A Randomized, Double-Blind, Placebo-Controlled Multicenter Phase III Trial of Alpha 1 – Antitrypsin (AAT) Combined with Corticosteroids vs Corticosteroids Alone for the Treatment of High Risk Acute Graft-versus-Host Disease (GVHD) Following Allogeneic Hematopoietic Stem Cell Transplant

BMT CTN 1705/CSL964_5001

IND#: 18195

NCT#: 04167514

STATISTICAL ANALYSIS PLAN

Version 3.0

November 30, 2023

Effective by: December 7, 2023

VERSION HISTORY

Version numbers will always be whole numbers and any changes to the document require up-versioning.

Version #	Creation/ Revision Date (DDMMMYYYY)	Approved By	Effective Date (DDMMMYYYY)	Reason for Change
1.0	24FEB2022	PPD	11MAR2022	Initial document
2.0	04JUL2022	PPD	14JUL2022	Protocol Amendment
3.0	30NOV2023	PPD	07DEC2023	Protocol Amendment

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviations	Description of Abbreviations
aGVHD	Acute Graft-Versus-Host Disease
AAT	Alpha 1 - Antitrypsin
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine Aminotransferase
AREG	Amphiregulin
AST	Aspartate Aminotransferase
BMT	Blood and Marrow Transplant
BMT CTN	Blood and Marrow Transplant Clinical Trials Network
BSA	Body Surface Area
Cmax	Maximum Concentration
Ctrough	Trough Concentration
CBC	Complete Blood Count
CI	Confidence Interval
CLL	Chronic Lymphocytic Leukemia
CMV	Cytomegalovirus
COVID-19	Coronavirus Disease 2019
CR	Complete Response
CRF	Case Report Form
CS	Corticosteroids
CT	Computerized Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DOR	Duration of Response
DSMB	Data and Safety Monitoring Board
EOT	End of Therapy
FDG	Fluoro-Deoxyglucose
GI	Gastrointestinal
GVHD	Graft-Versus-Host Disease
HSCT	Hematopoietic Stem Cell Transplantation
HQL	Health Quality of Life
IL	Interleukin
IMP	Investigational Medicinal Product
IND	Investigational New Drug
IRT	Interactive Response Technology
ITT	Intention-to-Treat
IV	Intravenous
mITT	Modified Intention-to-Treat
MAGIC	Mount Sinai Acute GVHD International Consortium
MDASI	MD Anderson Symptom Inventory

Abbreviations	Description of Abbreviations
MDS	Myelodysplastic Syndromes
MN	Refined Minnesota Risk
MOP	Manual of Procedures
MP	Methylprednisolone
MR	Mixed Response
NCT	National Clinical Trial
NHLBI	National Heart, Lung, and Blood Institute
NIH	National Institutes of Health
NK	Natural Killer
NR	No Response
NRM	Non-Relapse Mortality
ORR	Overall Response Rate
OS	Overall Survival
PCR	Polymerase Chain Reaction
PET	Positron Emission Tomography
PFS	Progression-Free Survival
PK	Pharmacokinetics
PR	Partial Response
PROMIS	Patient-Reported Outcomes Measurement Information System
PTM	Placebo to Match
REG3α	Regenerating Family Member 3 Alpha
SAE	Serious Adverse Events
SAP	Statistical Analysis Plan
ST2	Suppression of Tumorigenicity 2
TD	Treatment Discontinuation
Teff	T Effector Cell
TF	Treatment Failure
TLFs	Tables, Listings, and Figures
TNFα	Tumor Necrosis Factor Alpha
Treg	T Regulatory Cell
VGPR	Very Good Partial Response

1 INTRODUCTION

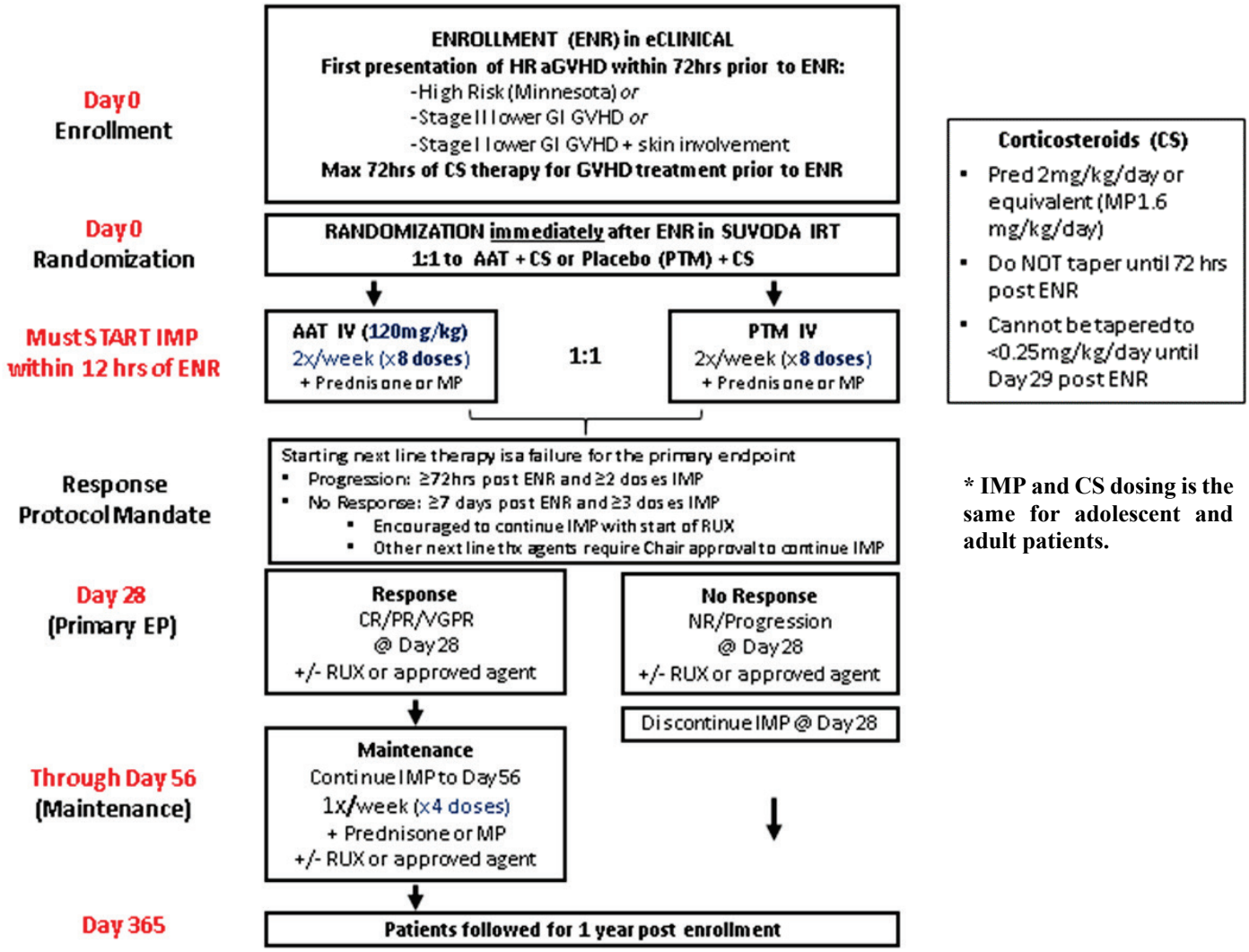
This Statistical Analysis Plan (SAP) elaborates upon the analysis strategy introduced in the study protocol and includes detailed procedures for completing the statistical analysis of efficacy and safety endpoints.

The content herein is based on BMT CTN Protocol 1705 version 5.0. In order to prevent bias from arising in the analysis, Versions 1.0 and 2.0 of the SAP were finalized and signed before the first analysis of study data at the planned interim analysis. If required, further revisions to the approved SAP may be made prior to the database hard lock. Future revisions will be version controlled.

This statistical analysis is coordinated by the responsible biostatistician at Emmes. Any changes to the analyses described in the SAP will be detailed and justified in the final analysis report.

2 STUDY SCHEMA AND PATIENT ASSESSMENT SCHEDULE

2.1 STUDY SCHEMA



2.2 SCHEDULE OF ASSESSMENTS

Table 1: Required Assessments Post-Enrollment

			Days												
	Pre-Enrollment	**Pre-Therapy[1]**	**7***	**14***	**21***	**28***	**35***	**42***	**49***	**56 ***	**86 (3 mos)****	**120****	**6 mos*****	**EOT[2] (+ 30 days)**	**12 mos********
History, physical exam, weight and height	X														
Karnofsky/Lansky Performance Status	X					X				X	X		X		X
Pregnancy test if applicable[3]	X					X				X					
Acute GVHD evaluation	X	X	X	X	X	X	X	X	X	X	X		X		X
Liver Function Tests (alkaline phosphatase, bilirubin, AST, ALT)	X		X	X	X	X	X	X	X	X	X		X		X
CBC with differential, platelet count	X			X		X				X	X				
Immunosuppressive drug monitoring (e.g., sirolimus, tacrolimus, cyclosporine)		X	X	X	X	X	X	X	X	X	X		X	X	X
Determination of corticosteroid dose, including administered dose and patient weight	X	X	X	X	X	X	X	X	X	X	X		X	X	X
Chronic GVHD evaluation[7]						X				X	X		X		X
Toxicity assessment												X	X		
Infection monitoring[4]		Reporting of infections at time of occurrence from time of enrollment through 3 months													
AE/SAE assessments[5]		X	X	X	X	X	X	X	X	X	X		X[5]	X[5]	X[5]
Patient-reported outcomes[6]: MDASI, PROMIS		X				X				X			X		

Notes:

* +/- 3 days from the target day to allow for scheduling flexibility, holidays, etc.

** +/- 14 days from the target day to allow for scheduling flexibility

*** +/- 28 days to allow for scheduling flexibility

**** 60 days to allow for scheduling flexibility

1 Pre-therapy assessments preferably should be performed prior to enrollment; however, up to 36 hours post-enrollment is allowed.

2 End of Treatment assessments (except specimen collection) for patients that terminate treatment early will be done 30 days after the last dose of study drug. A +/- 3-day window is allowed.

3 Pregnancy test (by blood) for females of childbearing potential within 30 days prior to enrollment. Serum or urine test on Day 28 and Day 56.

4 Includes anti-viral prophylaxis

5 AEs and SAEs will be reported from randomization until 30 days post last dose. Beginning at 31 days from the last dose of study drug through the end of study follow-up, SAEs deemed to be related to the study drug require reporting, and any unanticipated SAE regardless of relationship to study drug require reporting.

6 Only English- and Spanish-speaking patients are eligible to participate in the Health Quality of Life (HQL) component of this trial. MDASI surveys to be completed by both English- and Spanish-speaking patients. PROMIS surveys to be completed by English speaking patients only. Surveys may not be completed by surrogates.

7 The Chronic GVHD Provider Survey is due at each Chronic GVHD evaluation and will record GVHD symptoms present in the last week and must be completed by a clinician on the day of the assessment.

Table 2: Required and Optional Assessments Post-Treatment Initiation

Type	Purpose	Day							
		Pre-Therapy[2]	0	8	16	24	28	56	TD[3]
Required	Study drug levels[1]	X	X	X	X	X	X	X	
	Serum Inflammatory Cytokines and GVHD Biomarkers[4]	X		X			X	X	X
	Flow Cytometry[4]	X			X		X	X	X
Optional *(at select centers)*	Stool research samples[5]	X		X			X		X

Notes: Enrollment is Day 0. First dose of study drug must occur within 12 hours after enrollment and may fall on Day 0 or Day 1.

1 Samples should be drawn just prior to each study drug infusion (and also within 15 to 45 minutes post-end of infusion for the Day 0, 16, 28, and 56 samples).

2 Collect sample prior to administering study drug.

3 The sample should be collected at treatment discontinuation (TD) if that occurs prior to Day 28 (stool)/56 (blood). (Note: This is at the time that patient discontinues the study drug.)

4 The collection of required research blood samples for serum biomarkers and flow cytometry studies should be coordinated with the collection of the pre-drug infusion blood samples for AAT level measurements. If the study drug is administered before or after the indicated day, but within the treatment window (Section 2.5.1.7), the timing of the research sample collection can be adjusted accordingly. If not possible for some reason to not coordinate the research sample collection, the window of collection for research blood samples are ± 3 day.

5 Stool samples being collected from consenting patients at selected centers with active Microbiome laboratories should be collected if possible, on the target visit day indicated or within ± 3 days of the scheduled visit.

3 STUDY OBJECTIVES AND DESIGN

3.1 STUDY OBJECTIVE

BMT CTN protocol #1705 a phase III, multicenter, double-blinded, randomized, placebo-controlled trial designed to compare AAT and corticosteroids (CS) to matched placebo/PTM and CS as first-line therapy for patients with high-risk acute GVHD.

3.2 PRIMARY OBJECTIVE

The primary objective of this trial is to compare the rate of overall (complete or partial) response on Day 28 post-randomization between AAT and CS versus placebo to match (PTM) and CS in patients with high-risk acute GVHD.

3.3 SECONDARY OBJECTIVES

Secondary objectives of the study are to assess the following:

1. Duration of response at 6 and 12 months post-randomization.
2. Cumulative incidence of non-relapse mortality (NRM) at 6 and 12 months post-randomization.
3. Overall survival (OS) and progression-free survival (PFS) at 6 and 12 months post-randomization.
4. GVHD-free survival at Day 56 post-randomization.
5. Proportions of complete response (CR), very good partial response (VGPR), partial response (PR), and treatment failure (TF) at Days 7, 14, 21, 28, 56, and 86 post-randomization.
6. Proportions of CR, very good partial response (VGPR), PR, and treatment failure (TF) at Days 7, 14, 21, 28, 56, and 86 post-randomization for patients who receive Ruxolitinib or other second line therapies approved by the protocol Chairs as next-line therapy and remain on AAT/PTM.
7. Incidence of systemic infections to assess safety.
8. Incidence of adverse events (AEs) at 30 days post last dose of the drug to assess safety.
9. Incidence of chronic GVHD at 6 and 12 months post-randomization.
10. Incidence of disease relapse at 6 and 12 months post-randomization.

3.4 EXPLORATORY OBJECTIVES

The exploratory objectives of the study are to assess the following:

1. AAT levels in serum at Days 0, 8, 16, 24, 28, and 56 post-treatment initiation.
2. Stool concentrations of AAT at baseline and at Days 8 and 28 post-treatment initiation.
3. Blood ratios of T regulatory to T effector cells (Treg/Teff), Natural Killer (NK) cells, and cellular immune subsets at baseline and at Days 16, 28, and 56 post-treatment initiation.

4. Serum levels of inflammatory cytokines and biomarkers at baseline and at Days 8, 28, and 56 post-treatment initiation.
5. Overall and organ-specific response rates comparison based on Refined Minnesota (MN) risk groups (Revised Minnesota high vs. standard) and organ-specific response rates comparison based on biomarker-based risk groups.
6. Corticosteroid-dose at baseline, Days 7, 14, 21, 28, 56, 86, 6 months and 12 months post-randomization.
7. CMV reactivation requiring therapy by Day 56 post-randomization.
8. Change in patient-reported outcomes from baseline to Days 28, 56 and 6 months post-randomization.

3.5 STUDY DESIGN

This study is a phase III, multicenter, double-blinded, randomized, placebo-controlled trial designed to compare AAT and CS to PTM and CS as first-line therapy for patients with high-risk acute GVHD. The targeted enrollment is 136 patients, accrued over three years from approximately 25 treatment centers with patients randomly allocated at a 1:1 ratio to treatment arms (68 patients per arm). Patients will be followed for one year following enrollment.

The primary endpoint is the overall (complete or partial) response to acute GVHD treatment at Day 28 post-randomization. Patients without a Day 28 response assessment will be considered failures. Acute GVHD will be graded and assessed for response based on Mount Sinai Acute GVHD International Consortium (MAGIC) Criteria (Harris et al. 2016).

The primary hypothesis is that the overall (CR+PR) response rate of high risk acute GVHD will be improved under treatment by AAT and CS compared to PTM and CS. Statistically, this will be evaluated by testing the null hypothesis that the odds ratio of overall response for the AAT+CS versus PTM+CS treatment arms is equal to 1 versus the alternative hypothesis that the odds ratio is greater than 1 after adjustment for the treatment center-group (see definition in Section 7.4.1) and Minnesota (MN) risk category (high vs. standard). An odds ratio greater than 1 indicates that the response rate is higher in the AAT+CS group compared with the PTM+CS group. This null hypothesis will be tested using a one-sided significance level of 2.5%.

One interim analysis for futility is included, to be performed once 76 patients become evaluable for the primary endpoint. There will be no early stopping for efficacy based on the interim analysis results.

To be eligible for this study, a patient must be experiencing an initial, clinically diagnosed presentation of acute GVHD after allogeneic transplant that requires systemic therapy with CS. Moreover, the acute GVHD must exhibit one of the following clinical features within 72 hours prior to enrollment: (A) high risk by Refined Minnesota Criteria (Macmillan et al. 2015, defined below) or (B) isolated stage 2 involvement of the lower gastrointestinal (GI) tract or stage 1 lower GI tract disease with skin involvement.

High-risk by Refined Minnesota Criteria (any one below):

- **Single organ involvement**
  a. Stage 4 skin
  b. Stage 3-4 lower GI
  c. Stage 1-4 liver
- **Multiple organ involvement**
  a. Stage 1-2 lower GI plus any liver
  b. Stage 2 lower GI plus any skin
  c. Stage 3-4 lower GI plus any liver or skin
  d. Any three-organ involvement

Under Versions 1.0 - 3.0 of the protocol, patients 18 years of age or older at enrollment were eligible for this study. Following release of Version 4.0, patients 12 years or older are eligible for this study.

Other inclusion/exclusion criteria can be found in section 2.4 of the study protocol.

Study drug (AAT/PTM) must be initiated as soon as possible after enrollment and no longer than 12 hours after enrollment. Each dose of study drug will be 120 mg/kg (actual weight) based on the patient's weight recorded at enrollment. Subsequent doses should be given on Days 4, 8, 12, 16, 20, 24, and 28. The patient's weight recorded at enrollment will be used to determine the dose for all infusions (i.e., dose adjustments for changes in weight during the treatment period will not be made). Responding patients (CR/PR at Day 28), compared to maximum GVHD organ staging within 72 hours prior to enrollment, that are receiving study drug alone or study drug plus Ruxolitinib or other approved next-line therapy will continue to receive treatment at a dose of 120mg/kg (actual weight) on Days 35, 42, 49, and 56. Patients who are not in CR/PR at Day 28 will not continue to receive the study drug. Responding patients at Day 28 who progress after Day 28 may continue to receive weekly doses through Day 56. Patients that receive next-line therapy other than ruxolitinib may not continue receiving study drug unless approval is granted by the BMT CTN 1705 Protocol Chairs. All treatments will have a window of +/- 3 days to allow for efficient scheduling of patients for holidays/weekends/etc.

3.6 RANDOMIZATION

Participants will be randomized at a 1:1 ratio between the AAT and PTM treatment arms using permuted blocks of random sizes. Randomization will be stratified by treatment center. Randomization will be conducted in the Suvoda IRT system after enrollment has been completed in Advantage eClinical.

4 SAMPLE SIZE AND POWER CONSIDERATIONS

4.1 HISTORICAL DATA

The high risk acute GVHD patients included in the study population include patients with Minnesota (MN) high risk classification per the Refined Minnesota criteria, MN standard risk patients with isolated stage 2 involvement of the lower GI tract or stage 1 lower GI tract disease with skin involvement. To determine the appropriate sample size for this study, historical data were used to determine the Day 28 overall response rate expected in this population under CS treatment. In developing the Refined Minnesota risk criteria, Macmillan et al. 2015 used acute GVHD data collected from 1723 patients across multiple centers to estimate GVHD response rates in their acute GVHD risk classification of Minnesota (MN) high vs. standard risk. Among MN high risk patients in this cohort, the Day 28 overall response rate was 43% (Figure 1).

Additionally, data from patients presenting with lower GI acute GVHD in the BMT CTN 0302 and 0802 acute GVHD treatment trials and at MD Anderson Cancer Center during 2009-2012 were examined. Among these patients, those with isolated stage 2 lower GI or stage 1 lower GI and skin involvement had observed Day 28 overall response rates of 50-57% (highlighted rows in Table 3), indicating a similar level of risk as to the MN high risk category despite being classified as “standard risk” per these criteria. Therefore, they are included in the study population to increase the set of study eligible patients that may benefit from AAT+CS treatment. As a whole, these historical control data suggest a Day 28 overall response rate of approximately 50% for the study population.

Figure 1: Comparison of Day 28 Overall Response (CR+PR) Rates in MN Standard vs. High Risk acute GVHD patients (Macmillan et al. 2015).

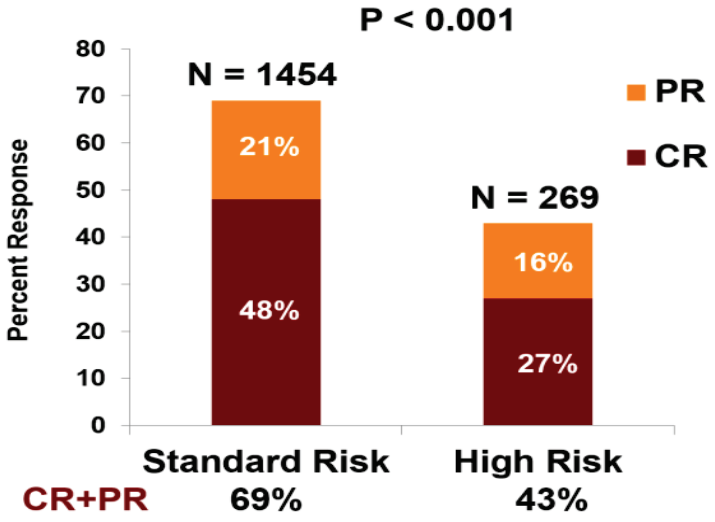


Table 3: Observed Day 28 Response Rates for Patients with Lower GI GVHD from BMT CTN 0302/0802 Trials and MD Anderson Cancer Center

	MD Anderson Cancer Center		BMT CTN 0302/0802 Trials	
Organ Involvement	**N**	**Day 28 CR/PR Rate**	**N**	**Day 28 CR/PR Rate**
Isolated Lower GI:				
Stage 1	37	73%	8	75%
Stage 2	14	50%	9	56%
Stage 3-4	23	48%	15	67%
Lower GI + any liver:				
Stage 1	9	56%	7	43%
Stage 2	5	40%	8	50%
Stage 3-4	12	17%	5	0%
Lower GI + any skin:				
Stage 1	19	53%	30	57%
Stage 2	10	50%	17	41%
Stage 3-4	17	35%	17	41%

* Highlighted rows show Minnesota standard risk subgroups that were selected for inclusion in BMT CTN 1705 based in similar observed CR/PR rates to Minnesota high risk patients

4.2 STATISTICAL DESIGN AND SAMPLE SIZE DETERMINATION

The primary hypothesis is that the odds ratio for the overall response (CR+PR) at Day 28 between the treatment groups, adjusted for treatment center and GVHD risk (MN high vs. standard risk), is greater than 1 (AAT+CS versus PTM+CS). The primary analysis will be based on a Wald test from a logistic mixed model with the overall response as the outcomes, testing the odds ratio of treatment assignment, as described in section 6.4.1.1. The required sample size for this study is approximated using a two-sample Z-test of the difference in response rates. Historical control data indicate that the overall response rate at Day 28 for high risk acute GVHD is approximately 50%. We anticipate that AAT+CS treatment will increase the Day 28 overall response rate to 75%, corresponding to a 25% effect size. To perform a one-sided test of the superiority of AAT+CS over PTM+CS with one interim analysis for futility, we use a two-stage design with one interim analysis for futility based on a Gamma family error spending function with $\gamma = -8$. The interim futility analysis was chosen to occur after 76 patients are evaluable for Day 28 response, at a 56% information fraction. This design specification and the two-sample Z-test sample size calculation imply that a sample size of 136 patients (68 per arm) is required to sufficiently maintain a one-sided type I error of 2.5% while providing 86% statistical power for a one-sided test to detect a 25% effect size.

Operating characteristics of this sequential design with a two-sample Z-test of response rates are shown in Table 4 as an approximation to the operating characteristics provided by a logistic mixed model in this setting. Kahan and Morris 2012 demonstrated that a stratified analysis improves statistical power for a specific set of analysis techniques. Based on their findings, we expect that the power provided by an unadjusted analysis via a two-sample Z-test comparing proportions is a lower bound to the power provided by the specified logistic mixed model.

Table 4: Group Sequential Design Stopping Thresholds for Z-test of Proportions with One-Sided Type I Error 2.5%, Power 86% and an Effect Size of 25%

			Futility Boundary Scales			Probability of Stopping Early for Futility	
Info. Fraction	Sample Size	Nominal Cumulative Type II Error Rate	Difference in Proportions	Z-Statistic	Conditional Power at Alternative	Under Null	Under Alternative
0.56	76	0.004	-0.034	-0.31	0.089	0.378	0.004
1.00	136	0.139	0.163	1.96	-	-	-

4.3 SIMULATION STUDY

A simulation study was conducted to evaluate the power provided by a logistic mixed model in this sequential design. Randomized treatment assignments were generated for each simulated trial using stratified randomized block design at a 1:1 ratio with permuted blocks of size 6, with center accrual numbers obtained from those observed in participating centers of the BMT CTN 1501 study. Centers with fewer than 10 patients were combined, resulting in 6 center-groups of size 23, 19, 17, 15, 14, and 48. We assumed these groups' random effects are a random sample from a zero-mean normal distribution. Additionally, we estimated that 65% of the patients would be classified as Refined Minnesota high risk, with GVHD risk (MN high vs. standard) included as a fixed effect in the logistic model for all simulations. We considered possible effect sizes of 10%, 15%, 20%, and 25% and standard deviations for the random effects of 0.05, 0.60, and 1.00, with 1000 replicated trials simulated for each scenario.

Simulation results in Table 5 indicate that, in the near absence of a random effect (i.e., when the standard deviation of the random effect is close to zero), the statistical power is at least 94.2% to detect an odds ratio of 3.0, corresponding to an effect size of 25%. Further, when the random effect is present (standard deviation of 0.6 or 1.0), statistical power is at least 92% and 90% to detect an odds ratio of 3.0. For any other given effect size, the power levels are similar regardless of the variability of the random effects. These results support the ability of the logistic mixed model to provide at least 86% power in this group sequential design.

Table 5: Empirical Power for Logistic Mixed Model (1000 simulations per scenario)

Effect Size (AAT vs. PTM)		Power (%)		
		Standard Deviation of Random Effects		
Difference in Day 28 CR/PR	Odds Ratio	0.05	0.60	1.00
10%	1.500	25.5	24.2	20.8
15%	1.857	53.8	51.3	48.4
20%	2.500	85.0	83.4	79.6
25%	3.000	94.2	92.9	90.1

5 ANALYSIS POPULATION

5.1 INTENTION TO TREAT POPULATION

The intention-to-treat (ITT) population consists of all randomized subjects, classified according to their randomized treatment assignments. All randomized subjects are included, regardless of whether the assigned study treatment was truly administered. This population comprises the primary efficacy analysis population for the study.

5.2 MODIFIED INTENTION TO TREAT POPULATION

The modified intention-to-treat (mITT) population consists of all subjects from the intention-to-treat population who initiate their randomized treatment.

5.3 SAFETY ANALYSIS POPULATION

The safety analysis population will include all randomized subjects who received at least one dose of study treatment. Subjects will be classified according to the study treatment received. If a patient receives a dose of the incorrect study treatment (study treatment mismatched to the randomized assignment), they will be classified as follows:

- If incorrect study treatment is received for the entire treatment period, the patient will be included in the treatment group corresponding to the study treatment received.
- If incorrect study treatment is received for only one dose and the correct treatment is received for all other doses, the patient will be included in the correct treatment group with a narrative provided for any events occurring during the period for which the subject is incorrectly dosed.
- If incorrect study treatment is received for more than one dose and the correct treatment is received for at least one other dose, the patient will be included in group corresponding to the majority of received doses; if equal numbers of doses of both treatments were received, they will be classified by their randomized assignment. A narrative will be provided for any events occurring during periods for which the subject was incorrectly dosed.

5.4 PHARMACOKINETIC ANALYSIS POPULATION

The pharmacokinetic (PK) analysis population will include all subjects who received at least one dose of study treatment and with at least 1 quantifiable serum AAT antigen concentration or serum AAT concentration with functional activity. This population will be analyzed using the treatment the subject actually received.

6 ANALYSIS VARIABLES

6.1 ACUTE GVHD RESPONSE DEFINITIONS

The response of acute GVHD to study treatment is evaluated by comparing the organ staging at the assessment time to the maximum GVHD organ staging within 72 hours prior to enrollment. Organ staging is performed in the acute GVHD target organs skin, upper GI, lower GI, and liver according to MAGIC criteria (Harris et al. 2016), which are shown in Table 6.

Table 6: Acute GVHD Organ Staging per MAGIC Criteria (Harris et al. 2016)

Stage	Skin	Lower GI	Upper GI	Liver
0	No active (erythematous) GVHD rash	Adult: Diarrhea 0-499mL/day or 0-3 episodes/day Child: 0.0-9.9mL/kg/day or 0-4 episodes/day	No or intermittent nausea, vomiting, or anorexia	Bilirubin < 2.0 mg/dL
1	Maculopapular rash < 25% BSA	Adult: Diarrhea 500-999mL/day or 3-4 episodes/day Child: 10.0-19.9mL/kg/day or 4-6 episodes/day	Persistent nausea, vomiting, or anorexia	Bilirubin 2.0-3.0 mg/dL
2	Maculopapular rash 25-50% BSA	Adult: Diarrhea 1000-1500mL/day or 5-7 episodes/day Child: 20.0-30.0mL/kg/day or 7-10 episodes/day	---	Bilirubin 3.1-6.0 mg/dL
3	Maculopapular rash > 50% BSA	Adult: Diarrhea > 1500mL/day or > 7 episodes/day Child: > 30.0mL/kg/day or > 10 episodes/day	---	Bilirubin 6.1-15.0 mg/dL
4	Generalized erythroderma (> 50% BSA) plus bullous formation and desquamation > 5% BSA	Severe abdominal pain with or without ileus or grossly bloody stool (regardless of stool volume)	---	Bilirubin >15.0 mg/dL

Response Definitions:

- Complete response (CR) is defined as having a stage of 0 in all evaluable organs.
- Very good partial response (VGPR) is defined as a quantitative and functional response that closely approximates CR. Patients must have all of the following clinical features:
    - No rash or rash involving <25% of body surface area (BSA) without bullae (residual faint erythema and hyperpigmentation do not count)
    - Serum bilirubin <2mg/dl or <25% of baseline at enrollment
    - Ability to tolerate food or enteral feedings with predominantly formed stool (no overt GI bleeding or cramping and no more than occasional nausea or vomiting)
- Partial response (PR) is defined as improvement in one or more acute GVHD target organs without progression in others. Note that VGPR is a subset of PR.
- Mixed response (MR) is defined as improvement in one or more organs with deterioration in another organ manifesting acute GVHD symptoms or development of acute GVHD symptoms in a new organ.
- No response (NR) is defined as the absence of improvement or progression in any organ.
- Progression is defined as deterioration in at least one organ without any improvement in others.

Patients who die will be classified as non-responders (NR). Patients who have an escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or higher will be classified as NR with the following exception: if a patient received a prednisone-equivalent dose of 2.5mg/kg/day or more on the randomization day, maintaining this starting dose is not automatically considered an NR, but escalation to a dose higher than the starting one

is an NR; moreover, if their dose is tapered below 2.5 mg/kg/day and later re-escalated to 2.5 mg/kg/day or above, they will be classified as an NR.

The receipt of next-line GVHD therapy also may impact how a patient is classified for GVHD response. This affects the following endpoints as described:

- **Overall Response at Day 28, Duration of Response, Acute GVHD Response to AAT/PTM, Overall and Organ Specific Response Rate by GVHD and Biomarker-based risk Classifications**: If a patient receives next-line GVHD therapy prior to a GVHD assessment visit, they will be classified as a non-responder (NR) for that assessment time.

- **Acute GVHD Response Allowing for Approved Next Line Therapy**: If a patient receives next-line GVHD therapy that is not approved by the protocol chairs prior to a GVHD assessment visit, they will be classified as a non-responder (NR) for that assessment time. However, the receipt of protocol chair-approved next-line therapy has no impact on the grading of GVHD response for this endpoint.

- **GVHD-free Survival:** If a patient receives next-line GVHD therapy while being event-free, they will be classified as an event at the time point of the therapy's initiation.

6.2 PRIMARY ENDPOINT – OVERALL RESPONSE AT DAY 28

Overall response at Day 28 is defined as having a complete or partial response at Day 28, i.e. having partial or complete resolution of acute GVHD symptoms in some organ(s) without worsening in any other(s) along with being alive, free of any next-line GVHD therapy, and free of escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more. Patients who have an escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or higher will be classified as NR with the following exception: if a patient received a prednisone-equivalent dose of 2.5mg/kg/day or more on the randomization day, maintaining this starting dose is not automatically considered an NR, but escalation to a dose higher than the starting one is an NR; moreover, if their dose is tapered below 2.5 mg/kg/day and later re-escalated to 2.5 mg/kg/day or above, they will be classified as an NR.

Organ staging is performed by MAGIC criteria, shown in Table 6. Patients with a missing Day 28 acute GVHD assessment will be classified as non-responders (NR). A Day 28 acute GVHD assessment conducted out of window (refer to Section 7.1 for visit window definition) will be considered as missing.

A supplementary analysis described in Section 7.4.3.4 will include adjustment for the refined Minnesota risk category and the MAGIC biomarker risk score (Srinagesh et al. 2019). The MAGIC biomarker risk score will be computed from baseline levels of ST2 and REG3α per the following algorithm:

- For each patient, compute their estimated 6 month non-relapse mortality risk $\hat{p}$ according to

$$log\big(-log(1-\hat{p})\big) = -11.263 + 1.844\, log_{10}\, ST2 + 0.577\, log_{10}\, REG3\alpha$$

- If $\hat{p} < 14.1\%$, biomarker score = 1; else if $\hat{p} > 29.0\%$, biomarker score = 3; else, biomarker score = 2

6.3 SECONDARY ENDPOINTS

For any time-to-event outcomes that follow, the analysis cutoff date will be defined as the date upon which a data freeze is performed for an analysis.

6.3.1 DURATION OF RESPONSE (DOR)

DOR is defined as the time from Day 28 response (CR or PR) until the first of the following events occurs: progression of acute GVHD, death, any next-line GVHD therapy, or escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more. DOR will be evaluated in the set of patients that attain a response (CR or PR) on Day 28. Progression is defined as the worsening by one stage or more from nadir in any acute GVHD target organ without improvement in other organs in comparison with the prior response. In the DOR context, the prior response refers to Day 28 response (CR or PR); nadir in an organ refers to Day 28 stage of this organ. For a patient in CR at Day 28, nadir is stage 0 in all target organs.

DOR, in days, is calculated as (Date of acute GVHD progression, death, next-line acute GVHD therapy, escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more, or last known to be alive) – (Date of Day 28 Response) +1.

Table 7: DOR Definition

Situation	Date of Event or Censoring	Outcome
Death, acute GVHD progression, next-line acute GVHD therapy initiation, or steroid dose escalation ≥ 2.5mg/kg/day before or on analysis cutoff date	Date of event	DOR event
Death, acute GVHD progression, next-line acute GVHD therapy initiation, or steroid dose escalation ≥ 2.5mg/kg/day; all events happened after analysis cutoff date	Analysis cutoff date	Censored
No event occurred, last known date alive is before or on analysis cutoff date	Last known date alive	Censored
No event occurred; last known date alive is after analysis cutoff date	Analysis cutoff date	Censored

DOR = Date of Event or Censoring – Date of Day 28 Response + 1

6.3.2 NON-RELAPSE MORTALITY (NRM)

NRM is defined as death without prior relapse/progression for the primary disease. Relapse/progression is treated as a competing risk for NRM and defined in section 6.3.4. The NRM time is defined as the time from the date of randomization until the documented date of relapse/progression or death from any cause, whichever happens first. The outcome type is defined as shown in Table 8. All relapses/progressions and deaths occurring on or after the randomization date will be included. Patients who are event-free at the time of analysis will be censored at the last day known to be alive or at the analysis data freeze date, whichever is earlier. Patients experiencing event(s) after the analysis cut-off date will be censored at the analysis cut-off date.

Table 8: NRM and Relapse/progression Competing Risks Definition

Situation	Date of Event or Censoring	Outcome
Relapse/progression occurred before or on analysis cutoff date	Date of relapse/progression	Relapse/ progression
Death without prior relapse/progression occurred before or on analysis cutoff date	Date of death	NRM
Death and/or relapse/progression occurred, all events happened after analysis cutoff date	Analysis cutoff date	Censored
No death or relapse/progression occurred, last known date alive is before or on cutoff date	Last known date alive	Censored
No death or relapse/progression occurred; last known date alive is after the cutoff date	Analysis cutoff date	Censored

Time of NRM = Date of Event or Censoring – Date of Randomization + 1

6.3.3 OVERALL SURVIVAL (OS)

OS is defined as the absence of death; failures for OS include deaths from any causes. The OS time is defined as the time from the date of randomization until the documented date of death from any cause. All deaths occurring on or after the randomization date will be included. Patients who are still alive at the time of analysis will be censored at the last day known to be alive or at the analysis data freeze date, whichever is earlier. Patients who died after the analysis cut-off date will be censored at the analysis cut-off date.

Time of OS, in days, is calculated as (Date of death or last known to be alive) – (Date of Randomization) +1.

Table 9: OS Definition

Situation	Date of Event or Censoring	Outcome
Death before or on analysis cutoff date	Date of death	Death
Death after analysis cutoff date	Analysis cutoff date	Censored
Last known date alive is before or on cutoff date	Last known date alive	Censored
Last known date alive is after the cutoff date	Analysis cutoff date	Censored

Time of OS = Date of Event or Censoring – Date of Randomization + 1

6.3.4 PROGRESSION FREE SURVIVAL (PFS)

PFS is defined as being alive and free of relapse/progression of the primary disease. Failures for PFS include death from any cause and disease relapse/progression. The PFS time is defined as the time from the date of randomization until the documented date of disease relapse/progression or death from any cause, whichever happens first.

Malignancy Relapse/progression Definition:

A subject's underlying disease (i.e., malignancy) relapse/progression information is collected in the case report form. Relapse is defined by either morphological or cytogenetic evidence of acute leukemia or myelodysplastic syndrome (MDS), consistent with pre-

transplant features, or radiologic evidence of lymphoma, documented or not by biopsy. Progression of disease applies to patients with lymphoproliferative diseases (lymphoma or chronic lymphocytic leukemia) not in remission prior to transplantation. The event is defined as an increase in the size of prior sites of disease or evidence of new sites of disease, documented or not by biopsy. In addition, the institution of any therapy to treat persistent, progressive, or relapsed malignancy, including the withdrawal of immunosuppressive therapy or donor lymphocyte infusion, will be considered evidence of relapse/progression regardless of whether the criteria described above were met.

- ***Acute Leukemia:*** Relapse will be diagnosed when there is:
  - Reappearance of leukemia blast cells in the peripheral blood; or
  - > 5% blasts in the bone marrow not attributable to another cause (e.g. bone marrow regeneration)
  - The development of extramedullary leukemia or leukemic cells in the cerebral spinal fluid
- ***MDS:*** Relapse will be diagnosed when there is:
  - Satisfaction of criteria for evolution into acute myeloid leukemia (blasts greater than 20%), or
  - Reappearance of pre-transplant morphological abnormalities detected in bone marrow specimens combined with a decrease in PB chimerism, OR
  - Reappearance of pre-transplant cytogenetic abnormality in at least one metaphase by routine metaphase cytogenetics or a positive test by FISH on each of two separate consecutive examinations at least one month apart, regardless of the number of metaphases analyzed
  - For patients with:
    - Less than 5% blasts: greater or equal to 50% increase in blasts to greater or equal to 5% blasts confirmed by flow cytometry
    - 5%-10% blasts: greater or equal to 50% increase to greater or equal to 10% blasts confirmed by flow cytometry

Lymphoproliferative Diseases: Relapse or progression will be diagnosed when there is:

- The appearance of any new lesion more than 1.5 cm in any axis during or at the end of therapy, even if other lesions are decreasing in size. Increased fluoro-deoxyglucose (FDG) uptake in a previously unaffected site will only be considered relapsed or progressive disease after confirmation with other modalities. In patients with no prior history of pulmonary lymphoma, new lung nodules identified by CT are mostly benign. Thus, a therapeutic decision should not be made solely based on the positron emission tomography (PET) without histologic confirmation.

- At least a 50% increase from the nadir in the sum of the product diameters of any previously involved nodes, or in a single involved node, or the size of other lesions (e.g., splenic or hepatic nodules). To be considered a progressive disease, a lymph node with a diameter of the short axis of less than 1.0 cm must increase by ≥ 50% and to a size of 1.5 x 1.5 cm or more than 1.5 cm in the long axis.
- Lesions should be PET positive if observed in a typical FDG-avid lymphoma or the lesion was PET positive before therapy unless the lesion is too small to be detected with current PET systems (<1.5 cm in its long axis by CT).
- In addition to the criteria above, patients with chronic lymphocytic leukemia (CLL) who present in complete remission prior to transplantation may fulfill the relapse definition if there is a reappearance of circulating malignant cells that are phenotypically characteristic of CLL.

- ***Other Malignant Diseases:*** Probable or definitive evidence of progression must be documented by biopsy, imaging, or peripheral blood findings that are diagnostic or otherwise consistent with that disease state.

Non-malignant Diseases: Patients with non-malignant diseases will be considered to have a transplant status of persistent/active disease. Graft failures will be considered a relapse of the disease.

All relapses/progressions and deaths occurring on or after the randomization date will be included. Patients who are still alive and free of relapse/progression at the time of analysis will be censored at the last day known to be alive and progression-free or at the analysis data freeze date, whichever is earlier. Patients experiencing relapse/progression or death after the analysis cut-off date will be censored at the analysis cut-off date.

Time of PFS, in days, is calculated as (Date of relapse/progression, death, or last known to be alive) – (Date of Randomization) +1.

Table 10: PFS Definition

Situation	Date of Event or Censoring	Outcome
Death or relapse/progression occurred before or on analysis cutoff date	Date of relapse/progression or death	PFS event
Death and/or relapse/progression occurred, all events happened after analysis cutoff date	Analysis cutoff date	Censored
No death or relapse/progression occurred, last known date alive is before or on analysis cutoff date	Last known date alive	Censored
No death or relapse/progression occurred; last known date alive is after analysis cutoff date	Analysis cutoff date	Censored

Time of PFS = Date of Event or Censoring – Date of Randomization + 1

6.3.5 GVHD-FREE SURVIVAL

GVHD-free survival is defined as being alive, free of active acute and chronic GVHD, free of any next-line GVHD therapy, and free of escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more. The following exception is permitted for steroid escalation: if a patient received a prednisone-equivalent dose of 2.5mg/kg/day or more on the randomization day, maintaining this starting dose is not automatically considered a failure, but escalation to a dose higher than the starting one is a failure; moreover, if their dose is tapered below 2.5 mg/kg/day and later re-escalated to 2.5 mg/kg/day or above, they will be classified as a failure. Acute GVHD is defined per the MAGIC criteria (Harris et al. 2016) with grade 0 referring to no acute GVHD. Chronic GVHD is defined per the 2014 NIH Consensus Criteria. The proportion of patients with GVHD-free survival will be evaluated at Day 56 post-randomization. For the analysis of GVHD-free survival, a patient with a missing acute or chronic GVHD assessment will be considered to be a failure at the corresponding GVHD assessment Day. An acute or chronic GVHD assessment conducted out of window (refer to Section 7.1 for visit window definition) will be considered as missing.

6.3.6 ACUTE GVHD RESPONSE TO AAT/PTM

This endpoint focuses specifically upon the effect of the randomized treatment (AAT or PTM) alone on acute GVHD response, following a similar approach as for the primary endpoint at Day 28. The proportions of patients with CR, PR (including the subset with VGPR), and treatment failure (TF) will be described at Days 7, 14, 21, 28, 56, and 86 post-randomizations. The designation of TF consists of patients with NR, MR, and acute GVHD progression. The Day 86 response will be summarized only in patients remaining on study maintenance treatment (AAT/PTM) after Day 28. Patients who die, initiate either approved or unapproved next-line GVHD therapy, or have escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more prior to an acute GVHD assessment will be classified as non-responders (NR) at the corresponding Day with the following exception: if a patient received a prednisone-equivalent dose of 2.5mg/kg/day or more on the randomization day, maintaining this starting dose is not automatically considered an NR, but escalation to a dose higher than the starting one is an NR; moreover, if their dose is tapered below 2.5 mg/kg/day and later re-escalated to 2.5 mg/kg/day or above, they will be classified as an NR. Patients with a missing acute GVHD assessment will be classified as non-responders (NR) at the corresponding Day. An acute GVHD assessment conducted out of window (refer to Section 7.1 for visit window definition) will be considered as missing.

6.3.7 ACUTE GVHD RESPONSE ALLOWING FOR APPROVED NEXT-LINE THERAPY

This endpoint focuses upon the effect of the randomized treatment (AAT or PTM) on acute GVHD response, *possibly in conjunction with an approved next-line therapy*. The proportions of patients with CR, PR (including the subset with VGPR), and treatment failure (TF) will be described at Days 7, 14, 21, 28, 56, and 86 post-randomization. The designation of TF consists of patients with NR, MR, and acute GVHD progression. The Day 86 response will be summarized only in patients remaining on study maintenance treatment (AAT/PTM) after Day 28. Patients who die, *initiate next-line GVHD therapy that is not approved by the protocol chairs*, or have escalation of prednisone-equivalent steroid

dose to 2.5mg/kg/day or more prior to an acute GVHD assessment will be classified as non-responders (NR) at the corresponding Day with the following exception: if a patient received a prednisone-equivalent dose of 2.5mg/kg/day or more on the randomization day, maintaining this starting dose is not automatically considered an NR, but escalation to a dose higher than the starting one is an NR; moreover, if their dose is tapered below 2.5 mg/kg/day and later re-escalated to 2.5 mg/kg/day or above, they will be classified as an NR. Patients with a missing acute GVHD assessment will be classified as non-responders (NR) at the corresponding Day. The acute GVHD assessment conducted out of window (refer to Section 7.1 for visit window definition) will be considered as missing.

Note that this endpoint uses the same ITT population for analysis as the secondary endpoint defined in section 6.3.6, while the categorization of response for patients who receive a protocol chair approved next-line therapy differs between these two endpoints. The receipt of an approved next-line therapy will be categorized as a failure (NR) for the endpoint in 6.3.6, while for this secondary endpoint 6.3.7 the receipt of an approved next-line GVHD therapy is not automatically considered a failure (NR) and, consequently, such a patient is still assessed for GVHD response as specified in Section 6.1.

6.3.8 SYSTEMIC INFECTIONS

The incidence of Grade 2 to 3 systemic infections occurring from randomization until 30 days after the last dose of the study drug (AAT or PTM) will be described. Grade 2 to 3 systemic infections are defined according to BMT CTN Manual of Procedures (MOP). Death before infection will be treated as a competing risk. Infection time is defined as the time from the date of randomization until the documented date of infection onset or death from any cause, whichever happens first. The outcome type is defined as shown in Table 11. All infections and deaths occurring on or after the randomization date will be included. Patients who are alive and free of infection onset at the time of analysis will be censored at the last day known to be alive or at the analysis data freeze date, whichever is earlier. Patients experiencing event(s) after the analysis cut-off date will be censored at the analysis cut-off date.

Table 11: Systemic Infections Competing Risks Definition

Situation	Date of Event or Censoring	Outcome
Infection occurred before or on analysis cutoff date	Date of infection onset	Infection
Death without prior infection onset occurred before or on analysis cutoff date	Date of death	Death
Death and/or infection onset occurred, all events happened after analysis cutoff date	Analysis cutoff date	Censored
No death or infection onset occurred, last known date alive is before or on analysis cutoff date	Last known date alive	Censored
No death or infection onset occurred; last known date alive is after analysis cutoff date	Analysis cutoff date	Censored

Time of Infection = Date of Event or Censoring – Date of Randomization + 1

6.3.9 ADVERSE EVENTS

The incidence of Grade 3-5 treatment-emergent adverse events occurring from randomization until 30 days after the last dose of study drug (AAT or PTM) will be described, where the grade is defined per Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0.

6.3.10 CHRONIC GVHD

Chronic GVHD is defined per the 2014 NIH Consensus Criteria (Jagasia et al. 2015), with death before chronic GVHD treated as a competing risk. Chronic GVHD time is defined as the time from the date of randomization until the documented date of chronic GVHD onset of any severity (mild, moderate, severe, or unknown) or death from any cause, whichever happens first. All chronic GVHD onsets and deaths occurring on or after the randomization date will be included. Patients who are event-free at the time of analysis will be censored at the last day known to be alive or at the analysis data freeze date, whichever is earlier. Patients experiencing event(s) after the analysis cut-off date will be censored at the analysis cut-off date.

Table 12: Chronic GVHD Competing Risks Definition

Situation	Date of Event or Censoring	Outcome
Chronic GVHD onset occurred before or on analysis cutoff date	Date of chronic GVHD onset	Chronic GVHD
Death without prior chronic GVHD onset occurred before or on analysis cutoff date	Date of death	Death
Death and/or chronic GVHD onset occurred, all events happened after analysis cutoff date	Analysis cutoff date	Censored
No death or chronic GVHD onset occurred, last known date alive is before or on analysis cutoff date	Last known date alive	Censored
No death or chronic GVHD onset occurred, last known date alive is after analysis cutoff date	Analysis cutoff date	Censored

Time of chronic GVHD = Date of Event or Censoring – Date of Randomization + 1

6.3.11 RELAPSE/PROGRESSION OF THE PRIMARY DISEASE

A relapse/progression of the primary disease is defined as in section 6.3.4. Death without prior relapse/progression is treated as a competing risk for relapse/progression. Relapse/progression time is defined similarly as PFS time in section 6.3.4, with the outcome type defined as shown in Table 8. All relapses/progressions and deaths occurring on or after the randomization date will be included. Patients who are event-free at the time of analysis will be censored at the last day known to be alive or at the analysis data freeze date, whichever is earlier. Patients experiencing event(s) after the analysis cut-off date will be censored at the analysis cut-off date.

6.4 EXPLORATORY ENDPOINTS

6.4.1 SERUM AAT LEVELS

AAT levels will be assessed using measurements of serum antigen concentration and serum AAT concentration with functional activity (functional AAT concentration) just prior to infusion and 30 min (+/- 15 min) post-end of infusion on Day 0 and on Days 16, 28, and 56 post-treatment initiation; these levels will also be assessed prior to infusion on Days 8 and 24 post-treatment initiation. Ctrough value is the serum AAT concentration of the sample collected prior to drug infusion and Cmax value is the serum AAT concentration of the sample collected within 15 to 45 minutes post-end of infusion. Post-infusion measurements are limited to subjects receiving a dose of AAT/PTM.

6.4.2 STOOL AAT LEVELS

Stool concentrations of AAT at baseline and at Days 8 and 28 post-initiation of treatment (or at treatment discontinuation (TD) if before Day 28) will be collected at a subset of centers for a limited cohort of patients.

6.4.3 IMMUNE CELL SUBSETS

Peripheral blood ratios of T regulatory to T effector (Treg/Teff) cells and Natural Killer (NK) cell populations will be assessed at baseline and on Days 16, 28, and 56 post-treatment initiation (or at treatment discontinuation (TD) if before Day 56).

6.4.4 INFLAMMATORY CYTOKINES AND GVHD BIOMARKERS

Serum levels of IL-1β, IL-6, IL-10, TNFα, amphiregulin (AREG), heparin sulfate, ST2, and REG3α will be evaluated at baseline and at Days 8, 28, and 56 post-treatment initiation (or at treatment discontinuation (TD) if prior to Day 56).

6.4.5 OVERALL AND ORGAN SPECIFIC RESPONSE RATES BY GVHD AND BIOMARKER-BASED RISK CLASSIFICATIONS

The proportions of patients with an overall response (complete or partial) at Day 28 post-randomization will be described by the MN risk category (high vs. standard) and by Ann Arbor biomarker risk score (1 vs. 2 vs. 3). Additionally, the response rate of each GVHD target organ will be described on Day 28 by treatment arm, MN risk and Ann Arbor biomarker risk. An organ overall response is defined as a decrease in its staging from baseline, along with being alive, free of any next-line GVHD therapy, and free of escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more. The baseline staging refers to the maximum GVHD organ staging within 72 hours prior to enrollment. Patients who have an escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or higher will be classified as a non-responder with the following exception: if a patient received a prednisone-equivalent dose of 2.5mg/kg/day or more on the randomization day, maintaining this starting dose is not automatically considered a non-response, but escalation to a dose higher than the starting one is an non-response; moreover, if their dose is tapered below 2.5 mg/kg/day and later re-escalated to 2.5 mg/kg/day or above, they will be classified as an non-response. Patients will be classified as responders or non-responders for each target organ accordingly. Organ staging is performed by MAGIC criteria, shown in Table 6. Patients with a missing Day 28 acute GVHD assessment will be classified as non-responders for all organs. The Day 28 acute GVHD assessment

conducted out of window (refer to Section 7.1 for visit window definition) will be considered as missing. The organ-specific response rate will be evaluated only using patients with a staging greater than 0 in the corresponding organ at baseline.

The Ann Arbor biomarker risk score (Levine et al. 2015) will be computed from baseline levels of TNFR1, ST2, and REG3α. The algorithm is as follows:

- For each patient, compute their estimated 6 month non-relapse mortality risk $\hat{p}$ according to

$$log\big(-log(1-\hat{p})\big) = -9.169 + 0.598\, log_2\, TNFR1 - 0.028\, log_2\, REG3\alpha + 0.189\, log_2\, ST2$$

- If $\hat{p} \leq 10\%$, biomarker score = 1; else if $\hat{p} \geq 40\%$, biomarker score = 3; else, biomarker score = 2

6.4.6 SYSTEMIC CORTICOSTEROID DOSE

The prednisone-equivalent corticosteroid dose will be evaluated at baseline; on Days 7, 14, 21, 28, 56, 86; and at 6 and 12 months post-randomization. Methylprednisolone doses will be converted to prednisone-equivalent doses through multiplication by a factor of 1.25. Dexamethasone will be converted to prednisone-equivalent doses through multiplication by a factor of 6.67.

6.4.7 CMV REACTIVATION

Cytomegalovirus (CMV) reactivations to be assessed include those requiring new systemic treatment for a CMV PCR level per institutional practice (patients receiving only standard of care viral prophylaxis will not be included in this assessment). Death prior to CMV reactivation will be treated as a competing risk. CMV reactivation time is defined as the time from the date of randomization until the documented date of CMV reactivation or death from any cause, whichever happens first. All CMV reactivations and deaths occurring on or after the randomization date will be included. Patients who are alive and free of CMV reactivation at the time of analysis will be censored at the last day known to be alive or at the analysis data freeze date, whichever is earlier. Patients experiencing event(s) after the analysis cut-off date will be censored at the analysis cut-off date.

Table 13: CMV Reactivation Competing Risks Definition

Situation	Date of Event or Censoring	Outcome
CMV reactivation occurred before or on analysis cutoff date	Date of CMV reactivation	CMV reactivation
Death without prior CMV reactivation occurred before or on analysis cutoff date	Date of death	Death
Death and/or CMV reactivation occurred, all events happened after analysis cutoff date	Analysis cutoff date	Censored
No death or CMV reactivation occurred, last known date alive is before or on analysis cutoff date	Last known date alive	Censored
No death or CMV reactivation occurred, last known date alive is after analysis cutoff date	Analysis cutoff date	Censored

Time of CMV reactivation = Date of Event or Censoring – Date of Randomization + 1

6.4.8 PATIENT REPORTED OUTCOMES

Patient self-reported measures for both adolescents (participants age 12.00 - 17.99 years at enrollment) and adults (age 18.00 or older) will be assessed at baseline, Days 28, 56, and 6 months post-randomization using the MD Anderson Symptom Inventory (MDASI) and selected PROMIS subscales assessing the domains of GI symptoms, physical function, and satisfaction with participation in social roles. Adolescent versions of these assessments will be completed by adolescents.

6.5 SAFETY VARIABLES

Safety outcomes of interest include adverse events (AEs), serious adverse events (SAEs), and deaths. SAEs are defined as AEs that resulted in one of the following outcomes: death, a threat to life, requiring or prolonging inpatient hospitalization, causing persistent or significant disability, causing a congenital anomaly or birth defect, are assessed as important medical events, or requiring intervention to prevent any of the aforementioned outcomes. Treatment-emergent adverse events (TEAEs) are defined as AEs starting on or after the first infusion date of study drug.

An AE/SAE can be classified as anticipated or unanticipated:

- **Anticipated AEs/SAEs** are those that have been previously identified as resulting from the underlying disease, the HSCT, or acute GVHD and are not related to the study drug.
- **Unanticipated AEs/SAEs** are those that vary in nature, intensity, or frequency from information in the current anticipated event list, the Investigator's Brochure, the package insert, or when it is not included in the informed consent document as a potential risk. Unanticipated events would also include those that have not been previously described as a result of the underlying disease requiring HSCT, the HSCT, or acute GVHD**.**

*** The following are anticipated adverse events and do NOT require reporting as an AE/SAE:**

- Disease progression of primary malignancy
- GVHD and symptoms of GVHD (these are recorded in acute and chronic GVHD eCRFs)
- Infections
- Graft failure
- Medical or surgical procedures (condition that leads to the procedure is the AE)
- Situations where an untoward medical occurrence has not taken place. For example:
  - Planned hospitalizations due to pre-existing conditions, which have not worsened (e.g. HSCT)
  - Hospitalizations that occur for procedures not due to an AE (e.g. cosmetic surgery)
  - Hospitalizations for a diagnostic procedure where the hospital stay is less than 24 hours in duration or for normal management procedures (e.g. chemotherapy)

- Overdose of AAT or any concomitant therapy that does not result in any adverse signs or symptoms

*** Laboratory findings do NOT need to be reported as AEs in the following cases**:

- Laboratory parameters already beyond the reference range at screening, unless a further increase/decrease can be considered an exacerbation of a pre-existing condition.
- Abnormal hematological laboratory parameters considered anticipated due to myeloablative conditioning regimen, or other permitted chemotherapy treatments are not considered AEs.
- An abnormal laboratory value that cannot be confirmed after repeat analysis, preferably in the same laboratory (i.e., the previous result could be marked as not valid and should not necessarily be reported as an AE).

AEs and SAEs will be classified per CTCAE v5.0. All non-serious adverse events will be reported from the time of randomization until 30 days after the last dose of study drug (AAT/PTM). Anticipated non-serious adverse events will be collected on calendar- or event-driven forms. All SAEs are required to be reported from the time of randomization until 30 days following the last dose of study drug. In addition, any SAEs occurring after that 30-day period and assessed as unanticipated or related to the investigational product must be reported. All SAEs are to be followed up until resolved, judged to be no longer clinically significant, or until they become chronic to the extent that they can be fully characterized.

Any unanticipated SAEs from time of enrollment through the study defined follow up are required to be reported following the BMT CTN Administrative MOP Chapter 6. Additionally, any grade 4 anticipated AE not collected on the calendar-driven toxicity or a specified event-driven form must also be reported through the expedited AE reporting system in Advantage eClinical.

Additionally, CSL Behring has a list of events classified as Adverse Events of Special Interest (AESI). These include “anaphylactic infusional reaction” and “suspected transmission of a viral agent via a medicinal product.” Both of these events must be reported through the expedited AE reporting system via Advantage eClinical and must be reported within 24 hours of knowledge of the event even if it does not meet the serious criteria.

Toxicities are a subset of anticipated adverse events determined to be from the underlying disease process, HSCT, GVHD prophylaxis, or corticosteroid use, and generally, are not considered to be related to the study product. The toxicities listed in Appendix E of the study protocol were pre-determined by the Protocol Team. Once the participant has reached 31 days from the last dose of study drug, toxicities listed in Protocol Appendix E should be reported at the specified time points on the Toxicity CRF.

The primary cause of death will be described for patients that expired on the study.

7 STATISTICAL METHODOLOGY

7.1 GENERAL GUIDELINES

Counts and percentages will be used to describe categorical variables, while the number of subjects (N), median, mean, standard deviation, interquartile range and range will be used to summarize continuous variables.

The study day for most efficacy and safety assessments will be computed in reference to the date of randomization, with the date of randomization defined as Study Day 0 for each patient. For all assessments, the (Study) Day will be calculated as (assessment date – randomization date). Unless otherwise specified, the visit window will be defined similarly to those specified in the protocol with the target day calculated in reference to the date of randomization. For outcomes where data are summarized in reference to a different date, such as treatment initiation, the assessment time point is clarified explicitly.

For each patient-reported outcome, the baseline value is defined as the outcome score assessed at the Pre-Therapy visit on the schedule of assessments (**Section 2.2**). For all other variables, unless otherwise specified, the baseline value is defined as the last available measurement prior to the first infusion of study drug if the measurement time was collected and the last available measurement before or on the date of the first infusion of study drug if the measurement time was not collected.

Hypothesis testing for the primary endpoint's analysis is one-sided, evaluating the superiority of AAT to PTM for Day 28 overall response, at a 2.5% significance level. The testing for secondary endpoints will be two-sided at a 5% significance level and is equivalent to assessment by two one-sided tests, each at a 2.5% level. Testing of exploratory endpoints will be two-sided at a 1% significance level. **Table 14** summarizes the endpoints to be analyzed for each analysis population.

A hierarchical testing procedure will be employed to ensure control of the familywise type I error rate at a one-sided, 2.5% level when evaluating the primary and key secondary endpoints of interest. A gatekeeping approach will be used with the following tests of superiority of AAT to PTM for the corresponding endpoints:

1. Overall Response at Day 28: The Wald test from the primary analysis as specified in **Section 7.4.3.1** (one-sided at 2.5% level)
2. Duration of Response, Primary Definition: The log-rank test for comparing the DOR from Day 28 to 12 months post-randomization from the primary analysis as specified in **Section 7.5.1** (two-sided at 5% level)
3. Duration of Response, Supplementary Definition: The log-rank test for comparing the DOR from Day 28 to 12 months post-randomization from the supplementary analysis as specified in **Section 7.5.1** (two-sided at 5% level)
4. Non-relapse Mortality: The Gray's test for comparing the cumulative incidence of NRM from randomization to 12 months post-randomization as specified in **Section 7.5.2** (two-sided at 5% level)
5. Overall Survival: The log-rank test for comparing the OS from randomization to 12 months post-randomization from the primary analysis as specified in **Section 7.5.3** (two-sided at 5% level)

6. Progression-free Survival: The log-rank test for comparing the PFS from randomization to 12 months post-randomization from the primary analysis as specified in **Section 7.5.4** (two-sided at 5% level)
7. GVHD-free Survival at Day 56: The Wald test from the logistic mixed model as specified in **Section 7.5.5** (two-sided at 5% level)

The gatekeeping procedure evaluates the endpoints sequentially in the order displayed and ends if the one-sided test or two-sided test for an endpoint fails to find a significant result showing superiority of AAT to PTM, precluding further testing of any remaining endpoints in the hierarchy. This strategy ensures control of the one-sided familywise type I error rate at 2.5% and permits claims of confirmatory benefit for any endpoint(s) where superiority of AAT is demonstrated. Any hypothesis testing performed outside the hierarchy is deemed descriptive in this study.

Analyses for several endpoints involve adjustment for center effects through regression modeling. Because the average number of patients enrolled per center is expected to be small (136/25 ≈ 5.5), centers will be combined into center-groups for these analyses to ensure the stability of the models. This will be done by including the top 5 enrolling centers each as a separate group and combining all other centers into a sixth catchall group.

All efficacy analyses will be conducted using the ITT population unless otherwise specified.

Table 14: Endpoints for Analysis Under Each Analysis Population

Analysis Population	Primary Endpoint	Secondary Efficacy Endpoints	Exploratory Endpoints* except Serum AAT Levels	Serum AAT Levels	Safety Endpoints
Intention-to-Treat (ITT)	X	X	X		
Modified ITT	X				
Pharmacokinetic				X	
Safety					X

* The analysis populations related with biomarkers and stool AAT levels, if different from ITT population, will be defined and presented in a separate document depending on the data availability.

All data processing, summarization, and analyses will be performed using SAS Version 9.4 or higher and R version 3.5 or higher. Specifications for the table, figure, and data listing formats can be found in the supplementary TLF shell document for this SAP.

7.2 DEMOGRAPHICS AND DISPOSITION

7.2.1 DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Demographic and baseline characteristics will be summarized using descriptive statistics for the ITT population.

Categorical variables to be examined include age at randomization (less than 18 years vs. 18 - <55 years vs. 55 - <60 years vs. 60 years or older), gender, race, ethnicity, pre-transplant Karnofsky/Lansky performance status (<90 vs. >=90), pre-transplant primary disease, pre-transplant disease type (malignant vs. non-malignant), pre-enrollment MAGIC

acute GVHD grade (I vs. II vs. III vs. IV), pre-enrollment MAGIC acute GVHD skin stage (0 vs. 1 vs. 2 vs. 3 vs. 4), pre-enrollment MAGIC acute GVHD upper GI stage (0 vs. 1), pre-enrollment MAGIC acute GVHD lower GI stage (0 vs. 1 vs. 2 vs. 3 vs. 4), pre-enrollment MAGIC acute GVHD liver stage (0 vs. 1 vs. 2 vs. 3 vs. 4), pre-enrollment refined MN risk classification (high vs. standard), acute GVHD eligibility criterion, graft source (bone marrow vs. peripheral blood vs. umbilical cord), donor sex, donor relatedness (biologically related vs. not biologically related), baseline Ann Arbor biomarker risk score (1 vs. 2 vs. 3), baseline MAGIC biomarker risk score (1 vs. 2 vs. 3), pre-enrollment GVHD prophylaxis, and conditioning regimen at transplant. If and only if a categorical variable has missing data for some patient(s), a "Missing" category will be included as well for that variable with its frequency and percentage. Descriptive statistics for the baseline Ann Arbor biomarker risk score and the baseline MAGIC biomarker risk score may be provided in a separate report depending on the biomarker data availability.

Continuous characteristics considered include age at randomization, time from steroid initiation to randomization, time from disease diagnosis to transplant, time from transplant to enrollment, height at transplant, weight at enrollment and body mass index at enrollment. Descriptive statistics will summarize values among non-missing observations.

7.2.2 PARTICIPANT DISPOSITION

The number and percentage of patients experiencing the following events on study will be described:

- Enrollment
- Initiating study treatment (AAT or PTM)
- Initiating next-line therapy (before Day 28, at/after Day 28)
- Completing study treatment up to Day 28
- Completing maintenance study treatment up to Day 56
- Completing planned study follow-up alive (i.e., initiate study treatment and complete 1 year of follow-up alive)
- Completing planned study follow-up (i.e. initiate study treatment and either die within 1 year and did not terminate study prior to the death or complete 1 year of follow-up alive)

In addition, the reasons for patients failing to initiate study treatment, discontinuing study treatment after initiation, and withdrawing from the study will be tabulated.

7.2.3 PROTOCOL DEVIATIONS

The number and percentage of patients in the ITT population with any major protocol deviation will be tabulated by the deviation type. A listing of all protocol deviations for patients in the ITT population will be provided.

7.3 STUDY TREATMENT

Compliance with protocol-prescribed usage of study treatment will be described based on the Safety population. The recommended schedule for study treatment includes infusion at a dose of 120 mg/kg on Day 0, 4, 8, 12, 16, 20, 24, and 28. The patient's weight recorded at enrollment will determine the dose for all infusions.

Responding patients (CR/PR) will continue to receive study treatment at a dose of 120 mg/kg on Day 35, 42, 49, and 56.

All subsequent treatments will have a window of ± 3 days to allow for efficient scheduling of patients for holidays/weekends/etc.

An infusion may be held for the following reasons:

- CTCAE Grade 3+ adverse event that is not attributable to an anticipated post-transplant event
- Severe infection resulting in hemodynamic instability requiring the use of vasopressor medication
- At the treating physician's discretion

Additionally, an infusion should be discontinued immediately if anaphylactic or severe anaphylactoid reactions occur.

Study drug shall be discontinued and not re-instituted if any one of the following criteria is met:

- The patient does not receive study medication for 3 consecutive doses or 4 doses in total for any reason, including toxicity
- The patient receives next line therapy other than Ruxolitinib that is not approved by the Study Chairs
- Relapse or progression of underlying malignancy
- Pregnancy or initiation of breastfeeding during the study
- Withdrawal of consent from study and/or study drug

The number and percentage of patients who received all planned infusions overall and through Day 28 will be tabulated. Additionally, treatment compliance will be summarized overall and through Day 28. The number of patients who receive Ruxolitinib or other approved next-line therapies will be tabulated.

7.4 ANALYSIS OF PRIMARY ENDPOINT – OVERALL RESPONSE AT DAY 28

7.4.1 ANALYSIS STRATEGY

The primary objective of this study is to evaluate whether AAT treatment provides a superior Day 28 overall response (CR+PR) rate in contrast to placebo-to-match (PTM). The overall response rate is suspected to differ by treatment center and Minnesota (MN) risk category (high vs. standard). Because the average number of patients enrolled per center is

expected to be small (136/25 ≈ 5.5), centers will be combined into center-groups for these analyses to ensure the stability of the models. This will be done by including the top 5 enrolling centers each as a separate group and combining all other centers into a sixth catchall group.

A logistic mixed model will be used to evaluate the effect of study treatment on Day 28 overall response while accounting for these factors, with center-group included as a random effect and treatment arm (AAT vs. PTM) and MN risk as fixed effects.

The primary hypothesis is that the Day 28 overall response rate under AAT treatment exceeds this proportion under PTM after adjusting for center and MN risk. No interactions of center or MN risk with treatment assignments are expected. Therefore, the null and alternative hypotheses for this comparison are:

- H_0: Adjusted odds ratio of response for AAT vs. PTM ≤ 1
- H_A: Adjusted odds ratio of response for AAT vs. PTM > 1

This odds ratio will be evaluated in the logistic mixed model using a one-sided, group sequential test with an overall type I error rate controlled at 2.5%. The standardized log odds ratio estimate from this model will be used as the test statistic (Z-statistic) at each stage. Positive values of the Z-statistic correspond to higher estimated odds of overall response for the AAT group compared to PTM.

Sequential testing of the treatment effect of AAT relative to PTM will be performed using a two-stage design with one interim analysis for futility based on a Gamma family error spending function from Hwang et al. 1990 with $\gamma = -8$. The interim futility analysis will occur after 76 patients are evaluable for Day 28 response and the final analysis will happen after the total study population becomes evaluable.

The primary analysis of Day 28 response will use the efficacy and futility boundaries displayed in **Table 15**, with efficacy/futility indicated if the Z-statistic falls above/below the respective boundary values. Note that the futility boundary is non-binding in that the overall type I error rate will not be inflated beyond 2.5% if the trial continues after the futility boundary is reached. The estimand for this primary analysis is specified in **Table 16**. The clinical question of interest is: what is the difference in overall response rate on Day 28 post-randomization between AAT combined with corticosteroids versus PTM combined with corticosteroids in patients with high risk acute GVHD following allogeneic hematopoietic stem cell transplant (HSCT), regardless of discontinuation of study treatment?

Table 15: Group Sequential Design Stopping Thresholds for Primary Endpoint: Primary Analysis Approach

Analysis	Information Fraction	Sample Size	Futility Boundary for Z-Statistic*	Efficacy Boundary for Z-Statistic*
Interim	0.56	76	-0.31	-
Final	1.00	136	-	1.96

* Positive values of the Z-statistic correspond to higher estimated odds of overall response for the AAT arm compared to placebo.

Table 16: Estimand for Primary Endpoint - Day 28 Overall Response

Estimand Attribute	Definition		
Variable	Day 28 Overall Response, refer to Section 6.2 for details		
Treatment	AAT (+corticosteroids) or PTM (+corticosteroids), refer to Section 2.1 for details		
Population	Patients with high risk acute GVHD following allogeneic HSCT		
Intercurrent Events	**Event**	**Strategy**	**Description**
	1. Death before Day 28 visit	Composite strategy	Classified as a non-response
	2. Initiation of additional systemic aGVHD treatment[1] before Day 28 visit	Composite strategy	Classified as a non-response
	3. Discontinuation of study treatment	Treatment policy	No effect on determination of overall response status
Population-level Summary	Proportions and 95% CIs for each arm; adjusted odds ratio from logistic mixed model and its 95% CI		

[1] Includes next-line GVHD therapy and escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more.

Due to a process error that has been corrected, there is a small chance that the treatment assignments of some participants were unblinded to their treatment center staff. A maximum of 14 participants could have been affected. To assess the robustness of the primary analysis results in consideration of this:

1) In the event that the primary analysis reaches its futility boundary at the interim analysis, a sensitivity analysis would be conducted where these 14 participants will be omitted from the ITT analysis population; and

2) At the final analysis, a sensitivity analysis will be conducted where these 14 participants will be omitted from the ITT analysis population.

The sensitivity analyses would include 62 participants at interim and 122 at the final analysis. These analyses would be conducted using the ITT population and a similar estimand as in **Table 16**, with the only modification being the exclusion of the 14 potentially affected patients from the ITT population. The efficacy and futility boundaries for the sensitivity analyses are displayed in **Table 17**.

Table 17: Group Sequential Design Stopping Thresholds for Primary Endpoint: Sensitivity Analysis Approach

Analysis	Information Fraction	Sample Size	Futility Boundary for Z-Statistic*	Efficacy Boundary for Z-Statistic*
Interim	0.51	62	-0.61	-
Final	1.00	122	-	1.96

* Positive values of the Z-statistic correspond to higher estimated odds of overall response for the AAT arm compared to placebo.

Details on the conduct of the primary and sensitivity analyses of the Day 28 response rate and decisions on futility and/or efficacy to be made based on their findings are described in the subsequent sections.

7.4.2 INTERIM FUTILITY ANALYSIS

7.4.2.1 PRIMARY ANALYSIS

The interim futility analysis will be performed after 76 patients become evaluable for Day 28 response. The primary analysis approach will use the ITT population. A logistic mixed model will be used to evaluate the effect of study treatment on Day 28 overall response while accounting for treatment center and MN risk category (high vs. standard), with center-group included as a random effect and treatment arm and MN risk as fixed effects. The standardized log odds ratio estimate for treatment arm from this model will be used as the test statistic (Z-statistic) and will be compared to the futility boundary in **Table 15**.

7.4.2.2 SENSITIVITY ANALYSIS

If the Z-statistic from the primary analysis is less than or equal to the futility boundary, a sensitivity analysis will be conducted that excludes from the ITT population the 14 participants who could have potentially been affected by unblinding. This analysis would use a logistic mixed model with treatment arm, MN risk, and treatment center-group included as covariates. The standardized log odds ratio estimate for treatment arm would be used as the test statistic and compared to the futility boundary in **Table 17**.

7.4.2.3 DECISION CRITERION

If both the primary and sensitivity analysis results meet their corresponding futility boundaries, the trial will be stopped. Otherwise, if one or both of the analyses do not meet their futility boundaries, the trial will continue.

7.4.3 FINAL ANALYSIS

7.4.3.1 PIVOTAL ANALYSIS

The final analysis will be performed after 136 patients become evaluable for Day 28 response. The primary analysis approach will use the ITT population. A logistic mixed model will be used to evaluate the effect of study treatment on Day 28 overall response while accounting for treatment center-group and MN risk category (high vs. standard), with center-group included as a random effect and treatment arm and MN risk as fixed effects. The standardized log odds ratio estimate for treatment arm from this model will be used as the test statistic (Z-statistic) and will be compared to the efficacy boundary in **Table 15**. The proportion of patients with ORR at Day 28 will be described by treatment arm using sample proportions and 95% Wilson score CIs. Adjusted odds ratio estimates from the logistic mixed model of Day 28 overall response will be summarized in a regression table.

7.4.3.2 SENSITIVITY ANALYSIS REGARDING POTENTIAL UNBLINDING INCIDENT (SA1)

A sensitivity analysis will also be conducted that excludes from the ITT population the 14 participants who could have potentially been affected by unblinding. This analysis will use a logistic mixed model with treatment arm, MN risk, and treatment center-group included as covariates. The standardized log odds ratio estimate for treatment arm would be used as the test statistic and compared to the efficacy boundary in **Table 17**. The proportion of patients with ORR at Day 28 will be described by treatment arm using sample proportions and 95% Wilson score CIs. Adjusted odds ratio estimates from the logistic mixed model of Day 28 overall response will be summarized in a regression table.

7.4.3.3 SENSITIVITY ANALYSIS REGARDING JANUARY 2021 STUDY DRUG SUPPLY DISRUPTION (SA2)

Due to a quality control error during the manufacturing process of the study drug, the study drug was recalled, and enrollment was put on hold in January 2021. This hold was lifted in May 2021 with release of version 3.0 of the protocol.

The course of treatment was disrupted for two patients who were still receiving study drug at the time of the drug recall. To evaluate the sensitivity of analysis results to this disruption, we will conduct a sensitivity analysis of the primary endpoint according to the same procedure in Section 7.4.3.1 where any patients whose treatment courses were disrupted by the drug recall are excluded from the ITT analysis population. This analysis would be conducted using the ITT population and a similar estimand as in **Table 16**, with the only modification being the exclusion of the 2 patients affected by the supply disruption from the ITT population.

7.4.3.4 SUPPLEMENTARY ANALYSES

Two supplementary analyses will investigate the potential influence of other variables on the primary analysis' outcome in the ITT population. Two logistic mixed models will be fitted with Day 28 overall response as the outcome. One model will have fixed effects of treatment arm, MN risk, age (under 18 years vs. 18 years or older), and disease type (malignant vs. non-malignant) and a random effect of center-group; the second will also include these covariates in addition to the MAGIC biomarker risk score (1 vs. 2 vs. 3). If no patients under 18 years old are enrolled and randomized, the fixed effect for age will be omitted from the model. The results from the second model may be presented in a separate report depending on the biomarker data availability.

In addition to these protocol-specified supplementary analyses, an additional supplementary analysis will assess Day 28 overall response in comparison to the aGVHD organ staging at initiation of AAT/PTM (i.e., aGVHD organ staging collected at pre-therapy visit if available, and the maximum GVHD organ staging within 72 hours prior to enrollment if not available). The analysis population will include the mITT patients, i.e., those who initiate their assigned study therapy. A logistic mixed model will evaluate the effect of treatment arm on Day 28 overall response with adjustment for treatment center and MN risk as specified in Section 7.4.1.1.

Moreover, the effects of duration from the initiation of corticosteroid to the initiation of IMP (<= 24 hours vs. 24-48 hours vs. 48-72 hours vs. 72 hours or more) and dose of corticosteroid on Day 28 post-randomization (>= 2mg/kg/day vs. 1.00-1.99 mg/kg/day vs. 0.26-1.00 mg/kg/day vs. 0.25 mg/kg/day or less, prednisone-equivalent) on Day 28 overall response will be evaluated using logistic mixed models that adjust for treatment center-group and MN risk. These analyses will use the mITT population.

7.4.3.5 DECISION CRITERION

The findings from the primary analysis alone will be used to judge efficacy of AAT in comparison to PTM for the primary endpoint at the final analysis. If the primary analysis test statistic falls above the efficacy boundary, the study team will conclude that significant evidence of efficacy has been demonstrated; otherwise, the conclusion is that insufficient evidence of efficacy exists.

7.5 ANALYSIS OF SECONDARY ENDPOINTS

7.5.1 DURATION OF RESPONSE (DOR)

DOR will be evaluated in the set of patients in the ITT population who achieve a response (CR/PR) at Day 28. DOR will be estimated using the Kaplan-Meier estimator, with the probability of having a durable response, i.e. retaining the CR/PR, evaluated at 6 and 12 months post-randomization. Estimates and 95% CIs for the median and quartiles of the DOR will be obtained from the Kaplan-Meier estimator using the inverse Kaplan-Meier method described in Andersen et al. 1993. DOR will be compared from Day 28 to 12 months post-randomization between arms using a log rank test. DOR will also be displayed graphically using Kaplan-Meier curves with numbers of subjects at risk at specific time points presented for each treatment group.

Primary and supplementary analyses will be performed. In the primary analysis, DOR is defined as stated in section 6.3.1, as the time from Day 28 response to progression of acute GVHD, any next-line GVHD therapy, escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more, or death from any cause. Its estimand is displayed in **Table 18.** The clinical question of interest is: what is the difference in duration of response between AAT combined with corticosteroids versus PTM combined with corticosteroids in patients with high risk acute GVHD following allogeneic HSCT who attain a Day 28 overall response, regardless of discontinuation of study treatment?

Table 18: Estimand for Duration of Response - Primary Analysis

Estimand Attribute	Definition		
Variable	Duration of Response; refer to Sections 6.3.1 and 7.9, and above for details		
Treatment	AAT (+corticosteroids) or PTM (+corticosteroids), refer to Section 2.1 for details		
Population	Patients with high risk acute GVHD following allogeneic HSCT who attain a Day 28 overall response		
Intercurrent Events	**Event**	**Strategy**	**Description**

Estimand Attribute	Definition		
	1. Acute GVHD progression	Composite strategy	Classified as an event
	2. Death	Composite strategy	Classified as an event
	3. Initiation of additional systemic aGVHD treatment[1]	Composite strategy	Classified as an event
	4. Discontinuation of study treatment	Treatment policy	No effect on variable
Population-level Summary	Median and quartiles with 95% CIs for each arm; probabilities of durable response at 6 and 12 months by arm; log rank test p-value comparing arms from Day 28 to 12 months post-randomization		

[1] Includes next-line GVHD therapy and escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more.

For the supplementary analysis, DOR is defined as the time from Day 28 response to any next-line GVHD therapy, escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more, or death from any cause; the corresponding estimand is shown in **Table 19**. The clinical question of interest is similar to the one for the primary analysis of DOR with the only difference being the definition of DOR.

Table 19: Estimand for Duration of Response - Supplementary Analysis

Estimand Attribute	Definition		
Variable	Duration of Response, refer to Sections 6.3.1 and 7.9, and above for details		
Treatment	AAT (+corticosteroids) or PTM (+corticosteroids), refer to Section 2.1 for details		
Population	Patients in Intention-to-treat Population who attain a Day 28 overall response		
Intercurrent Events	**Event**	**Strategy**	**Description**
	1. Acute GVHD progression	Treatment policy	No effect on variable
	2. Death	Composite strategy	Classified as an event
	3. Initiation of additional systemic aGVHD treatment[1]	Composite strategy	Classified as an event
	4. Discontinuation of study treatment	Treatment policy	No effect on variable
Population-level Summary	Median and quartiles with 95% CIs for each arm; probabilities of durable response at 6 and 12 months by arm; log rank test p-value comparing arms from Day 28 to 12 months post-randomization		

[1] Includes next-line GVHD therapy and escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more.

7.5.2 NON-RELAPSE MORTALITY (NRM)

The time from randomization until NRM will be described in the ITT population graphically for each treatment arm using the Aalen-Johansen estimator (Aalen and Johansen 1978), with numbers of subjects at risk at specific time points presented for each treatment group

and relapse/progression of the primary disease treated as a competing risk. Estimates of the cumulative incidence of NRM will be provided at 6 and 12 months post-randomization along with 95% CIs computed using the complementary log-log transformation. Gray's test will be used to compare the cumulative incidence of NRM within 12 months post-randomization between arms as the primary analysis. The estimand for NRM is presented in **Table 20**. The clinical question of interest is: what is the difference in cumulative incidence of NRM between AAT combined with corticosteroids versus PTM combined with corticosteroids in patients with high risk acute GVHD following allogeneic HSCT, regardless of discontinuation of study treatment?

Table 20: Estimand for Non-relapse Mortality

Estimand Attribute	Definition		
Variable	Non-relapse Mortality, refer to Sections 6.3.2 and 7.9 for details		
Treatment	AAT(+corticosteroids) or PTM (+corticosteroids), refer to Section 2.1 for details		
Population	Patients with high risk acute GVHD following allogeneic HSCT		
Intercurrent Events	**Event**	**Strategy**	**Description**
	1. Death	Composite strategy	Classified as an event
	2. Relapse / progression	Composite strategy	Classified as a non-event (competing risk)
	3. Discontinuation of study treatment	Treatment policy	No effect on variable
Population-level Summary	Cumulative incidence of NRM at 6 and 12 months with 95% CIs for each arm; Gray's test p-value comparing arms from randomization to 12 months post-randomization		

In a supplementary analysis, a Cox proportional hazards model with mixed effects will be used to compare the cause-specific hazards of NRM between treatment arms while adjusting for an MN risk fixed effect and center-group lognormal frailties. The Cox regression model will be used to evaluate the adjusted hazard ratio of treatment (AAT vs. PTM) using a point estimate and a 95% CI.

7.5.3 OVERALL SURVIVAL (OS)

The time from randomization until death from any cause will be described in the ITT population graphically for each treatment arm using the Kaplan-Meier estimator, with numbers of subjects at risk at specific time points presented for each treatment group.

Estimates of OS at 6 and 12 months post-randomization will be provided along with 95% CIs computed using the complementary log-log transformation as detailed in Kalbfleisch and Prentice 1980. For the primary analysis, the log-rank test will be used to compare OS within 12 months of randomization between arms as the primary analysis. The estimand for OS is shown in **Table 21**. The clinical question of interest is: what is the difference in overall survival between AAT combined with corticosteroids versus PTM combined with corticosteroids in patients with high risk acute GVHD following allogeneic HSCT, regardless of discontinuation of study treatment?

Table 21: Estimand for Overall Survival

Estimand Attribute	Definition		
Variable	Overall Survival, refer to Sections 6.3.3 and 7.9 for details		
Treatment	AAT (+corticosteroids) or PTM (+corticosteroids), refer to Section 2.1 for details		
Population	Patients with high risk acute GVHD following allogeneic HSCT		
Intercurrent Events	**Event**	**Strategy**	**Description**
	1. Death	Composite strategy	Classified as an event
	2. Discontinuation of study treatment	Treatment policy	No effect on variable
Population-level Summary	OS probabilities at 6 and 12 months with 95% CIs for each arm; log rank test p-value comparing arms from randomization to 12 months post-randomization		

In a supplementary analysis, a Cox proportional hazards model with mixed effects will be used to compare OS between treatment arms while adjusting for an MN risk fixed effect and center-group lognormal frailties. The Cox regression model will be used to evaluate the adjusted hazard ratio of treatment (AAT vs. PTM) using a point estimate and a 95% CI.

7.5.4 PROGRESSION-FREE SURVIVAL (PFS)

PFS is defined as being alive and free of relapse/progression of the primary disease. The time from randomization until PFS failure will be described in the ITT population graphically for each treatment arm using the Kaplan-Meier estimator, with numbers of subjects at risk at specific time points presented for each treatment group. Estimates of PFS at 6 and 12 months post-randomization will be provided along with 95% CIs computed using the complementary log-log transformation as detailed in Kalbfleisch and Prentice 1980. For the primary analysis, the log-rank test will be used to compare PFS within 12 months post-randomization between arms as the primary analysis. **Table 22** shows the estimand for PFS. The clinical question of interest is: what is the difference in PFS between AAT combined with corticosteroids versus PTM combined with corticosteroids in patients with high risk acute GVHD following allogeneic HSCT, regardless of discontinuation of study treatment?

Table 22: Estimand for Progression-free Survival

Estimand Attribute	Definition		
Variable	Progression-free Survival, refer to Sections 6.3.4 and 7.9 for details		
Treatment	AAT (+corticosteroids) or PTM (+corticosteroids), refer to Section 2.1 for details		
Population	Patients with high risk acute GVHD following allogeneic HSCT		
Intercurrent Events	**Event**	**Strategy**	**Description**
	1. Death	Composite strategy	Classified as an event
	2. Relapse / progression	Composite strategy	Classified as an event
	3. Discontinuation of study treatment	Treatment policy	No effect on variable

Estimand Attribute	Definition
Population-level Summary	PFS probabilities at 6 and 12 months with CIs for each arm; log rank test p-value comparing arms from randomization to 12 months post-randomization

In a supplementary analysis, a Cox proportional hazards model with mixed effects will be used to compare PFS between treatment arms while adjusting for an MN risk fixed effect and center-group lognormal frailties. The Cox regression model will be used to evaluate the adjusted hazard ratio of treatment (AAT vs. PTM) using a point estimate and a 95% CI.

7.5.5 GVHD-FREE SURVIVAL

GVHD-free survival is defined as being alive, free of active acute or chronic GVHD, free of any next-line GVHD therapy, and free of escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more. The GVHD-free survival at Day 56 will be computed as a binary outcome. The proportion of patients in the ITT population with GVHD-free survival at Day 56 will be evaluated by the treatment arm. These proportions will be described by sample proportions and 95% Wilson score CIs. A logistic mixed model will be used to evaluate the effect of study treatment on Day 56 GVHD-free survival while accounting for center-group and MN risk, with center-group included as a random effect and treatment arm and MN risk as fixed effects. **Table 23** contains the estimand for GVHD-free survival. The clinical question of interest is: what is the difference in Day 56 GVHD-free survival between AAT combined with corticosteroids versus PTM combined with corticosteroids in patients with high risk acute GVHD following allogeneic HSCT, regardless of discontinuation of study treatment?

Table 23: Estimand for GVHD-free Survival

Estimand Attribute	Definition		
Variable	Day 56 GVHD-free Survival, refer to Sections 6.3.5 and 7.9 for details		
Treatment	AAT(+corticosteroids) or PTM (+corticosteroids), refer to Section 2.1 for details		
Population	Patients with high risk acute GVHD following allogeneic HSCT		
Intercurrent Events	**Event**	**Strategy**	**Description**
	1. Death before Day 56	Composite strategy	Classified as a failure
	2. Development of aGVHD or cGVHD before Day 56	Composite strategy	Classified as a failure
	2. Initiation of additional systemic aGVHD treatment[1] before Day 56	Composite strategy	Classified as a failure
	3. Discontinuation of study therapy	Treatment policy	No effect on variable
Population-level Summary	Proportion of patients with GVHD-free at Day 56 with 95% CIs for each arm; adjusted odds ratio from logistic mixed model with 95% CI and p value		

[1] Includes next-line GVHD therapy and escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more.

7.5.6 ACUTE GVHD RESPONSE TO AAT/PTM

The proportions of patients with CR, PR, and TF will be tabulated by treatment arm at Days 7, 14, 21, 28, and 56 for all patients in the ITT population and at Day 86 for the subset of patients that remained on study maintenance treatment after Day 28. Goodman simultaneous 95% CIs (Goodman 1965) will be provided for the response category proportions. These proportions will be compared between treatment arms using Fisher's exact test. If a patient does not complete the acute GVHD assessment on an evaluation day for any reason (e.g. missed visit, withdrawal from the study, loss to follow-up), they will be assigned as a TF on that Day. Moreover, patients who die, initiate either approved or unapproved next-line GVHD therapy, or have escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more prior to an acute GVHD assessment will be classified as non-responders (NR) at the corresponding Day.

7.5.7 ACUTE GVHD RESPONSE ALLOWING FOR APPROVED NEXT-LINE THERAPY

The proportions of patients with CR, PR, and TF will be tabulated by treatment arm at Days 7, 14, 21, 28, and 56 for all patients in the ITT population and at Day 86 for the subset of patients that remained on study maintenance treatment after Day 28. Goodman simultaneous 95% CIs (Goodman 1965) will be provided for the response category proportions. These proportions will be compared between treatment arms using Fisher's exact test. If a patient does not complete the acute GVHD assessment on an evaluation day for any reason (e.g. missed visit, withdrawal from the study, loss to follow-up), they will be assigned as a TF on that Day. Moreover, patients who die, *initiate next-line GVHD therapy that is not approved by the protocol chairs*, or have escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more prior to an acute GVHD assessment will be classified as non-responders (NR) at the corresponding Day.

The overall response (CR+PR) rate at Day 28 allowing for approved next-line treatments will be compared between treatment arms. A logistic mixed model will be used to evaluate the effect of study treatment on Day 28 overall response with adjustment for treatment center-group and MN risk category (high vs. standard), with center-group (defined in section 7.1) included as a random effect and treatment arm and MN risk as fixed effects. This analysis will use the ITT population. A two-sided Wald test of the treatment arm odds ratio in the logistic mixed model will be performed at a significance level of 5%. The standardized log odds ratio estimate from this model will be used as the test statistic (Z-statistic). The proportion of patients with an overall response at Day 28 will be described by treatment arm by sample proportions and 95% Wilson score CIs. Adjusted odds ratio estimates from the logistic mixed model of Day 28 overall response will be summarized in a regression table.

7.5.8 SYSTEMIC INFECTIONS

The frequency of Grade 2-3 infections occurring from randomization until 30 days post-last dose of AAT/PTM will be tabulated in the safety population by treatment arm, disease site, date of onset, and severity, with Grade defined per the BMT CTN Technical MOP.

The time from randomization until the first Grade 2-3 infection will be described graphically using the Aalen-Johansen estimator, with death before infection treated as a competing risk. Estimates of the cumulative incidence of Grade 2-3 infection will be provided at 3 months post-randomization along with 95% CIs computed using the complementary log-log transformation. Gray's test will be used to compare the cumulative incidence of Grade 2-3 infection within 3 months of randomization between arms.

7.5.9 ADVERSE EVENTS (AES)

The frequency of Grade 3-5 TEAEs per CTCAE version 5.0 occurring from randomization until 30 days post-last dose will be tabulated by the organ system for each treatment arm in the safety population.

The proportions of patients experiencing a Grade 3-5 TEAE will be described by arm using 95% Wilson score CIs and compared between treatment arms using Barnard's exact test.

7.5.10 CHRONIC GVHD

The time from randomization until the onset of chronic GVHD of any severity (mild, moderate, or severe per 2014 NIH Consensus Criteria) will be described in the ITT population graphically using the Aalen-Johansen estimator, with death prior to chronic GVHD onset treated as a competing risk. Estimates of the cumulative incidence of chronic GVHD will be provided at 6 and 12 months post-randomization along with 95% CIs computed using the complementary log-log transformation. Gray's test will be used to compare the cumulative incidence of chronic GVHD within 12 months of randomization between arms as the primary analysis.

In a supplementary analysis, a Cox proportional hazards model with mixed effects will be used to compare the cause-specific hazards of chronic GVHD between treatment arms while adjusting for an MN risk fixed effect and center-group lognormal frailties. The Cox regression model will be used to evaluate the adjusted hazard ratio of treatment (AAT vs. PTM) using a point estimate and a 95% CI.

The maximum severity of chronic GVHD through 12 months post-randomization will be tabulated by treatment arm.

7.5.11 RELAPSE/PROGRESSION OF PRIMARY DISEASE

The time from randomization until relapse/progression of the primary disease will be described in the ITT population graphically for each treatment arm using the Aalen-Johansen estimator, with death prior to relapse/progression treated as a competing risk. Estimates of the cumulative incidence of relapse/progression will be provided at 6 and 12 months post-randomization along with 95% CIs computed using the complementary log-

log transformation. Gray's test will be used to compare the cumulative incidence of relapse/progression within 12 months of randomization between arms as the primary analysis.

In a supplementary analysis, a Cox proportional hazards model with mixed effects will be used to compare the cause-specific hazards of relapse/progression between treatment arms while adjusting for an MN risk fixed effect and center-group lognormal frailties. The Cox regression model will be used to evaluate the adjusted hazard ratio of treatment (AAT vs. PTM) using a point estimate and a 95% CI.

7.6 ANALYSIS OF EXPLORATORY ENDPOINTS

7.6.1 SERUM AAT LEVELS

The pharmacokinetic parameters maximum concentration (Cmax) and trough concentration (Ctrough) for serum antigen AAT concentrations and serum functional AAT concentrations (AAT concentrations with functional activity) will be reported and summarized in the PK population at Days 0, 16, 28, and 56 post-treatment initiation by treatment arm using the n, mean, standard deviation (SD), coefficient of variation (CV%), median, range, geometric mean, and geometric CV%. Ctrough will also be summarized on Days 8 and 24.

7.6.2 STOOL AAT LEVELS

The stool concentration of AAT may be summarized at baseline and at Days 8 and 28 post-treatment initiation by treatment arm using the mean, median, quartiles, and range. The detailed analysis of this endpoint may be described in a separate document when data become available.

7.6.3 IMMUNE CELL SUBSETS

Peripheral blood ratios of T regulatory to T effector (Treg/Teff) cells and Natural Killer (NK) cell populations may be described by treatment arm using the sample mean, median, range, and quartiles at baseline and Days 16, 28, and 56 post-treatment initiation in tables of summary statistics and displayed graphically using box plots. Changes in levels from baseline to each subsequent assessment time may be described similarly in summary tables and compared between treatment arms using Wilcoxon rank sum tests. The detailed analysis of this endpoint may be described in a separate document when data become available.

7.6.4 INFLAMMATORY CYTOKINES AND GVHD BIOMARKERS

Serum levels of IL-1β, IL-6, IL-10, TNFα, amphiregulin (AREG), heparin sulfate, ST2, and REG3α may be described by treatment arm using the sample mean, median, range, and quartiles baseline and at Days 8, 28, and 56 post-treatment initiation in tables of summary statistics and displayed graphically using box plots. Changes in levels from baseline to each subsequent assessment time may be described similarly in summary tables and compared between treatment arms using Wilcoxon rank sum tests. The detailed analysis of this endpoint may be described in a separate document when data become available.

7.6.5 OVERALL AND ORGAN-SPECIFIC RESPONSE RATES BY GVHD AND BIOMARKERS BASED RISK CLASSIFICATIONS

The proportions of patients with an overall response (complete or partial) and the response rate of each GVHD target organ at Day 28 will be described in the ITT population by MN risk (high vs. standard) and by Ann Arbor biomarker risk score (1, 2, or 3) using sample proportions and 95% Wilson score CIs and compared on each factor using Fisher's exact test.

In addition, a subgroup analysis of overall response by treatment arm will be conducted. Response rates for each arm within levels of MN risk and Ann Arbor biomarker score will be described using sample proportions and 95% Clopper-Pearson CIs. Potential differential treatment effects between subgroups will be evaluated using logistic regression models with treatment arm, risk classification, and an interaction of the two included as effects. Wald test p-values of the interaction terms will be reported to assess the presence of differential effects across subgroups. The results from the analysis by Ann Arbor biomarker risk score may be presented in a separate report depending on the biomarker data availability.

7.6.6 SYSTEMIC CORTICOSTEROID DOSE

The prednisone-equivalent corticosteroid dose being administered (measured in prednisone equivalent) will be described at baseline, at Days 7, 14, 21, 28, 56, and 86 and at 6 and 12 months post-randomization by treatment arm in the ITT population using the sample mean, median, range, and quartiles in a table of summary statistics and displayed graphically using box plots. Changes in levels from baseline to each subsequent assessment time will be described similarly in summary tables and compared between treatment arms using Wilcoxon rank sum tests.

7.6.7 CMV REACTIVATION

Among patients who were CMV positive at transplant, the cumulative incidence of initiation of systemic treatment for CMV reactivation will be described in the safety population graphically using the Aalen-Johansen estimator, with death treated as a competing risk. Estimates of the cumulative incidence of CMV reactivation will be provided at Day 56 for each arm along with 95% CIs computed using the complementary log-log transformation. Gray’s test will be used to compare the cumulative incidence of CMV reactivation within 56 days post-randomization between arms.

7.6.8 PATIENT REPORTED OUTCOMES

Patient self-reported measures will be assessed at baseline, Days 28 and 56, and 6 months post-randomization using the MD Anderson Symptom Inventory (MDASI) and selected PROMIS subscales. The core symptom severity and mean interference subscales of the MDASI and the PROMIS domains will be scored according to the recommendations of the developers.

For each subscale/domain considered, scores at baseline, Days 28 and 56, and 6 months post-randomization will be described in the ITT population by treatment arm using the

sample mean, median, range, and quartiles in tables of summary statistics and displayed graphically using box plots. Linear regression models will be used to compare trajectories in scores over time between treatment arms while adjusting for baseline score; model effects will include treatment assignment, baseline score, and post-randomization assessment time (Day 28, Day 56, or 6 months).

Provided that at least 10 adolescent participants are enrolled, separate analyses of patient reported outcomes will be conducted for adolescents and adults. Otherwise, their data will be aggregated for one common analysis.

7.7 INTERIM ANALYSES

7.7.1 FUTILITY ANALYSIS OF PRIMARY ENDPOINT

This trial includes one formal interim analysis for futility, to occur once 76 patients are evaluable for the primary endpoint. The existence of a futility signal will be determined as specified in Section 7.4.1. Note that this is non-binding futility stopping rule, in that the type I error rate will not exceed 2.5% if the DSMB chooses not to stop the trial in the presence of a futility signal.

7.7.2 ANALYSIS FOR 6-MONTH CLINICAL STUDY REPORT

To expedite the evaluation of the primary endpoint and to support a potential regulatory submission, a clinical study report will be drafted once all enrolled participants complete 6 months of follow-up in the study. This report will include efficacy, safety, and PK data collected through the data cut-off date. Analysis included in this report will consist of descriptive statistics on baseline characteristics and outcomes as well as formal treatment comparison for the primary endpoint Day 28 overall response via hypothesis testing. The p value from the test of superiority of AAT to PTM will be calculated via a logistic mixed model for GVHD-free survival at Day 56. This p value is considered descriptive in the 6-month report and will not be formally evaluated for efficacy of AAT until all the testing entailed in the hierarchical testing procedure is completed at the end of the study. The analysis will be performed using the statistical methods detailed in Sections 7.4-7.6.

A detailed account of which TLFs will be included in the 6-month clinical study report is provided in the supplementary TLF shell document for this SAP.

7.8 SAFETY ANALYSIS

7.8.1 DATA AND SAFETY MONITORING BOARD

The policies and composition of the Data and Safety Monitoring Board (DSMB) are described in the BMT CTN MOP. The DSMB performs a review of key efficacy and safety data at least once a year at regularly scheduled meetings. Additional reviews may occur on an ad-hoc basis by request of the DSMB. The content of the data reports provided for DSMB reviews is not specified by this SAP.

7.8.2 STATISTICAL MONITORING RULES

The rate of non-relapse mortality occurring by Day 56 will be monitored monthly for each treatment arm. If either rate significantly exceeds pre-set thresholds, the NHLBI will be notified so that the DSMB can be advised. These stopping guidelines serve as triggers for consultation with the DSMB for guidance and are not formal “stopping rules” that would mandate automatic closure of study enrollment. The monitoring scheme is detailed in section 5.3.4 of the study protocol.

7.8.3 SAFETY EVENT SUMMARIES FOR ANALYSIS REPORTS

All safety analyses will be conducted using the safety population. All AEs and SAEs reported on the study will be categorized by MedDRA preferred terms using MedDRA version 23.0 and higher. The number of each type of event and number of patients affected will be described during two study periods: from randomization until 30 days following the last dose of study drug (AAT/PTM), and from 30 days following the last dose until 12 months post-randomization. These events will be reported using the safety analysis population.

In addition, AEs and SAEs will be described in the same manner for the set of patients who received Ruxolitinib or other approved next-line therapy and those who did not. Causes of death for expired participants will be tabulated.

7.9 MISSING DATA AND SENSITIVITY ANALYSIS

Regarding endpoints involving acute GVHD response to study treatment (overall response, DOR, GVHD-free survival, acute GVHD response, organ-specific response), any patient with a missing acute GVHD assessment at the evaluation day will be classified as a non-responder (NR) at that time point.

For the duration of response (DOR), a patient with a missing acute GVHD assessment will be considered as a failure at that time point.

For GVHD-free survival, a patient with either a missing acute or chronic GVHD assessment will be considered as a failure at that time point.

For time-to-event outcomes other than DOR and GVHD-free survival (NRM, OS, PFS, chronic GVHD, relapse/progression), patients that withdraw or are lost to follow-up are assumed to be censored at random.

For other endpoints, the occurrence of missing data, whether due to the patient’s missing assessment(s) or withdrawal from the study, is assumed to occur randomly.

7.10 ASSESSMENT OF IMPACT OF COVID-19 ON TRIAL RESULTS

7.10.1 COVID-19 RELATED MODIFICATIONS TO DATA COLLECTION

The BMT CTN considers the COVID-19 era to have potentially impacted its trials effective March 10, 2020 unless otherwise specified. Data collection was adjusted, and guidance was given to centers to track the impact of COVID-19. Specifically, the following modifications were made:

- Impact of COVID-19 on scheduled visits: Collection of information about whether visits were missed or scheduled out of the assessment window; the type of visit (clinic visit, telemedicine visit, phone contact, or combination); and if the visit was missed, what the reason for the missed visit was (which included COVID-19 as an option)
- COVID-19 infections: Guidance on the collection of COVID-19 infection occurrences was incorporated into the infection data collection form
- COVID-19 related deaths: Addition of COVID-19 as a potential primary cause of death

7.10.2 COVID-19 RELATED EVENTS AND VISIT DISRUPTIONS

Frequencies and percentages will be provided on the number of missed visits, out of window visits, type of visits, and reasons for missed visits, summarized overall and by treatment arm. The number of deaths attributed to COVID-19 and the number of COVID-19 infections will also be described by treatment arm.

7.10.3 SENSITIVITY ANALYSIS OF THE PRIMARY ENDPOINT

The primary endpoint of this study is overall response (CR+PR) at Day 28 post-randomization. To assess the impact of COVID-19 on the trial's pivotal analysis, the following COVID-19 sensitivity analyses (CA) are proposed.

- CA1: Exclude patients from the analysis who have missing Day 28 GVHD assessments attributed to a COVID-19 related event or for those who have a telemedicine or phone contact visit only
- CA2: Exclude patients from the analysis who had a COVID-19 infection or death attributed to COVID-19 before or on Day 28
- CA3: Exclude all patients who are excluded from analysis in CA1 and/or CA2

For each sensitivity analysis, the primary endpoint will be analyzed as described in Section 7.4.3.1 according to the ITT principle. The purpose of the sensitivity analyses is to assess the robustness of the pivotal analysis results. The impact of sensitivity analysis populations on the number of responses and widths of the confidence intervals should be considered in interpreting these analyses.

If thought to be important based on the results, similar sensitivity analyses of secondary endpoints may also be conducted.

7.11 CHANGES TO PROTOCOL-SPECIFIED ANALYSIS

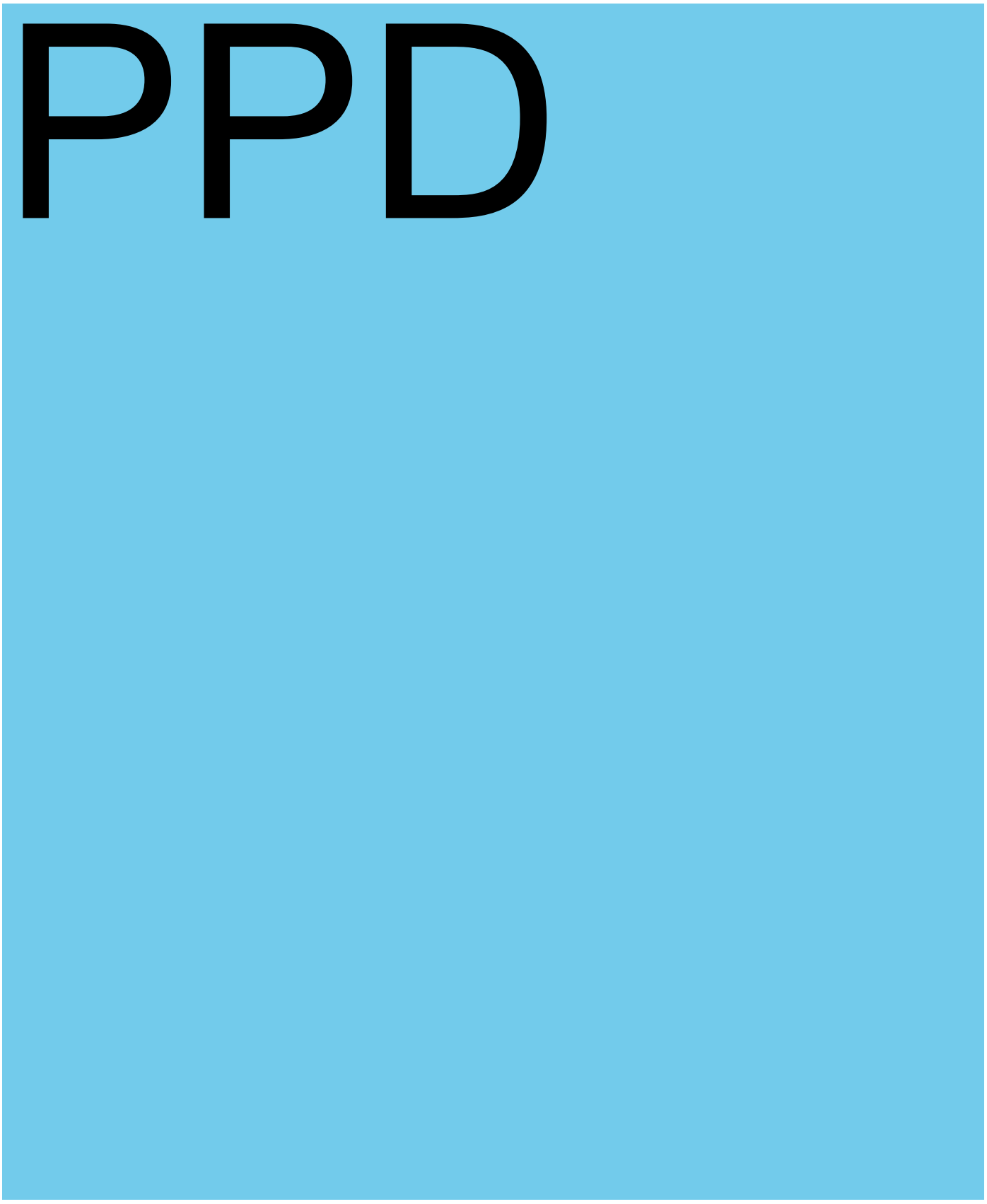
The current SAP elaborates on the protocol-specified analysis. The sole change in the SAP from the protocol-specified analysis is that the significance level of the two-sided testing of exploratory endpoints is planned to change from 5% to 1%. Any future revisions to the SAP that constitute changes from the protocol-specified analysis, and the justification for these changes, will be documented here.

8 REFERENCES

1. *Aalen, O. O. and S. Johansen (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scand J Stat 141-150.*
2. *Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). Statistical Models Based on Counting Processes. Springer Ser. Statist., Springer.*
3. *Goodman, L. A. (1965). On Simultaneous Confidence Intervals for Multinomial Proportions. Technometrics 7(2): 247-254.*
4. *Harris, A. C., R. Young, S. Devine, W. J. Hogan, F. Ayuk, U. Bunworasate, C. Chanswangphuwana, Y. A. Efebera, E. Holler, M. Litzow, R. Ordemann, M. Qayed, A. S. Renteria, R. Reshef, M. Wolfl, Y. B. Chen, S. Goldstein, M. Jagasia, F. Locatelli, S. Mielke, D. Porter, T. Schechter, Z. Shekhovtsova, J. L. Ferrara and J. E. Levine (2016). "International, Multicenter Standardization of Acute Graft-versus-Host Disease Clinical Data Collection: A Report from the Mount Sinai Acute GVHD International Consortium." Biol Blood Marrow Transplant 22(1): 4-10.*
5. *Jagasia, M. H., H. T. Greinix, M. Arora, K. M. Williams, D. Wolff, E. W. Cowen, J. Palmer, D. Weisdorf, N. S. Treister, G. S. Cheng, H. Kerr, P. Stratton, R. F. Duarte, G. B. McDonald, Y. Inamoto, A. Vigorito, S. Arai, M. B. Datiles, D. Jacobsohn, T. Heller, C. L. Kitko, S. A. Mitchell, P. J. Martin, H. Shulman, R. S. Wu, C. S. Cutler, G. B. Vogelsang, S. J. Lee, S. Z. Pavletic and M. E. Flowers (2015). National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group report. Biol Blood Marrow Transplant 21(3): 389-401 e381.*
6. *Kahan, B. C., & Morris, T. P. (2012). Improper analysis of trials randomised using stratified blocks or minimisation. Stat Med, 31(4), 328-340.*
7. *Kalbfleisch, J. D. and Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons.*
8. *Hwang, I. K., Shih W. J., and De Cani J. S. (1990). Group sequential designs using a family of type I error probability spending functions. Stat Med, 9(12), 1439-1445.*
9. *Levine, J. E., Braun T. M., Harris A. C., Holler E., Taylor A., Miller H., Magenau J., Weisdorf D. J., Ho V. T., Bolanos-Meade J., Alousi A. M., and Ferrara J. L. M. (2015) A prognostic score for acute graft-versus-host disease based on biomarkers: a multicentre study. Lancet Haematology, 2(1):e21-29.*
10. *MacMillan, M. L., Robin, M., Harris, A. C., DeFor, T. E., Martin, P. J., Alousi, A., Ho, V.T., Bolaños-Meade, J., Ferrara, J. L., Jones, R. and Arora, M. (2015). A refined risk score for acute graft-versus-host disease that predicts response to initial therapy, survival, and transplant-related mortality. Biol Blood Marrow Transplant 21(4), 761-767.*
11. *Srinagesh, H. K., Özbek, U., Kapoor, U., Ayuk, F., Aziz, M., Ben-David, K., Choe, H. K., DeFilipp, Z., Etra, A., Grupp, S. A., Hartwell, M. J., Hexner, E. O., Hogan, W. J., Karol, A. B., Kasikis, S., Kitko, C. L., Kowalyk, S., Lin, J.-Y., Major-Monfried, H., Mielke, S., Merli, P., Morales, G., Ordemann, R., Pulsipher, M. A., Qayed, M., Reddy, P., Reshef, R., Rosler, W., Sandhu, K. S., Schechter, T., Shah, J., Sigel, K., Weber, D., Wolfl, M., Wudhikarn, K., Young, R., Levine, J. E., and Ferrara, J. L. M. (2019) The MAGIC*

algorithm probability is a validated response biomarker of treatment of acute graft-versus-host disease. Blood Advances 3(23):4034-42.

12. *Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22: 209-212.*


PPD

APPENDIX B: INTERIM ANALYSIS FOR FUTILITY: DSMB DATA PRESENTATION PLAN

Protocol and Primary Endpoint

The primary endpoint for the protocol BMT CTN 1705 "A Randomized, Double-Blind, Placebo-Controlled Multicenter Phase III Trial of Alpha 1 – Antitrypsin (AAT) Combined with Corticosteroids vs Corticosteroids Alone for the Treatment of High-Risk Acute Graft-versus-Host Disease (GVHD) Following Allogeneic Hematopoietic Stem Cell Transplant" is overall response at Day 28. Overall response at Day 28 is defined as having a complete or partial response (CR or PR) at Day 28, i.e. having partial or complete resolution of acute GVHD symptoms in some organ(s) without worsening in any other(s) along with being alive, free of any next-line GVHD therapy, and free of escalation of prednisone-equivalent steroid dose to 2.5mg/kg/day or more. Acute GVHD organ staging at Day 28 is compared to maximum GVHD organ staging within 72 hours prior to enrollment to determine whether a complete or partial resolution has occurred. Organ staging is performed according to MAGIC criteria (Harris et al. 2016). Patients with a missing Day 28 acute GVHD assessment will be classified as non-responders (NR).

Study Design on Interim Analysis

This trial includes one formal interim analysis for futility, to occur once 76 patients are evaluable for the primary endpoint. The primary analysis will be done using the intention-to-treat (ITT) population. The intention-to-treat (ITT) population consists of all randomized subjects, classified according to their randomized treatment assignments.

A logistic mixed model will be used to evaluate the effect of study treatment on Day 28 overall response while accounting for treatment center and MN risk category (high vs. standard), with center-group included as a random effect and treatment arm and MN risk as fixed effects. The top 5 enrolling centers each will count as a separate group and all other centers will be combined into a sixth catchall group. So there will be a total of 6 center groups.

The standardized log odds ratio estimate for treatment arm from this model will be used as the test statistic (Z-statistic) and will be compared to the futility boundary in Table 1. The futility boundaries were obtained from the Gamma family spending function with shape parameters of -8.

Futility Interim Analysis Stopping Boundary

Table 1: Group Sequential Design Stopping Thresholds for Primary Endpoint: Primary Analysis Approach

Analysis	Information Fraction	Sample Size	Futility Boundary for Z-Statistic	Efficacy Boundary for Z-Statistic
Interim	0.56	76	-0.31	-
Final	1.00	136	-	1.96

Sensitivity Analysis for Interim Analysis

A sensitivity analysis will be conducted ONLY IN THE EVENT that the primary analysis reaches its futility boundary at the interim analysis as above mentioned in Table 1.

The sensitivity analyses would include 62 participants at interim analysis. Due to a process error that has been corrected, there is a small chance that the treatment assignments of some participants were unblinded to their treatment center staff. A maximum of 14 participants could have been affected. To assess the robustness of the primary analysis results in consideration of this, a sensitivity analysis will be conducted where these 14 participants will be omitted from the ITT analysis population. Therefore, the sensitivity analysis will include 62 subjects (76 randomized – 14 randomized subjects potentially affected)

The efficacy and futility boundaries for the sensitivity analyses are displayed in **Table 2**. Again, the futility boundaries were obtained from the Gamma family spending function with shape parameters of -8.

Table 2: Group Sequential Design Stopping Thresholds for Primary Endpoint: Sensitivity Analysis Approach

Analysis	Information Fraction	Sample Size	Futility Boundary for Z-Statistic	Efficacy Boundary for Z-Statistic
Interim	0.51	62	-0.61	-
Final	1.00	122	-	1.96

If the Z-statistic from the primary analysis is less than or equal to the futility boundary, a sensitivity analysis will be conducted that excludes from the ITT population the 14 participants who could have potentially been affected by unblinding. Similarly, this analysis would use a logistic mixed model and the standardized log odds ratio estimate for treatment arm would be used as the test statistic and compared to the futility boundary in Table 2.

Results and Data Presentation to DSMB

Primary results for the interim analysis will be presented as below:

Analysis	Information Fraction	Sample Size (#Patient Included)	Test Statistic	Reach Futility Boundary for Z-Statistic (Yes or No)
Interim	0.56	76		

Sensitivity analysis results for the interim analysis will be presented as below:

Analysis	Information Fraction	Sample Size (#Patient Included)	Test Statistic	Reach Futility Boundary for Z-Statistic (Yes or No)
Interim	0.51	62		

If both the primary and sensitivity analysis results meet their corresponding futility boundaries, the study team recommends closure of the trial. Otherwise, if one or both of the analyses do not meet their futility boundaries, the team recommends that the trial continue.

Besides the interim analysis results, below exhibits will also be included in the DSMB report to facilitate the review, and these are usually included in the regular DSMB closed data review. All data will be presented in **Closed** session by the Unblinded STAT lead.

Exhibit R1: Participant disposition table

Numbers of patients that were randomized, received study treatment, completed study treatment, stopped study treatment early, withdrew by treatment arm I and II

Exhibit R2: Overall response

(A) Overall response by Day 56 in Arm I

(B) Overall response by Day 56 in Arm II

Overall (complete + partial) response summarized at Days 14, 28, and 56

Timing of Interim Analysis

From operation perspective, DCC is monitoring the data submission for the enrolled patients to ensure we have 76 randomized patients with Day 28 evaluation for the primary endpoint for this interim analysis. A nightly monitoring is put in place and will send an automatic email alert to the STAT team when the time is approaching to see the data submission for the primary endpoint for the first 76 patients. Patient who would expire prior to Day 28 will be included in the ITT population for the primary endpoint. DCC will plan a data freeze for DSMB review when 76 patients reach 28 days follow up.

Signature Page

CSL964_5001 - Statistical Analysis Plan - 30 November 2023

Signed By	Date (GMT)
PPD	08-Feb-2024 23:33:39
ApprovedPPD	

--- Do Not Detach This Signature Page ---