/kg IV Q2W, 1 mg/kg IV Q2W, and 3 mg/kg IV Q4W, assessed by overall response rate (ORR) in all randomized participants with cGVHD after failure of at least 2 prior lines of therapy.

- Analysis population: All randomized participants at each dose level
- Variable: CR or PR at any timepoint in the first 6 cycles defined by the 2014 NIH Consensus Development Project on Criteria for Clinical Trials in cGVHD
- Intercurrent events: death, lost to follow-up, progressive disease (PD), and start of subsequent new systemic treatment for cGVHD, including an increase in the dose of corticosteroid (for a treatment of cGVHD, an increase in the dose of corticosteroids to prednisolone 1 mg/kg (+/- 10%) equivalent is considered a new systemic therapy). Only the CR or PR up to the time of these intercurrent events, whichever is earlier, are considered as responses. Participants below will be counted as non-responders:
    - Participants who do not have a post-baseline response assessment due to lost to follow-up, early PD
    - Participants who move on to subsequent systemic therapy prior to achieving CR or PR
    - Participants who die, experience PD, or stop response assessments for any reason prior to achieving CR or PR
- Population level summary: Proportion of participants with a best response of CR or PR at any timepoint in the first 6 cycles in the analysis population and its 95% exact binomial confidence intervals at each dose level

Subgroup analyses will be performed for the primary endpoint. The subgroups are:

- Age (<17 years, >=17 and <65 years, >=65 years)
- Sex
- Race
- Number of Lines of prior therapy
- Prior ibrutinib (Yes, No)
- Prior ruxolitinib (Yes, No)
- Prior belumosudil (Yes, No)
- Best response to the last prior cGVHD treatment (CR/PR, No change, PD, Unknown)
- Time from cGVHD diagnosis to enrollment
- Severity of cGVHD at screening (mild/moderate, severe)
- Baseline corticosteroid dose level
- Number of organs involved at baseline (<=4 vs >4)
- Lung involvement at baseline (Yes, No)

3.4.2 Key Secondary Efficacy Endpoints

The key secondary endpoint is defined as the proportion of patients with a ≥7-point improvement from baseline in normalized score using the cGVHD mLSS (Lee et al 2002, Teh et al 2020). The cGVHD mLSS contains 28 items grouped in 7 subscales (skin, eyes and mouth, breathing, eating and digestion, muscles and joints, energy, emotional distress). The cGVHD mLSS questionnaire asks patients to indicate the degree of "bother" that they experienced during the past 7 days due to symptoms in 7 domains potentially affected by chronic GVHD (Lee et al 2002, Teh et al 2020).

Subscales may be scored if 50% of more of the items in the subscale are completed. Scores are linearly transformed to a 0-100 scale where 0 means all answered items were a "0" and "100" means that all answered items were a "4". Missing items are not included in the scoring. The normalized summary score is calculated as the average of the subscale scores, as long as 4 or more subscales are available and normalizing to a 0 to 100 scale (Merkel et al 2016).

A summary table on compliance measures at each scheduled visit will be provided by dose cohort, where the percentage compliance at a visit will be calculated as the number of completed cGVHD mLSS questionnaires at the visit divided by the number of expected mLSS questionnaires at the visit and then multiplied by 100. The number of completed questionnaires is calculated as number of subjects who have normalized score available at the specified visit. The number of expected questionnaires is determined by the number of subjects who are ongoing with study treatment at the specified visit.

The key secondary endpoint will be summarized for each scheduled assessment, and best change over time during anytime on study. For the key secondary endpoint, the point estimates and 95% exact binomial confidence intervals will be provided. Waterfall plots of the best change from baseline in normalized mLSS score will be presented.

The estimand for the key secondary endpoint is the clinical benefit of axatilimab at 0.3 mg/kg IV Q2W, 1 mg/kg IV Q2W, and 3 mg/kg IV Q4W, measured by the proportion of participants with a ≥7 points improvement in modified Lee Symptom Scale score (mLSS) in the ITT population.

- Analysis population: All randomized participants at each dose level
- Variable: maximum reduction from baseline in the mLSS
- Intercurrent events: Subsequent new systemic treatment for cGVHD. mLSS assessment that occurred after the subsequent new systemic cGVHD treatment will be excluded
- Population level summary: Proportion of participants with a ≥7 points improvement in mLSS at any timepoint in the first 6 cycles in the analysis population and its 95% exact binomial confidence intervals at each dose level.

3.4.3 Other Secondary Efficacy Endpoints

3.4.3.1 Definition

The secondary endpoints include:

- ORR on study defined as proportion of patients with OR anytime on study. OR is defined as Section 3.4.1. Overall response will be derived as Section 3.4.1;
- BOR on study defined as best response assessment assigned to a patient at any timepoint during the study;
- DOR (days) defined as the interval from the date of first response (PR or better) to progression of cGVHD from disease nadir in any organ, new systemic therapy, or death from any cause, whichever comes first, in responders only. For a treatment of cGVHD, an increase in the dose of corticosteroids to prednisolone 1 mg/kg (+/- 10%) equivalent is considered a new systemic therapy. Refer to Appendix 3 for detailed derivation.
- Determination of events and censoring times for DOR will be performed according to the table in Section 3.4.3.2, depending on the situation.
- SRR defined as proportion of patients with sustained response. The sustained response is defined as OR lasting for at least 20 weeks (140 days) from the time of initial response. Responses by organ system will be assessed based on the 2014 NIH Consensus Development Project on Clinical Trials in cGVHD;
- Organ-specific response rate defined as the proportion of patients with OR for the nine individual organs based on 2014 NIH Consensus Development Project on Criteria for Clinical Trials in cGVHD (Skin, Eyes, Mouth, Esophagus, Upper GI, Lower GI, Liver, Lungs, Joints and Fascia) in patients with each specific organ involvement at baseline;
- Joints and fascia response rate defined as the proportion of patients with Joints and Fascia response based on refined NIH response algorithm for cGVHD in patients with joints and fascia involvement at baseline;
- Percent reductions in average daily doses (or equivalent) of corticosteroid from baseline to discontinuation of corticosteroid or end of treatment, whichever comes first, corticosteroid will be converted to an equivalent dose of prednisone in mg (refer to Appendix 2 for conversions);
- Proportion of patients who discontinue corticosteroid use after first dose of study intervention until end of treatment;

- Percent reductions in average daily doses (or equivalent) of calcineurin inhibitors from baseline to discontinuation of calcineurin inhibitors or study drug, whichever comes first;
- Proportion of patients who discontinue calcineurin inhibitors after first dose of study drug until end of treatment. Patients who discontinue calcineurin inhibitors due to intolerance and switch to sirolimus will not be considered as discontinuing calcineurin inhibitor.

3.4.3.2 Analyses of Other Secondary Efficacy Endpoints

ORR, SRR, Organ-specific response rate, Joints and fascia response rate based on refined NIH response algorithm for cGVHD

For the response rates (i.e., ORR, SRR, Organ-specific response rate, Joints and fascia response rate), the point estimates and 95% exact binomial confidence intervals may be provided.

Proportion of patients who discontinue corticosteroid, proportion of patients who discontinue calcineurin inhibitors

For the proportions, the point estimates and 95% exact binomial confidence intervals may be provided. The proportion of patients who discontinued corticosteroid and calcineurin inhibitors may also be summarized by cycle for patients who are ongoing with axatilimab treatment.

Percent reductions in average daily doses (or equivalent) of corticosteroid, percent reductions in average daily doses (or equivalent) of calcineurin inhibitors

The average daily dose within each cycle and percent reductions of average daily dose within each cycle may be summarized descriptively by cycle. Means and standard errors may be presented descriptively.

DOR

Time to event variable DOR will be summarized descriptively using the Kaplan-Meier method for each dose cohort for each definition separately in patients who are responders. Point estimates and 2-sided 95% CIs for median using the Brookmeyer-Crowley (1982) method will be estimated. Plots of the Kaplan-Meier estimate for the duration of response will be presented by dose cohort.

Situation	Date of event or censoring	Outcome
Death without new systemic treatment or disease progression	Date of death	Event
Subsequent new systemic treatment before disease progression or death	Date of initiation of subsequent systemic treatment	Event
Disease progression before start of new systemic treatment	Date of the disease assessment corresponding to the disease progression	Event
No death and without disease progression or subsequent new systemic treatment	Date of the last disease assessment	Censored
Discontinued treatment without a disease progression	Date of last disease assessment	Censored

The estimand for the primary analysis of duration of response (DOR) is the duration of treatment effect of axatilimab at 0.3 mg/kg IV Q2W, 1 mg/kg IV Q2W, and 3 mg/kg IV Q4W assessed by the median duration of response in the analysis population.

- Analysis population: All responders, defined as the randomized participants who have achieved CR or PR at any timepoint in the first 6 cycles by the 2014 NIH Consensus Development Project on Criteria for Clinical Trials in cGVHD, at each dose level.
- Variable: DOR (days) defined as the interval from the date of first response (PR or better) to progression of cGVHD from disease nadir in any organ, new systemic therapy, or death from any cause, whichever comes first, in responders only. More details are specified in Appendix 3.
- Intercurrent events: Deaths, PD, and the start of subsequent new systemic therapy for cGVHD prior to PD or death. The intercurrent events of death, PD and the subsequent therapy will be considered as “event” in the analysis. Responders who haven’t experienced any event will be censored at the date of the last disease assessment before data cut-off date.
- Population level summary: median DOR using the Kaplan-Meier method for each dose level including point estimate and 2-sided 95% CIs.

Sensitivity analyses will be performed and the DOR is defined as the interval from the date of initial overall response of CR or PR until:

1. Start of new anti-GVHD systemic therapy, or death from any cause, whichever is earlier (alternative measure of durability of response). Patients who haven’t started new therapy and are still alive will be censored at the last contact date. Similar methods as the primary analysis of DOR will be applied.
2. The first progression of cGVHD from nadir organ-level response, or start of new anti-GVHD systemic therapy, or death from any cause, whichever is earlier.
3. The first progression of cGVHD from organ status at baseline, or start of new anti-GVHD systemic therapy, or death from any cause, whichever is earlier.

3.4.4 Exploratory Efficacy Endpoints

3.4.4.1 Definition

The exploratory efficacy endpoints are as follows:

- Changes from baseline in patient-reported symptom activity using the Lee cGVHD symptom scale;
- Failure Free Survival (FFS) defined as the time from randomization to addition of new systemic anti-cGVHD therapy, relapse of underlying malignancy, or death, whichever is earlier. Determination of events and censoring times for FFS will be performed according to the table blow, depending on the situation.

Situation	Date of event or censoring	Outcome
Addition of new systemic anti-cGVHD therapy before or without relapse of underlying malignancy or death	Date of the initiation of systemic therapy	Event
Relapse of underlying malignancy before or without addition of new systemic anti-cGVHD therapy or death	Date of relapse	Event
Death without addition of new systemic anti-cGVHD therapy y or relapse of underlying malignancy	Date of death	Event
No death and without any event defined above	Date of the last date known to be alive	Censored

- Overall survival defined as time from randomization to the date of death from any cause. Patients who are alive or lost to follow-up (as of the data analysis cutoff date) will be censored at the date patient was last known to be alive. The last date the patient was known alive will be determined based on the latest of the following: study assessment date (e.g. physical examination, vital signs, KPS score, neurological examination, ECG, laboratory tests, response assessment, physician or patient self-assessment, etc.), study drug dosing date, AE/medication start and stop dates, last contact of survival follow-up;
- Time to response defined as time from randomization to the first date the patient achieved a PR or CR, only calculated for patients who achieve a CR or PR;
- Time to next treatment defined as time from randomization to the first date on which a patient receives new systemic anti-cGVHD therapy;
- Changes from baseline in cGVHD severity as based on the physician-reported global cGVHD activity assessment;
- Duration of >5 point improvement in mLSS defined as the summation of the time in the post-baseline cycles for which the decrease in the total score from baseline is >5 points, the summation is calculated as the sum of anytime in post-baseline cycles with a >5 points decreasing in the total score compared to baseline;
- Duration of ≥7 point improvement in mLSS defined as the summation of the time in the post-baseline cycles for which the decrease in the total score from baseline is ≥7 points, the summation is calculated as the sum of anytime in post-baseline cycles with a ≥7 points decreasing in the total score compared to baseline;
- Changes from baseline in the 10-point NIH Patient Global Rating Score.

3.4.4.2 Analyses of Exploratory Efficacy Endpoints

Changes from baseline in patient-reported symptom activity using the Lee cGVHD symptom scale

The changes from baseline in patient-reported symptom activity using Lee cGVHD symptom scale will be summarized descriptively for each scheduled assessment, and best change over time during anytime on study.

FFS

The FFS rate at 6 months and 1 year will be summarized descriptively using Kaplan-Meier method, and the Kaplan-Meier estimates will be provided along with the 95% CI.

Overall survival

OS will be analyzed using the Kaplan-Meier method for each dose cohort. The median (and its 95% CI), number and percentage of patients who have died and patients being censored will be provided. The Kaplan-Meier estimates of overall survival at 6-month intervals will also be presented. Plots of the Kaplan-Meier estimate for the overall survival will be presented by dose cohort.

Time to response

The TTR will be summarized descriptively for patients who achieve a CR or PR.

Time to next treatment

The duration of time to next treatment will be summarized descriptively for patient who receive new systemic anti-cGVHD therapy (including increase of steroid dose by 10% from baseline).

Changes from baseline in physician-reported global cGVHD activity assessment

The changes from baseline in physician-reported global cGVHD activity assessment will be summarized descriptively for each scheduled assessment.

Duration of >5 point improvement from baseline in mLSS

Number of cycles with of >5 point improvement from baseline in mLSS will be summarized descriptively. Spider plot may be presented for mLSS change from baseline (x axis will show duration in days).

Duration of ≥7 point improvement from baseline in mLSS

Number of cycles with of ≥7 point improvement from baseline in mLSS will be summarized descriptively. Spider plot may be presented for mLSS change from baseline (x axis will show duration in days).

Changes from baseline in the 10-point NIH Patient Global Rating Score

The changes from baseline in the 10-point NIH patient global rating score will be summarized descriptively.

3.5 Pharmacokinetic Assessment

The details of the analysis methodology and results for PK assessment will be described in a separate population PK report.

3.6 Pharmacodynamic Assessment

The PD effect will be assessed for the following biomarkers. For each of the following PD parameters, the change will be listed and summarized using descriptive statistics by dose cohort and total.

- Changes in skin macrophages, Langerhans cells, and dendric cells in skin or pulmonary biopsy prior to SNDX-6352 (baseline) and after 2 cycles of axatilimab treatment.
- Changes in circulating immune cells in peripheral circulation, including NK cells, T-cells, B-cells prior to and after axatilimab treatment.
- Changes in CSF-1, IL-34 and/or inflammatory cytokine concentrations prior to axatilimab (baseline) and after treatment and its association with cGVHD response.
- Changes in circulating monocytes and phenotype (CD14/16) prior to axatilimab and after axatilimab treatment.

3.7 Safety Assessment

Safety assessments include adverse events, AEs of CTCAE Grade >= 3, AESI, as well as physical and neurological examinations, vital signs, clinical laboratory assessments, ECGs, Karnofsky or Lansky (for patients under 16 years of age) Performance Status score, and COVID-19 test.

All safety analyses will be summarized by actual initial treatment received on C1D1 and in total based on the Safety Analysis Set.

3.7.1 Adverse Events (AEs)

An AE is any untoward medical occurrence in a patient or clinical investigation patient administered a pharmaceutical product and that does not necessarily have a causal relationship with this treatment (ICH E6:1.2). An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease (new or exacerbated) temporally associated with the use of study intervention. All AEs will be coded to system organ class and preferred term using MedDRA v 24.0. The severity of all AEs will be graded according to the NCI CTCAE v5.0. If a patient experiences repeat episodes of the same AE, then the event with the highest severity grade and strongest causal relationship to study treatment will be used for purposes of incidence tabulations.

Treatment emergent adverse events (TEAEs) are defined as AEs either reported for the first time or worsened from a pre-existing event after the first dose of study drug, and until 90 days post last dose of study drug.

Adverse events of special interest are defined as: infusion-related reactions including hypersensitivity reactions, and infections.

An overview of AEs will be provided including counts and percentages of subjects with the following:

- Any TEAEs (overall and by worst CTCAE grade)
- Any study drug related TEAEs (overall and by worst CTCAE grade)
- Any TEAEs of special interest (overall and by worst CTCAE grade)
- TEAEs with CTCAE Grade >=3
- TEAEs with study drug related CTCAE Grade >=3
- TEAEs with action of dose interrupted
- TEAEs with action of dose reduced
- TEAEs with action of study treatment discontinued
- Any treatment-emergent serious AEs (TESAEs)
- Any AEs leading to death

Counts and percentages of subjects will also be tabulated by system organ class and preferred term for each of the categories in the overview. For those summaries, subjects with multiple adverse events will be counted only once per system organ class and preferred term.

Counts and percentages of subjects will be tabulated by worst CTCAE grade, system organ class, and preferred term for TEAEs, study drug related TEAEs, and TEAEs of special interests. For these summaries, subjects with multiple adverse events will be counted only once by the worst CTCAE grade within a system organ class and preferred term.

Supportive analyses that will assess the AEs prior to and after the IDMC recommendation of the IA#1 may be conducted to better characterize the safety profile of the 1 mg/kg or 3 mg/kg doses.

Listings will be presented for all TEAEs, TEAEs with CTCAE Grade >=3, SAEs and TEAEs leading to dose interrupted, or dose reduced, TEAE leading to study treatment discontinuation.

3.7.2 Clinical Laboratory Tests

Clinical laboratory tests will be collected according to Schedule of Activities (SOA) (Study Protocol Section 1.2), and may be repeated during the treatment period per the Investigator's clinical judgment. The laboratory tests will be performed by the local laboratory, as listed in the table below.

Descriptive statistics for absolute value and change from baseline will be provided for clinical laboratory tests (hematology, clinical chemistry, and coagulation parameters) by visit. Changes from baseline for the maximum post-treatment value, minimum post-treatment value for selected clinical laboratory parameters (hematology, clinical chemistry,) will also be summarized descriptively. Similar analysis may be performed for [redacted] and T-Score and Z-Score from bone density scan.

Abnormal laboratory results will be graded according to the NCI CTCAE v5.0, if applicable. Shift tables for baseline and worst post-treatment values according to NCI CTCAE grade, will be provided for selected hematology and clinical chemistry laboratory parameters.

The count and percentage of subjects with potential Hy's Law cases: ALT or AST ≥3×ULN, total bilirubin ≥2×ULN, and ALP < 2×ULN (at the same visit) will be tabulated. The evaluation of Drug-Induced Serious Hepatotoxicity plot may be presented for all subjects.

Hematology	
White blood cell count (WBC) with differential (Neutrophils, Lymphocytes, Monocytes, Eosinophils, Basophils)	Coagulation factors, including PT or International Normalized Ratio (INR) and aPTT (Screening only)
Hemoglobin	Hematocrit
Platelet count	Red blood cell (RBC)
Clinical Chemistries	
Alanine Aminotransferase (ALT)	Aspartate Aminotransferase (AST)
Alkaline phosphatase	Gamma-glutamyl transferase (GGT)
Total bilirubin (fractionated)	Albumin
Calcium	Blood urea nitrogen (BUN)
Sodium	Creatinine
Chloride	Potassium
Glucose	Bicarbonate
Phosphorus/phosphates	Lactate dehydrogenase
Uric acid	Total protein
Amylase	Magnesium (Screening only, unless clinically indicated)
Lipase	CPK
Urinalysis	
Basic Urinalysis (dipstick, including macroscopic appearance, bilirubin, blood, color, glucose, ketones, leukocyte esterase, nitrite, pH, protein, specific gravity, urobilinogen; prior to each administration of study intervention treatment after C1 if clinically indicated).	Full urinalysis (dipstick plus microscopic evaluation) to be performed only at the Screening and EOT visits).
Other	
[redacted]	
Coagulation studies, including PT or INR and aPTT	Human chorionic gonadotropin (hCG) pregnancy test (as needed for women of childbearing potential
HCV RNA	

3.7.3 Vital Signs

Vital signs will be measured at scheduled visits according to the Schedule of Activities (SOA) (Study Protocol Section 1.2). Measurements will include height, weight, systolic and diastolic blood pressure (mmHg), pulse rate (beats per minute), respiration rate (breaths per minute), and temperature (F or C; oral, axillary, temporal, or auricular).

Descriptive statistics will be provided for the vital sign measurements (weight, systolic and diastolic blood pressure, pulse rate, respiration rate and temperature). Changes from baseline to the maximum, minimum post-treatment values will also be summarized descriptively. Both scheduled and unscheduled post-treatment visits values will be considered for the maximum or minimum post-treatment values.

Criteria for clinically notable vital sign abnormalities are defined as follows. The abnormal values for participants exhibiting clinically notable vital sign abnormalities will be listed. Alert vital signs are defined as an absolute value outside the defined range and percentage change from baseline greater than 25%. The abnormal values for participants exhibiting alert vital sign abnormalities will be listed.

Parameter	High Threshold	Low Threshold
Systolic blood pressure	> 155 mmHg	< 85 mmHg
Diastolic blood pressure	> 100 mmHg	< 40 mmHg
Pulse	> 100 bpm	< 45 bpm
Respiratory rate	> 24 breaths/min	< 8 breaths/min
Temperature	> 38°C	< 35.5°C

All vital sign data will be listed by subject.

3.7.4 12-Lead Electrocardiograms

Triplicate 12-lead electrocardiograms (ECG) will be obtained at time points that given in the Schedule of Activities (SOA) (Study Protocol Section 1.2), including heart rate and measures PR, QRS, QT, and QTc intervals.

All ECG data will be listed by subject for each individual reading and the average of the triplicates.

Criteria for clinically notable ECG abnormalities are defined as follows. Participants exhibiting clinically notable ECG abnormalities will be listed with study visit. Abnormal values for participants with alert ECG values, defined as both the absolute value and the percentage change from baseline being outside normal ranges, will be identified and listed.

Parameter	High Threshold	Low Threshold
QTc(QTcF/QTcB)	> 480 ms	< 295 ms
PR	> 220 ms	< 75 ms
QRS	> 120 ms	< 50 ms
QT	> 500 ms	< 300 ms
RR	> 1330 ms	< 600 ms
Heart rate	> 100 bpm	< 50 bpm

3.7.5 Physical Examinations

Physical examinations will be performed at visits according to Schedule of Activities (SOA) (Study Protocol Section 1.2). All physical examinations will be listed by subject.

3.7.6 Neurological Examinations

A neurological examination will include general appearance, posture and gate, motor activity, assessment of mental status, cranial nerves, sensory and neuromuscular function, and reflexes. Neurological examinations will be performed at visits according to Schedule of Activities (SOA) (Study Protocol Section 1.2).

All neurological examinations will be listed by subject.

3.7.7 Karnofsky/Lansky Performance Status Score

Karnofsky/Lansky performance status score will be collected at time points as outlined in Schedule of Activities (SOA) (Study Protocol Section 1.2). KPS data will be summarized descriptively for actual values and for changes from baseline by scheduled visits, including maximum and minimum post-treatment values. Both scheduled and unscheduled post-treatment visits values will be considered for the maximum or minimum post-treatment values.

All KPS data will be listed by subject.

3.7.8 ADA Analysis

The details of the analysis methodology and results for ADA assessment will be summarized in the integrated summary of immunogenicity.

4 INTERIM ANALYSIS AND FINAL ANALYSIS

4.1 Interim Analysis

Two IAs addressing efficacy and safety are planned for this study. An Independent Data Monitoring Committee will evaluate all data that are available at the time of the data cut for the interim analyses, and provide a recommendation on continued use of the available axatilimab dose levels.

Interim analyses will occur when the first 27 patients in each cohort have had the opportunity to complete 3 (IA #1) and 6 (IA #2) cycles of therapy. The addition of the first IA to the Simon's 2-stage design does not inflate Type I error. With the assumption that 70% of responses occur in the first 3 cycles, adding this futility assessment slightly reduces the overall study power from 90% to 83% but provides a 67% probability of early stopping under the null to reduce potential over-randomization to a futile dose.

Toxicity evaluation will coincide with the futility assessments. Toxic event is defined as any serious or severe (≥Grade 3) TEAE that is attributed to study drug. Grade 2 events that are considered treatment-related and result in medical intervention or hospitalization will be counted as a toxic event.

For IA #1, the stopping rules for unacceptable toxicity are established using Bayesian method using a maximum acceptable toxicity rate of 20%, a θ prior distribution (1,1), and a posterior probability of 90% (Lee et al 2021). The operating characteristics of this design are calculated using Bayesian Toxicity Monitoring Software (BTM) by The University of Texas MD Anderson Cancer Center.

IA #1 Toxicity Stopping Boundaries (Rate = 0.20 with Prior [1,1] and Posterior Probability of 0.90)

Total Number of Patients	Number of Patients with Toxicities
27	≥8

Simon's optimal 2-stage design will be implemented within each dose cohort. In the first stage 27 patients will be randomized to each of the 3 dose cohorts. To limit the potential exposure of patients to an inefficacious dose and obviate the need for a pause in accrual, the initial futility analysis will be based on an early endpoint (ie, overall response in the first 3 cycles). Each dose will be evaluated for futility and unacceptable toxicity as follows:

- IA #1: Futility assessment based on responses in the first 3 cycles will be conducted. This assessment will occur when the first 27 patients in each cohort have had the opportunity to complete 3 cycles of therapy. If ≤6 patients achieve a response after the first 3 cycles of axatilimab (up to and including Cycle 4, Day 1 assessment), the randomization to this dose level may be stopped for futility. Safety assessment will occur at this IA. For IA #1, the boundary for unacceptable toxicity is ≥8 out of the first 27 patients having a toxic event defined as any serious or severe (≥Grade 3) TEAE that is attributed to study drug. Grade 2 events that are considered treatment-related and result in medical intervention or hospitalization will be counted as a toxic event.
- IA #2: IA #2 will consist of an overall evaluation of available clinical data. It will be conducted when the first 27 patients in each cohort have had the opportunity to complete 6 cycles of therapy. If ≤9 patients achieve a response to axatilimab, the randomization to this dose level will be stopped for futility. For IA #2, given the full enrollment of the study as of this amendment, the rate of toxicity events will be provided for all treated patients and will be reviewed as part of the overall benefit/risk assessment. For the overall benefit/risk assessment, a comprehensive safety and efficacy analysis will be provided to the IDMC:
    - Safety for all participants
    - Efficacy for all participants who have had the opportunity to receive 3 cycles of axatilimab.

4.2 Final Analysis

A final efficacy analysis will be performed when all patients have had the opportunity to complete 6 cycles of treatment with axatilimab. A dose level will be considered successful if ≥29 patients have had a response to axatilimab (PR or CR), as defined by NIH 2014 cGVHD criteria. The final boundary for efficacy will be re-calculated if there is over enrollment in the second stage of the design. For example, if the total number of randomized patients in a dose level is 80, this dose level will be considered successful if ≥32 patients have had a response to axatilimab. The calculation of the efficacy boundary is presented in Appendix 4.

At the final analysis, if 19 or more patients out of the total 70 patients experience a toxic event (the boundary will be recalculated if there is over enrollment), this will indicate a 90% posterior probability of the toxicity rate exceeding 20% and the final determination of acceptability of a given dose will be based on the overall benefit/risk indicated by review of the totality of evidence.

5 CHANGES FROM PROTOCOL-SPECIFIED STATISTICAL ANALYSES

No changes from Protocol-specified statistical analysis.

6 PROGRAMMING SPECIFICATIONS

Analyses will be performed using SAS® version 9.4 or higher. Detailed Programming Specifications will be provided in a separate document.

7 REFERENCES

Brookmeyer, R., Crowley, J. (1982). A confidence interval for the median survival time. Biometrics. 1982 Mar 1:29-41.

Inamoto, Y., S. J. Lee, L. E. Onstad, M. E. D. Flowers, B. K. Hamilton, M. H. Jagasia, et al. (2020). Refined National Institutes of Health response algorithm for chronic graft-versus-host disease in joints and fascia. Blood Adv 4(1): 40-46.

Lee, J., Y. Kuo, D. Liu, N. Chen. Bayesian Toxicity Monitoring. Department of Biostatistics, MD Anderson Cancer Center Houston, TX; (biostatistics.mdanderson.org/shinyapps/BTOX). PID 880, v2.2.3.0; Last updated 4/22/2021; accessed 9 June 2023.

Lee, S., E. F. Cook, R. Soiffer, J. H. Antin (2002). Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. Biol Blood Marrow Transplant 8(8): 444-452.

Lee, S. J., D. Wolff, C. Kitko, J. Koreth, Y. Inamoto, M. Jagasia, et al. (2015). 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 21(6): 984-999.

Merkel, E. C., S. A. Mitchell, S. J. Lee (2016). Content Validity of the Lee Chronic Graft-versus-Host Disease Symptom Scale as Assessed by Cognitive Interviews. Biol Blood Marrow Transplant 22(4): 752-758.

Teh, C., L. Onstad, S. J. Lee (2020). Reliability and Validity of the Modified 7-Day Lee Chronic Graft-versus-Host Disease Symptom Scale. *Biol Blood Marrow Transplant* 26(3): 562-567.

APPENDIX 1: INCOMPLETE DATES CONVENTIONS

Algorithm for Treatment-emergent Adverse Events

AE Start Date	AE Stop Date	Action
Known	Known/Partial/Missing	• If start date < study drug start date, then not TEAE • If start date ≥ study drug start date and ≤ (end of treatment +90 days), then TEAE • If start date > (end of treatment +90 days), then not TEAE
Partial, but the known date components show that it cannot be on or after study drug start date	Known/Partial/Missing	Not TEAE
Partial, could be on or after study drug start date	Known	• If stop date < study drug start date, then not TEAE • If stop date ≥ study drug start date, then TEAE
	Partial	Impute stop date as latest possible date (ie, last day of month if day is unknown or 31-Dec if day and month are unknown), then: • If stop date < study drug start date, then not TEAE • If stop date ≥ study drug start date, then TEAE
	Missing	Assumed TEAE
Missing	Known	• If stop date < study drug start date, then not TEAE • If stop date ≥ study drug start date, then TEAE
	Partial	Impute stop date as latest possible date (ie, last day of month if day is unknown or 31st December if day and month are unknown), then: • If stop date < study drug start date, then not TEAE • If stop date ≥ study drug start date, then TEAE
	Missing	Assumed TEAE

TEAE = treatment-emergent AE

Algorithm for Concomitant Medications

CM Start Date	CM Stop Date	Action
Known	Known	• If stop date < study drug start date, assign as PRIOR • If stop date ≥ study drug start date, and start date ≤ end of treatment +90 days, assign as CONCOMITANT • If stop date ≥ study drug start date and start date > 90 days after the end of treatment, assign as POSTTREATMENT
	Partial	Impute stop date as latest possible date (ie, last day of month if day unknown or 31-Dec if day and month are unknown), then: • If stop date < study drug start date, assign as PRIOR • If stop date ≥ study drug start date, and start date ≤ end of treatment +90 days, assign as CONCOMITANT • If stop date ≥ study drug start date, and start date > 90 days after the end of treatment, assign as POSTTREATMENT
	Missing	• If stop date is missing, then PRIOR will never be assumed or assigned • If start date ≤ end of treatment, assign as CONCOMITANT • If start date > 90 days after the end of treatment, assign as POSTTREATMENT
Partial	Known	Impute start date as earliest possible date (ie, first day of month if day unknown or 01-Jan if day and month unknown), then: • If stop date < study drug start date, assign as PRIOR • If stop date ≥ study drug start date and start date ≤ end of treatment + 90 days, assign as CONCOMITANT • If stop date ≥ study drug start date and start date > 90 days after the end of treatment, assign as POSTTREATMENT
	Partial	Impute start date as earliest possible date (ie, first day of month if day unknown or 01-Jan if day and month are unknown) and impute stop date as latest possible date (ie, last day of month if day unknown or 31-Dec if day and month are unknown), then: • If stop date < study drug start date, assign as PRIOR • If stop date ≥ study drug start date and start date ≤ end of treatment + 90 days, assign as CONCOMITANT • If stop date ≥ study drug start date and start date > 90 days after the end of treatment, assign as POSTTREATMENT
	Missing	Impute start date as earliest possible date (ie, first day of month if day unknown or 01 Jan if day and month unknown), then: • If stop date is missing, then PRIOR will never be assumed or assigned • If start date ≤ end of treatment + 90 days, assign as CONCOMITANT • If start date > 90 days after the end of treatment, assign as POSTTREATMENT

CM Start Date	CM Stop Date	Action
Missing or Unknown	Known	• If stop date < study drug start date, assign as PRIOR • If stop date ≥ study drug start date, assign as CONCOMITANT • If start date is missing, then POST-TREATMENT will never be assumed or assigned
	Partial	Impute stop date as latest possible date (ie, last day of month if day unknown or 31-Dec if day and month are unknown), then: • If stop date < study drug start date, assign as PRIOR • If stop date ≥ study drug start date, assign as CONCOMITANT • If start date is missing, then POSTTREATMENT will never be assumed or assigned
	Missing	Assign as CONCOMITANT

APPENDIX 2: SYSTEMIC PREDNISONE EQUIVALENT DOSE CONVERSION

Systemic prednisone equivalent dose conversions in mg are presented in the table below in order to evaluate systemic steroid use for GvHD.

Medication	Route	Equivalent Dose (mg)	Prednisone Equivalent Conversion Factor
Betamethasone	IV	0.75	6.7
Cortisone	PO	25	0.2
Dexamethasone	IV or PO	0.75	6.7
Hydrocortisone	IV or PO	20	0.25
Methylprednisolone	IV or PO	4	1.25
Prednisolone	PO	5	1
Prednisone	PO	5	1
Triamcinolone	IV	4	1.25

Source: https://www.mdcalc.com/steroid-conversion-calculator

NOTES:

1. Prednisone equivalent conversion factor is calculated by dividing the equivalent dose for prednisone by the equivalent dose for each medication.

2. Systemic prednisone equivalent dose is calculated by multiplying the dose by the prednisone equivalent conversion factor.

APPENDIX 3: DERIVATION OF BEST OVERALL RESPONSE AND DURATION OF RESPONSE

Derivation of BOR:

Step 1: Overall response at each post-baseline visit

If “Organ-specific response is PD in any organ” then “overall response at that visit = PD” regardless of whether there is missing organ assessment, else if an organ assessment is missing, the response would be considered “Not Evaluable” at that visit, else if “Organ-specific response is CR in all involved organs” then “overall response at that visit = CR”, else if “Organ-specific response is PR in at least one involved organ AND Organ-specific response is NOT PD in any organ” then “overall response at that visit = PR”.

Step 2: Best overall response across all post-baseline visits

Select the best response across all post-baseline visits before the start of new anti-GVHD systemic therapy.

Derivation of DOR:

Step 1: For each participant, identify the individual organ raw data (cGVHD scores, laboratory, or FEV1 values) at the visit at which the initial overall response is documented, denote as “organ status at DOR start”.

Step 2: For each organ assessed at visits after the visit of overall response, compare the raw data with the best raw data observed at or after the initial overall response visit to determine organ level response status. If worsening in raw data from the best observed raw data is documented and meets the NIH progression criteria (Protocol, Table 9), then set the organ response at that visit to PD. Otherwise, keep the investigator-assessed organ response without change. Derive a new overall response to reflect updated organ response.

Step 3: Compute DOR as specified in Section 3.4.3.

APPENDIX 4: CALCULATION OF THE EFFICACY BOUNDARY

The primary endpoint of the study is ORR in the first 6 cycles. The early endpoint at IA #1 is ORR in the first 3 cycles. The overall type I error will need to be controlled at one-sided $\alpha = 0.025$.

Let p be the true response rate in the first 6 cycles. The null hypothesis to be tested is below:

$$H_0: p \leq 30\% \; vs. \; H_a: p > 30\%$$

Overall type I error α

$$= Prob\left(reject\ H_0 \mid H_0\ is\ true\right)$$

$$= Prob\left(The\ specific\ dose\ level\ is\ efficacious\ at\ the\ final\ primary\ analysis \mid H_0\ is\ true\right)$$

$$= Prob\begin{pmatrix} Pass\ the\ early\ futility\ boundary\ in\ Stage\ 1\ and \\ Pass\ the\ futility\ boundary\ at\ end\ of\ Stage\ 1\ and \\ Claim\ efficacy\ at\ end\ of\ Stage\ 2 \mid H_0\ is\ true \end{pmatrix}$$

According to the conditional probability chain rule, the above probability equals to the product of three probabilities:

Probability #1:

$$Prob\left(Pass\ the\ early\ futility\ boundary\ in\ Stage\ 1 \mid H_0\ is\ true\right) = Prob(r_{3c} > d_{3c} | p = p_0 = 30\%)$$

Probability #2:

$$Prob\begin{pmatrix} Pass\ the\ futility\ boundary\ at\ end\ of\ Stage\ 1 \mid H_0\ is\ true\ and \\ Pass\ the\ early\ futility\ in\ Stage\ 1 \end{pmatrix}$$
$$= Prob\left(r_1 > d_1 | r_{3c} > d_{3c}\ \&\ p = p_0 = 30\%\right)$$

Probability #3:

$$Prob\begin{pmatrix} Claim\ efficacy\ at\ end\ of\ Stage\ 2 \mid H_0\ is\ true\ and \\ Pass\ the\ early\ futility\ in\ Stage\ 1\ and \\ Pass\ the\ futility\ boundary\ at\ end\ of\ Stage\ 1 \end{pmatrix}$$
$$= Prob\left(r \geq d | r_1 > d_1\ \&\ r_{3c} > d_{3c}\ \&\ p = p_0 = 30\%\right)$$

r_{3c}, r_1, r_2, r represent the responders in the first 3 cycles among Stage 1 subjects, responders in the first 6 cycles at the end of Stage 1, responders in the first 6 cycles in Stage 2, total number of responders in Stages 1 and 2, respectively.

d_{3c}, d_1, d represent the futility boundary in the first 3 cycles among Stage 1 subjects, the futility boundary in the first 6 cycles at the end of Stage 1, and the efficacy boundary in the first 6 cycles at the end of Stage 2.

The number of responders in each stage follow the binomial distribution:

$r_{3c} \sim Binomial(n_1, p_{3c})$, $r_1 - r_{3c} \sim Binomial(n_1 - r_{3c}, \pi)$ and $r_2 \sim Binomial(n_2, p)$.

(π is the conditional probability of the response after 3 cycles given that no response in the first 3 cycles).

Based on the formulae and computational algorithm above, as well as assumptions in the study, If there is over enrollment in the second stage, the efficacy boundary can be re-calculated on the overall type I error $\alpha = 0.025$.