STATISTICAL ANALYSIS PLAN

Protocol 133646-2
Pomalidomide for the Treatment of Bleeding in Hereditary Hemorrhagic Telangiectasia (NCT03910244)

SAP VERSION: 1.1

SAP DATE: June 19, 2023 – Complete initial version
June 26-28, 2023 – TOC and formatting revisions
July 20, 2023 – updated information in concomitant medications
September 28, 2023 – minor updates after blinded data review meeting

PI/ IND SPONSOR: Keith McCrae, MD
Cleveland Clinic

IND number: 133646

FUNDER: NIH/NHLBI

PREPARED BY: RTI International.
3040 Cornwallis Rd
Research Triangle Park, NC 27709-2104

AUTHOR (S): Benjamin Carper, MS
Derek Marsh, MS
Sonia Thomas, DrPH

CONTENTS

LIST OF ABBREVIATIONS

Abbreviation:	Deciphered:
AE	Adverse Event
AVM	arteriovenous malformations
CRF	Case Report Form
CI	Confidence interval
CTCAE	Common Terminology Criteria for Adverse Events
DSMB	Data Safety and Monitoring Board
EOS	End of study visit, scheduled to be 4-weeks post-treatment
ESS	Epistaxis Severity Score
HCT	Hematocrit
HHT	Hereditary Hemorrhagic Telangiectasia
MedDRA	Medical Dictionary for Regulatory Activities
Neuro-QoL	Quality of Life in Neurological Disorders
NHLBI	National Heart, Lung, and Blood Institute
NIH	National Institutes of Health
PI	Principal Investigator
PROMIS	Patient Reported Outcomes Measurement Information System
QOL	Quality of Life
REMS	FDA mandated Risk Evaluation and Mitigation Strategy
SAE	Serious Adverse Event
SAS	Statistical Analysis System

1 BACKGROUND

This study will address the efficacy of pomalidomide in the treatment of epistaxis in patients with Hereditary Hemorrhagic Telangiectasia (HHT) who have anemia and/or require blood transfusion or iron infusion for treatment of bleeding-induced anemia and iron deficiency. HHT is an inherited disease clinically diagnosed using the Curacao criteria, which consists of 1) spontaneous and recurrent epistaxis, 2) telangiectasias at characteristic sites, 3) visceral arteriovenous malformations (AVMs) or telangiectasias, and 4) a first degree relative with HHT (inheritance is usually autosomal dominant). Patients with three criteria are considered to have definite HHT, and those with 2 criteria probable HHT, while one or no criteria make the diagnosis unlikely. Patients eligible for this study will have a diagnosis of definite or probable HHT. Estimates suggest that HHT affects between 1 in 1,330 and 1 in 10,000 individuals. Since many physicians are not familiar with the disease, only 10% of patients with HHT may actually be diagnosed. Significant manifestations of HHT often do not appear until the third or fourth decades, sometimes later.

There are no evidence-based guidelines for managing HHT, and there is no medical therapy universally accepted as efficacious; thus, treatment is not standardized. Historically, interventional procedures have formed the mainstay of HHT therapy. However, though often effective initially, such procedures are associated with a high incidence of recurrent bleeding from new AVMs or AVMs not visualized during endoscopic procedures. Epistaxis is the most common clinical manifestation of HHT, affecting ≥ 80% of patients, and leading to significant impairment in quality of life. There exists very little information concerning the effect of therapy on QOL in patients with HHT. To assess the effect of pomalidomide on QOL in HHT patients, we will administer QOL assessments before, during and after therapy.

2 PURPOSE OF THE ANALYSIS

This statistical analysis plan (SAP) contains detailed information about statistical analysis to be performed to assess the efficacy and safety of pomalidomide as a treatment for reducing the severity of epistaxis.

3 STUDY OBJECTIVES AND OUTCOMES

3.1 Study Objectives

3.1.1 Primary Objectives

1. To determine efficacy of pomalidomide compared to placebo for the reduction in severity of epistaxis after 24 weeks of treatment.

3.1.2 Secondary Objectives

1. To determine the safety and tolerability of pomalidomide for the treatment of HHT.
2. To determine if pomalidomide treatment improves quality of life in HHT.
3. To determine whether a continued response to pomalidomide is evident 4 weeks after treatment discontinuation.
4. To develop a biorepository for future studies to define biomarkers predictive of pomalidomide response and allow investigations into the biology of HHT and mechanisms of pomalidomide.

3.2 Outcomes

3.2.1 Primary Outcomes

The primary endpoint of this study is change of the Epistaxis Severity Score from baseline to week 24 in the placebo and pomalidomide-treated groups. The ESS at baseline and each follow-up visit will reflect the patient's symptoms and bleeding over the previous 4 weeks.

3.2.2 Key Secondary Outcomes

1. Amount of parenteral iron administered (in mg) during the 24-week treatment period in the pomalidomide and placebo groups (calculated as average mg/4-week interval to account for patients who discontinue early).
2. Amount of packed red blood cell transfusions (in units) during the 24-week treatment period in the pomalidomide and placebo groups (calculated as average units/4-week interval to account for patients who discontinue early).
3. Change in Neuro-QoL™ Satisfaction with Social Roles and Activities Short Form (V1.1) T-score from baseline (randomization visit) to weeks 12 and 24 (key timepoint), and the 4-week post-treatment follow-up visit in the pomalidomide and placebo groups.
4. Change in the HHT-specific QOL total score from baseline to weeks 12 and 24 (key timepoint), and the 4-week post-treatment follow-up visit in the pomalidomide and placebo groups.
5. Change in average daily epistaxis duration from the four-week screening period prior to baseline to weeks 8-12 and to weeks 20-24 (key timepoint), and to weeks 1-4 post-treatment in the pomalidomide and placebo groups.

3.2.3 Other Secondary Outcomes

1. Change in PROMIS® Emotional Distress-Depression Short Form (V1.0) T-score from baseline (randomization visit) to weeks 12 and 24 (key timepoint), and the 4-week post-treatment follow-up visit in the pomalidomide and placebo groups.
2. Change in PROMIS® Fatigue Short Form (V1.0) T-score from baseline (randomization visit) to weeks 12 and 24 (key timepoint), and the 4-week post-treatment follow-up visit in the pomalidomide and placebo groups.
3. Proportion of patients requiring no red blood cell transfusion or parenteral iron infusion during the 24-week treatment period in the pomalidomide and placebo groups.
4. Change in the ESS from baseline to that recorded at each individual patient assessment, including the 4-week post-treatment follow-up visit.
5. Change in the ESS from baseline to the average of the week 16, 20 and 24 assessments.
6. Proportion of patients requiring endoscopic interventions for management of bleeding during the 24-week treatment period in the pomalidomide and placebo groups.
7. Incidence and severity of adverse events in the pomalidomide and placebo groups including but not limited to:

    - Venous thromboembolism
    - Arterial thromboembolism
    - Thrombocytopenia
    - Neutropenia
    - Peripheral neuropathy
    - Fatigue
    - Constipation/diarrhea
    - Rash

- Any other AEs or SAEs of at least moderate severity that are possibly related to pomalidomide and deemed medically relevant at masked data review.

8. Change in iron/hemoglobin-related studies including iron saturation, ferritin, hemoglobin, hematocrit, MCV, and MCHC.

3.2.4 Exploratory Outcomes

1. Rate of relapse at the 4-week post-treatment follow-up visit, where relapse is defined as a return of the ESS to the baseline value or greater in pomalidomide (or placebo)-treated subjects.
2. Median time after initiation of pomalidomide to obtain a decrease in the ESS of at least 1.0 in the in pomalidomide-treated subjects compared to placebo.
3. Effect of genotype on the response to pomalidomide.

Outcome Modification History:
The protocol originally specified that the PROMIS® Satisfaction with Social Roles and Activities Short Form 8a (V2.0) was to be the 3rd key secondary outcome. Part-way through the study (summer 2022) it was determined that due to an inadvertent error, the CRF meant to collect the PROMIS instrument instead corresponded to the Neuro-QoL™ Short form V 1.1 instrument with the same name, number of items, and response categories, but different items. As the study was over 50% enrolled when this issue was identified, the decision was made to update the protocol to specify the Neuro-QoL instrument instead of the PROMIS instrument. This update was made in Protocol version 2.10.

Although developed as part of an assessment system for people with neurological conditions, the NeuroQoL Satisfaction with Social Roles and Activities is scored relative to a general population sample (Gershon et al 2012), supporting its use in non-neurologic health conditions. For both the PROMIS and Neuro-QoL instruments, a T-score of 50 represents the average for the United States general population with a standard deviation of 10, and calibration testing was performed on a large sample of the general population. PROMIS was developed in 2016. Neuro-QoL was developed in 2014, and the T-score calculations are based on a normative population of over 5,000 individuals. The protocol was initially designed with a 12-week follow-up period, which was modified to a 4-week period in Protocol version 2.8 (August 2, 2021). Prior to the change, 20 participants had completed the 12-week rather than 4-week follow-up.

Additionally, protocol update version 2.8 (8/2/2021) clarified that the early termination visit available in the EDC system was to be used only for participants who withdrew consent and did not agree to complete the 4-week post-treatment follow-up. Prior to the amendment, there were a number of subjects with an early termination (ET) visit which should have been recorded as a regular study visit or a 4 week (or 12 weeks) post-treatment visit. These cases were manually reviewed, and the ET visits will be re-classified in all reporting and statistical analysis to be either a C1-C6 treatment visit, or an end-of-study (EOS) visit. If the number of days since the prior visit and prior study drug dispensing was within 28 days, it was recoded to the next cycle visit. If the number of days since the prior visit or the prior study drug dispensing was > 28 days, then it was recoded as an EOS visit. Details are documented in the analysis file specification document.

4 STUDY METHODS

4.1 Overall Study Design and Plan

This is a multi-center, double blind, placebo-controlled study that will investigate the efficacy and safety of pomalidomide in patients with HHT and chronic epistaxis with iron-deficiency anemia and/or requiring intravenous iron infusions or blood transfusion.

After recruitment, patients will be screened for study eligibility. The screening evaluation will include the ESS with three-month recall, which will reflect the patient's history of epistaxis and bleeding over the prior three months, as well as detailed review of iron infusion and red cell transfusion over the preceding six months. Eligible patients will be provided a diary to record the duration of each epistaxis event that occurs during the 4-week period between screening and baseline visits. Eligible patients will then return for the randomization visit between 28 and 56 days later, at which time patients will undergo genetic testing if this has not been previously performed or if results for ENG, ACVRL1 and SMAD4 genes are not available. A repeat ESS with 4-week recall will be performed and will be considered the "baseline" ESS. Patients will also take QOL surveys at this visit. Patients will then be randomized 2:1, stratified by study site, to either pomalidomide, 4 mg/day for 28 days during each of six 28-day cycles (24 weeks), or placebo, administered on an identical schedule.

Patients will be seen every four weeks during the study (visits C1-C6 at weeks 4, 8, 12, 16, 20, 24 and 4-weeks post treatment follow-up) to measure the ESS (with 4-week recall, see **Appendix 1**), measure iron stores, and obtain laboratory studies, which will include CBC, metabolic profile, and iron studies. The iron replete state will be defined as a ferritin $\geq$ 50 ng/ml and a transferrin saturation $\geq$ 30% (since ferritin levels may be elevated due to concurrent inflammation). Investigators will be encouraged to replete iron stores in patients in both arms of the study based on these parameters, and the amount of iron infused (in mg) recorded. There is no mandate for the use of a particular iron preparation. Investigators will be encouraged to use iron infusions prior to blood transfusion, though the latter will be mandated for a hemoglobin below 6.5 gm/dl and recorded as units of packed red blood cells administered. Blood transfusion may also be given at the investigator's discretion based on symptoms that may include shortness of breath, severe fatigue, or other cardiovascular manifestations. Patients will be assessed for adverse events (AE) throughout the study. Treatment dosage may be reduced, or temporarily or permanently discontinued following AE-specific guidelines related to fatigue, cytopenia, or other toxicities.

The 4-week diary of epistaxis events (see **Appendix 2**) will be recorded at selected intervals between visits throughout the study (weeks 8-12 and 20-24 as well as 0-4 weeks post treatment follow-up). The diary was collected from participants via a phone app developed by RTI available for all smart phones, or as a paper diary for participants without access to a smartphone.

The effect of pomalidomide on QOL will be assessed at the baseline visit, at the 12- and 24-week visits, and at the 4-week post-treatment follow-up visit using validated NIH instruments of 1) NeuroQoL satisfaction with social roles and activities, 2) PROMIS emotional distress – depression, and 3) PROMIS fatigue. We will also assess responses to an HHT specific questionnaire developed specifically for this study and shown to correlate with these PROMIS items. The effect of pomalidomide on duration of epistaxis will be assessed at the 12 and 24-week treatment visits and at the 4-week post-treatment follow up visit.

4.2 Study Population

The study population is defined by the following eligibility criteria.

Inclusion Criteria:

1. A clinical diagnosis of HHT as defined by the Curacao criteria
2. Age > 18 years
3. Platelet count ≥ 100 x 109/L
4. WBC ≥ 2.5 x 109/L
5. INR ≤ 1.4 and normal ± 2 sec activated partial thromboplastin time (aPTT or PTT per local laboratory designation) by local laboratory criteria (except for patients on a stable dose of warfarin or direct oral anticoagulants)
6. Epistaxis severity score ≥ 3 measured over the preceding three months, measured at the screening visit
7. A requirement for anemia, as determined by local laboratory hemoglobin assessment and normal ranges, and/or parenteral infusion of at least 250 mg of iron or transfusion of 1 unit of blood over the 24 weeks preceding the screening visit
8. All study participants must agree to be registered into the FDA mandated POMALYST REMS® program, and be willing and able to comply with the requirements of the POMALYST REMS® program
9. Females of childbearing potential (FCBP) must adhere to the pregnancy testing schedule mandated by the POMALYST REMS® program.
10. Ability to understand and sign informed consent

Exclusion Criteria:

1. Women currently breast feeding or pregnant
2. Renal insufficiency, serum creatinine > 2.0 mg/dl
3. Hepatic insufficiency, bilirubin > 2.0 (or >4.0 in the setting of a prior clinical or genetic diagnosis of Gilbert's syndrome) or transaminases > 3.0x normal
4. Prior treatment with thalidomide or other Immunomodulatory imide drugs (IMiDs) within previous 6 months
5. Prior treatment with bevacizumab (systemic or nasal) within previous 6 weeks
6. Prior treatment with pazopanib within previous 6 weeks
7. The use of octreotide or oral estrogens within the previous month
8. History of prior unprovoked thromboembolism confirmed by venous ultrasound or other imaging modalities
9. Known peripheral neuropathy, confirmed by neurologic consultation
10. Known underlying hypoproliferative anemia (i.e., myelodysplasia, aplastic anemia)
11. Currently enrolled in other interventional trials
12. Known hypersensitivity to thalidomide or lenalidomide.
13. The development of erythema nodosum if characterized by a desquamating rash while taking thalidomide or similar drugs.
14. Known SMAD-4 mutation
15. Anything that in the investigator’s opinion is likely to interfere with completion of the study

4.3 Study Arm Assignment and Randomization

159 randomized individuals are required for this study. Randomization will be performed at a 2:1 (pomalidomide:placebo) manner to enhance patient enthusiasm for participation in the trial.

Randomization is completed within the Medidata EDC system. Randomization is stratified by site using a permuted block design with randomly varying block sizes of 3 and 6. For further details on randomization see the **PATH-HHT Randomization Specification**.

4.4 Masking and Data Lock

4.4.1 General Masking Procedures

This is a double-blind, placebo-controlled clinical trial. As such, neither the participant receiving study drug nor the site staff that administers study procedures and assessments will be aware of the treatment assignment. Additionally, all members of the clinical coordinating center at Cleveland Clinic including PI Keith McCrae and the data coordinating center at RTI including DCC PI Sonia Thomas are masked except for statisticians responsible for developing the randomization schedule, coordinating study drug supplies with Celgene, and producing DSMB reports. Any intentional or unintentional unmasking will be reported as a protocol deviation in the data management system. For further details on masking and procedures for unmasking see the **PATH-HHT Blinding Plan**.

4.4.2 Database Lock

The database will be locked at the completion of study follow-up. Randomization assignment will not be unmasked beyond the details in the blinding plan until after database lock is finalized.

4.5 Study Flow Chart of Assessments and Evaluations

	Screen and Baseline		Treatment Period						End of Study	
Visit	**SC**	**BL**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**EOS**	**ET**
Study Procedures	**Week -4 to -8**	**Week 0**	**Week 4 0-5 Days**	**Week 8 0-5 Days**	**Week 12 0-5 Days**	**Week 16 0-5 Days**	**Week 20 0-5 Days**	**Week 24 0-5 Days**	**Week 28 (4 week follow up) 0-7 Days**	**Only for withdrawn consent**
Informed Consent	X									
Enrollment Criteria	X									
Demographics	X									
Medical History	X									
Concomitant Medications	X	X	X	X	X	X	X	X	X	X
Physical Exam	X	X[1*]	X[1]	X[1]	X	X[1]	X[1]	X	X	X
Vital Signs[2]	X	X	X	X	X	X	X	X	X	X
Genetic Testing[3]		X								
Iron Infusion	X[4]	X	X	X	X	X	X	X	X	X
Blood Transfusion	X[4]	X	X	X	X	X	X	X	X	X
Adverse Events		X	X	X	X	X	X	X	X	X
ESS	X[5]	X	X	X	X	X	X	X	X[5]	X
Epistaxis Duration Diary	X	X[6]		X	X[6]		X	X[6]	X[6]	
CBC with differential	X	X	X	X	X	X	X	X	X	X
CMP	X	X	X	X	X	X	X	X		X
Iron Studies[7]	X	X	X	X	X	X	X	X	X	X

	Screen and Baseline		Treatment Period						End of Study	
Visit	**SC**	**BL**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**EOS**	**ET**
Study Procedures	**Week -4 to -8**	**Week 0**	**Week 4 0-5 Days**	**Week 8 0-5 Days**	**Week 12 0-5 Days**	**Week 16 0-5 Days**	**Week 20 0-5 Days**	**Week 24 0-5 Days**	**Week 28 (4 week follow up) 0-7 Days**	**Only for withdrawn consent**
INR and aPTT	X									
Biorepository Collection	X[8]	X[8*]			X[8]			X	X[8]	X
QOL[9] surveys		X			X			X	X	X
Pregnancy Testing[10]	X	X	X	X	X	X	X	X		
Randomization		X								
Drug Dispensation		X	X	X	X	X	X			
Drug Return			X[11]	X[11]	X	X[11]	X[11]	X		X

SC = Screening; BL = Baseline; C = End of cycle; ESS = Epistaxis severity score; CBC = Complete blood count; CMP = Complete metabolic panel; EOS = End of study Visit; ET = Early termination; QOL = Quality of life; Gray shaded boxes denote flexibility to perform visits remotely in which case physical exam may be modified; vital signs and biorepository samples may not be collected.

[1] Limited physical examination; *A detailed physical exam if screening visit was remote.

[2] Vital signs will include body weight, blood pressure, pulse and O2 saturation.

[3] Genetic testing will be performed only if not done previously and consent is obtained.

[4] Information on Iron infusion and Blood transfusion over the past 6 months to be collected at screening visit.

[5] ESS will be a 3-month recall for these visits; for all other visits, it will be a 4-week recall.

[6] Epistaxis Duration Diary will be collected at these visits and if applicable a new diary will be dispensed for the next recording.

[7] Iron studies include ferritin levels, serum iron and total iron binding capacity.

[8] One PAXgene RNA tube is included at these visits; [8*] One PAXgene RNA tube will be included if screening was remote.

[9] QOL questionnaires will include 3 NIH instruments (NeuroQoL Satisfaction with Social Roles and Activities Short Form V1.1, PROMIS Depression Short Form V1.0 and PROMIS Fatigue Short Form V1.0) and an HHT specific questionnaire.

[10] In FCBP, pregnancy tests will be performed 10-14 days before, and within 24 hours of study drug order. During the first treatment cycle, weekly pregnancy testing is mandated. Thereafter, pregnancy testing will be performed once every 4 weeks, during each cycle of treatment, preferably during the middle of the treatment cycle.

[11] Information on remnant drug and empty bottles will be collected over phone, email, or other means, if visit is remote.

5 ANALYSIS POPULATIONS

Randomization can occur either shortly before or at the baseline visit.

- If the screening visit was completed in-person, randomization can be completed before the baseline visit and study drug is ordered by the site from McKesson (manager of the study drug for Celgene and implementer of the Pomalidomide REMS system) to be shipped to the site so it is available at the site on the day of the baseline visit.
- If the screening visit was completed remotely, then eligibility is confirmed at the baseline visit and randomization occurs at baseline, and the drug is ordered to ship directly to the participant or to the site for dispensing to the participants within a few days of the baseline visit.

Under these scenarios, occasionally a randomized participant might withdraw consent before receiving the study medication from McKesson. Participants might also withdraw consent after receiving study medication but before the first follow-up visit and they may refuse to provide any follow-up assessments. Because these patients have no efficacy data to contribute to the analysis, they are excluded from the efficacy analysis populations, as follows:

Modified Intention-to-Treat Population (mITT)

The modified intention-to-treat population will comprise of all randomized participants who took at least one dose of the double-blind study medication AND have at least one post-baseline efficacy measurement.

Note that if a participant receives study drug sent by McKesson and does not return a full drug bottle to the site, it is assumed that they took at least some study drug.

Participants who receive study drug but do not complete the study will be used in all analyses for which data are available. Participants who withdraw consent for further follow-up will be treated as lost to follow-up at the time that consent was withdrawn, and data will be used up to the time of study withdrawal.

Per-Protocol Population (PP)

The per-protocol population will comprise of all randomized participants who complete the study with no major protocol violations and took at least 80% of the prescribed double-blind medication. Participants who were lost to-follow-up will be considered non-compliant for the purposes of the per-protocol population.

The blinded data review meeting identified the per-protocol exclusions as the following:

- Participants who did not take at least 34 pills of study drug (20%) for reasons other than AEs/side effects,
- Participants discontinued due to lost to follow-up or withdrawal of consent,
- One case of discontinuation due to joint decision of PI and participant, and one case due to PI decision,
- Two participants who discontinued study drug due to AEs but continued in the study with efficacy assessments for multiple months after study drug discontinuation.

Safety Population (SAF)

The safety population will comprise all participants that receive any study drug, irrespective of amount or duration of treatment.

6 SAMPLE SIZE DETERMINATION

We anticipate consenting up to 200 participants to reach 159 randomized participants.

A total sample size of 159 patients (106 pomalidomide, 53 placebo) will provide 94% power to identify a treatment group difference in ESS change from baseline at 24 weeks of 1.0 with $p<0.044$ (per interim analysis alpha spending rules), assuming a standard deviation of 1.7 and 10% early discontinuation rate with the planed MMRM analysis. Power will remain above 80% if there is a substantially higher than expected study discontinuation rate of 30% (with 10% providing no post-baseline data): 81% based on a conservative T-test, or 90% with the planned MMRM analysis. If the standard deviation is 1.8 instead of 1.7, the power for the planned MMRM analysis would be 91% with 10% discontinuation, and 87% with 30% discontinuation.

With 159 patients, we will also have at least 90% power to identify a treatment group difference of 30% for a binary secondary endpoint of proportion of participants requiring at least one packed red blood cell transfusion or iron infusion at any time during the 24-week treatment period with $p < 0.05$, whether assuming a rate in the placebo and pomalidomide groups of 90% vs. 60%, or 50% vs. 20%.

7 STATISTICAL/ANALYTICAL ISSUES

7.1 General Rules

All statistical computations will be performed, and data summaries created using SAS 9.4 or higher. For summaries of study data, categorical measures will be summarized in tables listing the frequency and the percentage of subjects in each study arm; continuous data will be summarized by presenting mean, standard deviation, median and range; and ordinal data will be summarized by presenting median and range.

Estimates from linear models will be presented by mean, standard error, and 95% CI. Estimates from logistic regression will be presented by odds ratios and 95% CIs. All statistical tests will be 2-tailed.

Checks of assumptions (e.g., normality) underlying statistical procedures will be performed and corrective procedures such as use of a transformation or application of nonparametric tests will be applied.

7.2 Adjustments for Covariates

Indicator variables for the study stratification of pooled clinical site (Section 7.6) will be included as a covariate in most efficacy analyses performed for this study (details in Section 9). Also, for all scores analyzed as a change from baseline, the baseline value will also be a model covariate.

Additionally, demographic and baseline characteristics for subjects will be compared between study arms using analysis of covariance techniques for continuous measures, Mantel-Haenszel mean score test using standardized mid-rank scores for ordinal measures, and chi-squared tests for general association for categorical measures. If sample sizes allow, these analyses will control for the study stratification factor of pooled site. If these analyses identify any substantial differences among arms, the use as covariates of these parameters on which the arms differ will be explored in sensitivity analyses of the primary efficacy outcome (see Section 9.3.1).

7.3 Handling of Dropouts and Missing Data

Missing data and missed visits will be monitored in real-time. In the event a scheduled visit is missed, site coordinators will contact the patient before the follow-up window closes to encourage them to at least obtain measurements relevant to the primary endpoint. For any missing data that do occur, we will record and describe the reasons the data are missing when reporting results

In the event that one or more ESS or QOL visit assessments is missing during the treatment or post-treatment periods, the MMRM analysis approach which assumes data are missing at random (MAR) will be used to compare treatments based on the available data and the correlation between observations within a person. To guard against bias if the MAR assumption is incorrect, missing data mechanisms will be explored, and sensitivity analyses will be conducted on the primary outcome to assess the robustness of the primary analysis (see Section 9.3.1). Sensitivity analyses are not planned for secondary analyses.

7.4 Interim Analyses and Data Monitoring

Safety outcomes will be assessed at each DSMB meeting. Rates of safety outcomes will be compared between treatment groups using Fisher's exact tests and provided to the DSMB. At each meeting, the DSMB will be presented with information about enrollment and outcome data attainment (for

example, the percent of expected visits that have been completed) to allow them to determine that the study is making reasonable progress. These review meeting will occur once each 6 months.

A formal interim analysis for efficacy after 50% and 75% of patients have completed the 24-week treatment period or discontinued from the study. See Section 9.6 for details.

7.4.1 Study Halting Rules

Randomization of new participants will be halted when eight thromboembolic events, defined as any venous thromboembolic event requiring medical intervention (grade 2 or higher) or any arterial thromboembolic event, that are determined to be related or "probably related" to the study drug are reported to the DCC.

The DCC will notify the study monitor, study sponsor, PI and investigators immediately when the eighth thromboembolic event (as defined above) is reported, and randomization functionality will be deactivated. The PI will inform the DSMB members within 24 hours of this occurrence and the DCC will provide the DSMB with AE listing reports. The DSMB will convene an ad-hoc meeting by teleconference or in writing as soon as possible. The DSMB will provide recommendations for proceeding with the study to the study sponsor/NIH. The clinical coordinating center will inform NHLBI and the FDA of the temporary enrollment halt and the disposition of the study.

7.5 Masked Data Review

A masked data review of the primary outcome and secondary outcomes for this study will be performed by the protocol team. This review will occur prior to completion of the primary analyses prior to any official unblinding. This will include a presentation of descriptive statistics (e.g., means, standard deviations, percentiles for continuous variables and counts and percentages of categorical variables) of the selected outcomes and model predictor variables.

7.6 Multicenter Studies

For this multicenter study, randomization of study participants was stratified within center. Consequently, for all model-based primary and secondary analyses, center will be included as a fixed effect in the models, with small centers pooled.

Sites will be pooled based on total number of randomized patients and geographical location prior to unblinding, such that sites with fewer than 10 participants will be pooled, and no pooled site is larger than the largest un-pooled site. If there are several sites with very low enrollment, one pooled site may consist of these sites rather than pooling based on geographical location.

Masked data review prior to database lock identified the following site pooling:

- 1 pooled group of small sites (n=24): Cleveland Clinic (n=7), Johns Hopkins (n=7), UCSF (n=2), Utah (n=6), Florida (n=1) and UCSD (n=1)
- 5 non-pooled sites: MGH (n=51), UNC (n=21), Penn (n=16), Wisconsin (n=12), Mayo (n=21)

As an ancillary analysis associated with the primary outcome, we will examine descriptively whether the treatment effect varies across sites; however, no other analyses will assess site differences in treatment effect because sample sizes are inadequate to support evaluation of site-level effects.

7.7 Multiple Comparisons and Multiplicity

For inferential tests, the p-value for determination of statistical significance (Type I error) for the single primary efficacy outcome (change in ESS from baseline to week 24) will be evaluated relative to 0.044 to adjust for the interim efficacy analysis (see Section 9.6). However, the 75% interim analysis was conducted at 80% due to timing of DSMB meetings. This changed the significance level for the primary outcome from 0.044 to **0.0425** (see Appendix 3).

Significance of the treatment group comparison for 5 key secondary outcomes will be evaluated as follows: If the primary endpoint is found statistically significant at $p < 0.0425$, then the key secondary endpoints will be evaluated for statistical significance with a Hochberg modification to the Bonferroni adjustment (Hochberg, 1988), in which the p-values of the 5 outcomes will be ordered. The largest p-value will be compared relative to $p < 0.05$, and if met, all 5 endpoints will be considered significant. If not, then the second largest p-value will be assessed relative to $p < 0.05/2 = 0.025$, and if met then it and the other three endpoints will be considered significant, and so on for the third p-value compared at $0.05/3=0.017$, the fourth compared to $0.05/4 = 0.012$, and the fifth compared to $0.05/5 = 0.010$. Results from analyses of all other secondary endpoints and subgroup analyses will not be evaluated relative to statistical significance.

7.8 Examination of Subgroups

We will evaluate treatment group comparisons for the primary outcome and key secondary outcomes (each in the primary specified analysis population using the primary model and primary timepoint) in subgroups of participants defined by:

- genetic confirmation of HHT (yes or no), defined as having one or more mutations of three primary genes associated with HHT: endoglin (ENG), ACVRL1 and SMAD4
- gender, and
- race (per NIH guidelines), defined as white versus non-white.

7.9 Assessment Windows

For the primary analysis, decisions about how to treat out-of-window visits will be made prior to unmasking data. We intend for all data to be used but will evaluate if any visits fall substantially within another visit window. For secondary analyses, all available data will be used.

8 STUDY PARTICIPANT CHARACTERIZATION

8.1 Participant Disposition

Participant eligibility status will be summarized, and overall disposition of study participants will be described using a standard CONSORT diagram. The number of participants randomized and those completing or discontinuing from study therapy will be summarized. Reasons for study withdrawal will be listed.

8.2 Protocol Deviations

Protocol deviations will be listed with information such as type of deviation, time of occurrence, and reason. The number of protocol deviations will also be summarized overall and for each protocol deviation category.

8.3 Study Drug Exposure and Compliance

All randomized participants will receive the double-blind study drug either at the baseline visit and at subsequent scheduled study visits. The first dose of study drug will be given in the amount of 4 mg pomalidomide or placebo, depending on randomization assignment. Dose can be modified by the investigator (blinded to active versus placebo) in response to side effects to 3 or 2 mg. All bottles contain 28 pills, for a 4-week supply.

The number and percentage of participants receiving each dose level will be presented by visit and treatment group.

Participants are instructed to return their medication bottles after each visit and remaining pills are counted by study staff. Overall study drug compliance will be calculated as the number of pills dispensed minus the number returned divided by the total number that should have been taken from the first dose up until study drug completion or early drug discontinuation. The number and percentage of participants that are at least 80% compliant will be presented. When a participant does not return a bottle, a note is recorded on the CRF as to why it was not returned. The notes will be used to assign an assumed number taken based on that note (for example, all pills taken, or no pills taken)

8.4 Demographic and Baseline Characteristics

Demographic and baseline clinical characteristics and HHT medical history for the study participants will be summarized by study arm using the general analysis rules describe above. Demographic variables of interest include gender, age (years), race and ethnicity), education, employment and marital status, BMI, and baseline levels of all vital signs and QOL measures. Extensive HHT medical history is recorded on a CRF, plus genetic testing.

9 EFFICACY ANALYSES

9.1 Overview of Efficacy Analysis Methods

- All efficacy analyses will be performed on the mITT population, the primary population.
- All efficacy variables will be summarized by treatment group at baseline and at subsequent study time points at which assessments were administered and collected. Count (N), mean, standard deviation, minimum, and maximum will summarize continuous efficacy variables, whereas number and percent will summarize categorical efficacy variables.
- Unless otherwise noted, all analyses of dichotomous outcomes, measured at respective time points, will be performed using models as specified in Section 9.3. Efficacy models will be adjusted for stratification by clinical site. Baseline value will be an additional covariate for all analyses of change from baseline scores. Consistent with the description of the primary analysis in the protocol, analyses of primary and secondary outcomes will also include an independent variable for study visit. For analysis of measures assessed at multiple time points (for example, C1 and C6), longitudinal modeling will be used and the interaction between time and treatment groups will be included. Unless otherwise noted, all analyses of continuous efficacy variables (e.g., ESS and QOL scales) will be performed using general linear mixed models. Variables with distributions substantially different from normal will be transformed prior to analysis.

  Models will be adjusted for pooled clinical site. If there are not enough patients per clinical site, similar sites will be combined (see Section 7.6)

9.2 Efficacy Variables

Primary and secondary efficacy variables as well as exploratory and safety outcomes are described in the table below.

Variable	Type	Definition
Primary Outcome		
Change from baseline in Epistaxis Severity Score at each post-baseline visit through 24 weeks and 4 weeks post-treatment	Continuous	The Epistaxis Severity Score will be computed at baseline and at each study visit during which assessments are administered and collected using the standard scoring algorithm (see **Appendix 1**). The outcome will then be computed as the difference in score at each visit and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing. The total score be calculated inside the EDC CRF page.
Key Secondary Outcomes		
Amount of iron administered through 24 weeks and 4 weeks post-treatment	Continuous	The 4-week average amount of iron transfusion (in mg/4 weeks) through 24 weeks will be calculated. For subjects who discontinue early, this value will be pro-rated to 24 weeks. Participants who receive no infusions during the 24 weeks will have a value of zero, In the screening period, the amount of iron received in the 6 months prior to screening is calculated as the number of infusions received multiplied by the amount infused at the last infusion (because the CRF only collected the amount infused at the most recent infusion prior to screening)
Amount of red blood cell administered through 24 weeks and 4 weeks post-treatment This was changed to: proportion of patients receiving at least 1 blood transfusion in the 24-week period.	Continuous Categorical (yes/no)	The 4-week average amount of red blood cell transfusion (in units/4-weeks) through 24 weeks will be calculated. For subjects who discontinue early, this vale will be pro-rated to 24 weeks. Note: at masked data review, it was determined the overall number of RBC transfusions was small, and this outcome will instead be modified to the percentage of participants that received any RBC transfusion during the 24-week treatment period (and separately, the 4-week post-treatment period)
Change from baseline in Neuro-QoL™ Satisfaction with Social Roles and	Continuous	The Neuro-QoL™ Satisfaction with Social Roles and Activities Short Form (V1.1) T-score will be computed at baseline and at each study visit during which assessments are administered and collected. Total score

Variable	Type	Definition
Activities Short Form (V1.1) T-score at 12 and 24 weeks and 4 weeks post-treatment		is calculated as a T score using Neuro-QoL online scoring tools (https://www.assessmentcenter.net/ac_scoringservice), where a score of 50 represents the average of the Neuro-QoL calibration sample, which is a general US population, with a standard deviation of 10. The outcome will then be computed as the difference in score at each post-baseline visit and the score at baseline. Missing items within an assessment are handled by the official online scoring tools.
Change from baseline in HHT QoL Total Score at 12 and 24 weeks and 4 weeks post treatment	Continuous	The HHT QoL Total Score will be computed at baseline and at each study visit during which assessments are administered and collected by summing all of the items (Kasthuri et al 2022). The outcome will then be computed as the difference in score at each post-baseline visit and the score at baseline. If any data items at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in average daily epistaxis duration 8-12 and 20-24 weeks and 4-weeks post-treatment	Continuous	The average daily epistaxis duration will be computed at baseline and at each study visit during which the epistaxis diary is collected (See **Appendix 2**). The outcome will then be computed as the difference in average at each post-baseline visit and the average at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing. The daily average is calculated by averaging all non-missing data across all available days in a dairy. At least 7 total days for each diary must be present in order to calculate the average for that time period. If the actual number of minutes for a nosebleed is not provided but the duration category is checked, then the midpoint of the duration category will be used as the duration for that nosebleed as follows: <1: 0.5, 1-5: 3, 6-15: 10.5, 16-30: 23, 31-60: 45, > 60: 60 Note: at masked data review, the unit of this outcome was modified from weekly to daily.
Other Secondary Outcomes		
Change from baseline in PROMIS® Emotional Distress-Depression Short Form (V1.0) T-score at 12 and 24 weeks	Continuous	The PROMIS® Emotional Distress-Depression Short Form (V1.0) T-score will be computed at baseline and at each study visit during which assessments are administered and collected using the PROMIS online scoring algorithm (https://www.assessmentcenter.net/ac_scoringservice).

Variable	Type	Definition
and 4 weeks post-treatment		A score of 50 represents the average of the PROMIS calibration sample, which is a general US population, with a standard deviation of 10. The outcome will then be computed as the difference in score at each post-baseline visit and the score at baseline. Missing items within an assessment are handled by the official online scoring tools.
Change from baseline in PROMIS® Fatigue Short Form (V1.0) T-score at 12 and 24 weeks	Continuous	The PROMIS® Fatigue Short Form (V1.0) T-score will be computed at baseline and at each study visit during which assessments are administered and collected using the PROMIS online scoring algorithm (https://www.assessmentcenter.net/ac_scoringservice). A score of 50 represents the average of the PROMIS calibration sample, which is a general US population, with a standard deviation of 10. The outcome will then be computed as the difference in score at each post-baseline visit and the score at baseline. Missing items within an assessment are handled by the official online scoring tools.
Proportion of patients with RBC transfusions or iron infusions through 24 weeks. The treatment goal is receiving none (a value of No).	Dichotomous (Yes/No)	The number of iron infusions and RBC transfusions will be counted throughout the 24-week treatment administration. The outcome is calculated as “No” if no transfusions/infusions, and “Yes” if ≥ 1 transfusion/infusion.
Proportion of patients requiring endoscopic intervention through 24 weeks	Dichotomous (Yes/No)	The number of endoscopic interventions for management of bleeding will be counted throughout the 24-week treatment administration. The outcome is calculated as “Yes” if ≥ 1 endoscopic intervention and “No” if no endoscopic interventions. Endoscopic interventions are recorded by the site on the concomitant medications form, can be either nasal or gastroenterological interventions. Identification of endoscopic interventions (such as nasal cautery) will be confirmed by medical review. These were identified prior to database lock and include endoscopy, cautery, and embolization.
Incidence and Severity of Specific adverse events	Discrete	The incidence rate and maximum severity of the following adverse events will be calculated: MedDRA terms identified in blinded data review noted for each.

Variable	Type	Definition

		• Venous thromboembolism
		  o Portal vein thrombosis, Deep vein thrombosis
		• Arterial thromboembolism
		  o None found
		• Thrombocytopenia
		  o Thrombocytopenia
		• Neutropenia
		  o Neutropenia
		  o also Leukopenia, and also a category of (Neutropenia or Leukopenia)
		• Peripheral neuropathy
		  o Neuropathy peripheral, Peripheral sensory neuropathy
		• Fatigue
		  o Fatigue
		• Constipation/diarrhea
		  o Constipation, Diarrhoea, Nausea, Abdominal pain
		• Rash
		  o Rash popular, Rash pruritic, Rash maculo-papular, Rash, Rash erythematous, Urticaria, Pruritus, Palmar erythema
		• Any other AEs or SAEs of at least moderate severity that are possibly related to pomalidomide and deemed medically relevant at masked data review:
		  o Pulmonary embolism
		  o Mouth ulceration
		  o Alanine aminotransferase increased
		  o Tremor
		  o Dyspnoea and exertional dyspnoea (only the cases that meet criteria of at least possibly and at least moderate)
		  o Tachycardia (only the cases that meet criteria of at least possibly and at least moderate)
		  o Muscle spasms (only the cases that meet criteria of at least possibly and at least moderate)

Change from baseline in iron (transferrin) saturation, ferritin, hemoglobin, hematocrit, MCV, and MCHC.	Continuous	Lab parameters are analyzed at each site (including at local labs) and entered onto CRFS in the EDC system. Transferrin Saturation is calculated within medidata from Iron and Iron Binding Capacity.

Variable	Type	Definition
		The outcome will then be computed as the difference in value at each post-baseline visit and the value at baseline.
Thrombocytopenia platelet categories	Categorical Normal < 150 K/uL < 100 K/uL < 50 K/uL	Added as additional planned safety presentation at blinded data review meeting. Categorical assessment of thrombocytopenia will be calculated based on all platelets collected post-baseline. A person will be classified into one of the 4 categories based on the lowest recorded post-baseline platelet value. The value of 150 represents the lower normal bound.
Neutropenia neutrophils categories	Categorical Normal < 1.45 K/uL < 1.00 K/uL < 0.50 K/uL	Added as additional planned safety presentation at blinded data review meeting. Categorical assessment of neutropenia will be calculated based on all neutrophils collected post-baseline. A person will be classified into one of the 4 categories based on the lowest recorded post-baseline neutrophil value. The value of 1.45 represents the lower normal bound.
Exploratory Outcomes		
Relapse at 4-weeks post treatment	Dichotomous (Yes/No)	The rate of relapse will be calculated at 4 weeks post-treatment. Relapse is defined as a return of the ESS to the baseline value or greater.
Treatment weeks to 1.0 decrease in ESS	Ordinal Categorical (4, 8, 12, 16, 20, 24, Not obtained)	The time to the first 1.0-unit decrease in ESS will be calculated for each participant by computing the difference between the baseline ESS and the ESS at each study visit and determining the first study visit at which the difference is <= -1.0, or “not obtained” if no visit had an improvement of at least 1.0
Oral iron taken during the treatment period	Categorical (Always taken, Never taken, Added after baseline, Stopped after baseline	All oral iron supplements will be identified from the concomitant medications form (using search for Iron, Fe, ferrous, Ferrous Sulfate) and confirmed by medical review. Oral iron usage will be categorized for each participant as always taken if it was taken during the screening period before baseline, and also taken post-baseline, added after baseline if it was started after the date of the baseline visit, stopped after baseline if it was stopped after the baseline visit, and never taken if a patient had no oral iron recorded.
Other outcomes		

Variable	Type	Definition
Other outcomes include individual questions on the ESS and the diary and iron studies not otherwise covered in a secondary or exploratory outcome		
Nosebleed frequency, duration, and intensity (ESS items 1-3) at each visit	Categorical	Reported as recorded on CRF, categories may be combined based on observed data and to assist in trend interpretation.
Sought medical attention for nose bleeding (ESS item 4) during the 24-week treatment period	Dichotomous (Yes/No)	“Yes” if reported at any visit C1-C6 and "No” if not reported at any visit C1-C6
Anemia (yes/no) (ESS item 5) at each visit	Dichotomous (Yes/No)	Reported as recorded on CRF
Iron replete status by visit, defined as a ferritin ≥ 50 ng/ml *and* a transferrin saturation ≥ 30%	Dichotomous (Replete/Not replete)	Iron replete status is defined as replete if ferritin ≥ 50 ng/ml *and* a transferrin saturation ≥ 30% and is defined as “not replete” if either ferritin < 50 ng/ml or transferrin saturation < 30%.
Hemoglobin below 6.5 gm/dl at any time during the 24-week treatment period	Dichotomous (Yes/No)	A blood transfusion was mandated for a hemoglobin below 6.5 gm/dl.
Diary frequency, duration and intensity questions	To be determined	Presentation of data on frequency, duration and intensity questions collected for each nosebleed on the diary will be considered as a secondary item after analyses specified in the SAP are completed.

9.3 Primary Analysis Methods

For the primary analysis, the change from baseline in epistaxis severity score after 24 weeks of study drug administration will be compared between the pomalidomide and placebo groups using a longitudinal general linear model, often referred to as a mixed model for repeated measures (MMRM) (Davis SM 2014), using the MITT population.

The observed change from baseline score at each scheduled post-baseline visit (Weeks 4, 8, 12, 16, 20, and 24 and 4-weeks post-treatment) is the dependent variable. For the purposes of this analysis, study visit will be treated as a categorical repeated measure. The model will include fixed effects for treatment group, study visit, and interactions between those variables and will include the baseline ESS score as a covariate. The model will also be adjusted for the design effect of stratification by pooled site. A heterogeneous Toeplitz covariance pattern will estimate the variance-covariance of the

within-subject repeated measures (study visit), and the Kenward-Roger method will be used to calculate denominator degrees of freedom. Estimates, p-values and 95% confidence intervals will be presented for treatment group comparisons at each visit. In an intent-to-treat fashion, all data will be included in the primary model, regardless of the treatment actually taken by the patient.

The 20 patients with 12-week post-treatment assessments will be included with all 4-week post-treatment assessments.

Treatment group comparisons of the change in ESS score averaged over the week 16, 20, and 24-week visit assessments (a secondary outcome) will be assessed via estimate and contrast statements within this model.

Sample SAS code for the primary model is as follows:

```
proc mixed covtest;
    where VISIT in ("C1","C2","C3","C4","C5","C6","EOS");
    class subject sitenum trtgp VISIT;
    model DELTA_ESS = ESS_BL sitenum trtgp VISIT trtgp*VISIT / ddfm=kenwardroger;
    repeated VISIT / subject=subject type=toeph R rcorr;
    title "Mixed model for change from baseline ESS";
    lsmeans trtgp*VISIT / cl pdiff=all;
run;
```

9.3.1 Sensitivity Analyses

Per-protocol population
The primary analysis will be repeated on the per-protocol population.

Imbalanced covariates
If there are important baseline characteristics that substantially differ between randomized treatment groups by chance, additional sensitivity analyses may be conducted to evaluate the impact of adjusting for those imbalances.

Clinical Site Effects
We will examine descriptively whether the treatment effect varies across sites.

Removing 12-week post-treatment data
A supportive analysis of the 4-week post-treatment assessment time will be completed for the MITT population with removal of the 20 patient visits that were conducted at 12 weeks post-treatment.

Missing data
The MMRM analysis approach assumes missing data due to missed visits or early study discontinuations are missing at random (MAR). If a statistically significant difference between treatment groups is observed in the primary analysis, then a sensitivity analysis will be conducted on the primary outcome to assess the robustness of the primary analysis to the MAR assumption using methods that assume that data are missing not at random (MNAR). Specifically, we will use multiple imputation of missing data to generate multiple imputed complete datasets (no missing data through 24 weeks), conduct analysis of covariance for change from baseline at 24 weeks on each dataset (model change from baseline at 24 weeks = baseline value + site + treatment group), and use Rubin's rule to combine results across the imputed datasets to obtain a comparison of treatment groups (Ratitich, 2014). We will explore 2 sensitivity analyses: (a) impute missing outcomes in the

pomalidomide treatment group (and the placebo group) based on data from the placebo group (control-based imputation), and (b) impute missing outcomes in the pomalidomide group at varying percentages of the observed values (while imputing missing placebo data based on observed placebo values) until a statistically significant treatment group difference is no longer identified (tipping point analysis) (O'Kelly and Ratitch, 2014). These sensitivity analyses will only be conducted if the primary analysis identifies a statistically significant treatment group difference at 24 weeks in the MITT population.

Intercurrent concomitant therapy

It is possible that participants in either treatment group may receive concomitant therapy that might impact the ESS score. For example, receiving an RBC transfusion would likely improve the anemia status; however, in this case the ESS score would have extra points for seeking medical care and receiving an RBC transfusion. Therefore, based on the ESS scoring weights, receiving an RBC transfusion would appropriately make the ESS score worse (higher score) in the month it is received. Iron infusions may also impact the anemia score, in which case there might be an extra ESS point for seeking the medical attention to obtain the iron infusion, thereby a slightly negative impact on the ESS score. On the other hand, nasal cautery or other endoscopic intervention may impact bleeding and therefore might impact the ESS score to indicate improved ESS due to the endoscopic treatment.

If applicable, we will descriptively evaluate the amount and timing of iron infusions and endoscopic interventions such as nasal cautery. If there are a substantial number of these events, we may descriptively evaluate if these events seem to have an impact on improving the ESS score, and if so, consider a sensitivity analysis of the impact of the intercurrent treatments on the primary treatment group comparisons. We will do this evaluation only if warranted by evaluation of the number of intercurrent events within the treatment groups.

Pre-specified Covariates/Subgroup analyses

We will investigate the primary outcome analysis for the MITT population on the pre-planned subgroups specified in the protocol: genetic confirmation of HHT (yes or no), gender and race (per NIH guidelines).

We will also investigate interactions for type of genetic mutation, and if the season of the year of enrollment is a relevant baseline covariate, given that there may be seasonal variation of epistaxis during colder months due to drier room air and increased prevalence of upper respiratory infections.

Interactions will be assessed within the primary mixed linear model by adding an interaction term between treatment group and subgroup and evaluating this multiple degree of freedom test at the 0.10 significance level, and if trends are identified then graphing the trajectories over time by subgroup, using a model with the treatment group x time x subgroup 3-way interaction.

9.4 Secondary Analysis Methods

Unless otherwise specified, changes from baseline in continuous secondary outcomes will be compared between treatment groups using an MMRM model as specified for the primary outcome and analogous generalized linear models will be used for any repeated dichotomous outcomes.

A Hochberg multiple comparison adjustment will be applied to the 5 key secondary endpoints to maintain the overall Type 1 error rate. Treatment groups will be compared for the secondary endpoints as described in Section 7.7.

Treatment groups will be compared for the total amount of packed red blood cell transfusions (units) and iron (mg) infused during the 24-week treatment period (calculated as average over a 4-week period), using an analysis of covariance or rank analysis of covariance if not normally distributed, with adjustment for the measure in the 6 months prior to the screening period and pooled clinical site. If packed red blood cell transfusions are rare, then we will switch to a zero-inflated Poisson regression model or logistic regression, depending on the distribution.

Blinded data review prior to database lock determined the number of RBC transfusions is rare and so RBC transfusions during the 24-week treatment period will be analyzed as yes/no using logistic regression with adjustment for transfusion in the 6 months prior to screening (yes/no) and pooled site.

The blinded data review also indicated the amount of iron infused during the 24-week treatment period is not normally distributed and will be analyzed with rank analysis of covariance (Stokes, Davis and Koch 2000). The rank ANCOVA is completed in 3 steps, (A) ranking the outcome variable and the baseline value separately within each pooled site, (B) obtaining the residuals from a regression of the ranked outcome variable on the ranked baseline variable, separately for each pooled site, and then (C) comparing the treatment groups for the residuals using a Mantel-Haenszel mean score test stratified by pooled site.

Sample SAS code for the rank ANCOVA is as follows:

```
proc rank nplus1 ties=mean out-ranks;
    by sitenum;
    var [outcome variable] [baseline variable]
run;

proc reg data=ranks;
    by sitenum;
    model [outcome ranked variable] = [baseline ranked variable;
    output out=redidual r=resid;]
run;

proc freq data=residual noprint;
    tables sitenum*trtgp*resid / cmh ;
run;
```

Groups will be compared for the proportion of patients relapsed, requiring no blood or iron, and requiring endoscopic interventions and other yes/no categorical outcomes using logistic regression with adjustment for baseline assessment (ESS score, whether the patient had required a blood transfusion or required an total iron infusion amount of at least 250 mg in the 6 months before the baseline visit, and ESS score, respectively), and study site. If events are rare, exact logistic regression will be used.

9.5 Exploratory Analysis Methods

Exploratory analyses include estimating the rate of relapse at the 4-week post-treatment follow-up visit, where relapse is defined as a return of the ESS to the baseline value or greater in pomalidomide-treated subjects and estimating the median time after initiation of pomalidomide to obtain a decrease of at least 1.0 in the ESS in pomalidomide-treated compared to placebo subjects. These outcomes will have no statistical testing.

To learn which demographic or biomarkers are predictive of greater response to pomalidomide we will fit generalized linear models to the primary outcome in the pomalidomide treatment group to predict treatment response based on presence of the specific mutations in the endoglin (ENG) or

activin-like kinase (ACVRL1) genes, as well as a set of pre-defined baseline covariates., to include age, gender, and baseline disease severity.

An exploratory analysis will evaluate the correlation of the ESS score and the average daily duration of bleeding from the diary.

An exploratory analysis will also compare groups for total iron taken during the treatment period, combining both oral iron and iron infusions.

9.6 Formal Interim Efficacy Analysis

We will perform a formal interim analysis to evaluate for early overwhelming efficacy based on the change in ESS score from baseline to 24 weeks after 50% and 75% of subjects have completed the 24-week visit or discontinued the study. Futility will also be assessed at 75%, and each formal interim analysis will include a full DSMB safety review.

For assessing early efficacy, the significance level for the interim and final analysis will be based on Lan-DeMets α-spending functions with O'Brien-Fleming boundaries in order to maintain the study-wise α level at 0.05. The α level will be 0.0031 at 50%, 0.0183 at 75% and 0.0440 at the final analysis. The exact α level at the interim will be determined using Lan-DeMets α-spending functions based on the actual percent of data available at the interim.

Assuming 90% power at the end of the study, there is a 26% chance of meeting the stopping boundary at the 50% interim, and a 69% chance at the second interim. If the study actually has 80% power (for example, due to higher-than-expected discontinuation rate or ESS standard deviation), then there is a 54% chance of meeting the stopping boundary at the second interim. Conversely, if the study actually has power > 95% (for example if treatment group difference in ESS change is 1.2 rather than 1.0 or lower than expected standard deviation), then the chance of stopping is 41% at the first interim and 84% at the second interim.

Futility will be assessed at 75% based on conditional power of the primary efficacy variable. The DSMB may recommend study stop due to futility if the upper limit of the 80% confidence interval for conditional power does not exceed 50%.

With our planned enrollment rate, we expect approximately 6 months of randomization visits to take place after the 50% interim analysis, and so there is a potential to decrease the overall study sample size in the event that the DSMB recommends stopping enrollment based on the interim analysis results. If enrollment proceeds as planned, then randomization will be essentially complete by the time of the 75% interim analysis. In this case, we will request the DSMB to skip the second interim analysis in order to maximize the alpha level at the end of the study. Otherwise, if enrollment is slower than expected, then the results of the second interim analysis still provide a chance to decrease the study sample size.

Both interim analyses were carried out. Details are provided in **Appendix 2** and **Appendix 3**.

10 SAFETY ANALYSES

10.1 Overview of Safety Analysis Methods

All safety analyses will be performed using all participants in the safety population, regardless of eligibility. The frequency of toxicities potentially attributable to pomalidomide treatment will be determined (e.g., venous/arterial thromboembolism, thrombocytopenia, fatigue, constipation/diarrhea, rash). Treatment and resolution of all safety endpoints will also be documented. Treatment groups will be compared for incidence of events with a chi-squared or Fisher's exact test, and incidence density (events per person-months of exposure) will be calculated for toxicities in the pomalidomide group. Adverse event incidence rates will be included in semi-annual safety reports to the DSMB and annually to the FDA.

10.2 Adverse Events

Per the protocol, participants were asked to report any adverse events from initiation of treatment through 24 weeks follow-up. All adverse events will be collected on an adverse event log and coded using MedDRA.

Adverse events will be listed and summarized by MedDRA system organ class and preferred term. Summaries will be of the number of events and number of individuals experiencing events by treatment group and will be created for all adverse events, by severity, and by relationship to treatment. Any events starting outside of the reportable time frame will be included in separate listings and will be excluded from summary tables. If a complete onset date is unknown and it cannot be confirmed that the event occurred during this time period, then the event will be considered a treatment-emergent adverse event.

10.3 Deaths and Serious Adverse Events

A serious adverse event (SAE) is any event that is life threatening, results in death, causes or prolongs hospitalization, leads to a disability or birth defect, or requires an intervention to prevent a disability. SAEs will be listed, and SAEs and treatment-related SAEs will be summarized in the manner mentioned in Section 10.2 if there are enough events to summarize. Deaths will be listed.

10.4 Laboratory Evaluations

Laboratory exams (include blood chemistry and hematology) will be collected from each patient and will be summarized by study visit. Box and whisker plots will be generated to show the changes across study visits for key safety labs. These have been produced for the DSMB reports.

If applicable, linear mixed effect models will be constructed modeling change from baseline for key safety lab values over time with study time treated categorically and will be used to compare any post-baseline measures to baseline. Change from baseline will be assessed for substantial violations of normality and if so, a transformation will be considered.

Incidence of treatment-emergent abnormal safety labs at any post-baseline timepoint for all collected lab parameters will be tabulated (similar to that produced for the DSMB reports but DSMB was not focused on treatment-emergent labs), and treatment groups will be compared with chi-squared or Fisher's exact test, if applicable. Treatment-emergent abnormal is defined as a lab value post-baseline that is outside the normal range during the treatment period but is within the normal range at baseline (or at screening, if the baseline value is missing).

10.5 Concomitant Medications

Patients will be allowed to use any concomitant medications that they require except those listed in the exclusion criteria or protocol section 7.6. They may remain on HHT-directed medications other than those listed in the exclusion criteria if they were taking these when eligibility was determined. If they remain on such medications, they must remain on a stable dose during the study. Concomitant medications will be reviewed at screening, and at each study visit.

Incidence of taking HHT-directed medications and any exclusionary medications will be presented by treatment group, and include the following:

Excluded medications:

- octreotide,
- erythropoetic agents,
- oral estrogen or progestin
- bevacizumab (Avastin) (systemic or nasal)
- pazopanib
- ciprofloxacin
- fluvoxamine
- ketoconazole
- carbamazepine

Medications allowed to continue if on stable dose at screening but not allowed to initiate or increase dose:

- oral epsilon-aminocaproic acid (Amicar)
- oral tranexamic acid
- propranolol
- timolol
- epsilon aminocaproic acid
- tranexamic acid nasal sprays
- nasal estriol

11 REPORTING CONVENTIONS

Unless required otherwise by journal standards, the following rules are standard:

- Moment statistics including mean and standard deviation will be reported at 1 more significant digit than the precision of the data.
- Order statistics including median, min and max will be reported to the same level of precision as the original observations. If any values are calculated out to have more significant digits, then the value should be rounded so that it is the same level of precision as the original data.
- Following SAS rules, the median will be reported as the average of the two middle numbers if the dataset contains even numbers.
- Test statistics including t and z test statistics will be reported to two decimal places.
- P-value will be reported to 3 decimal places if > 0.001. If it is less than 0.001 then report '<0.001'. Report p-values as 0.05 rather than .05.
- No preliminary rounding should be performed, rounding should only occur after analysis. To round, consider digit to right of last significant digit: if < 5 round down, if >=5 round up.

12 CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

The SAP expands the analysis of the primary and second endpoints described in the protocol. No changes from that specified in the protocol were made.

13 REFERENCES

Davis SM. Chapter 5: Mixed models for repeated measures with categorical time effects (MMRM). In Clinical Trials with Missing Data: a Guide for Practitioners. O'Kelly, M and Ratitch B. Wiley. 2014.

Gershon, R.C., Lai, J.S., Bode, R. *et al.* Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res* **21**, 475–486 (2012). https://doi.org/10.1007/s11136-011-9958-8

Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika.* 1988;75:800-803.

Intro to Neuro-QoL (healthmeasures.net) https://www.healthmeasures.net/explore-measurement-systems/neuro-qol/intro-to-neuro-qol. Accessed August 30, 2022.

Kasthuri R, Chaturvedi S, **Thomas S,** Vandergrift N, Bann c, Schaefer s, Pyeritz R, Clancy M, McCrae K, Development and Performance of a Hereditary Hemorrhagic Telangiectasia Specific Quality of Life Instrument. *Blood Advances.* 2022; 6 (14): 4301–4309. Epub 22 July 2021. DOI 10.1182/bloodadvances.2022007748

O'Kelly M, Ratitch, B. Chapter 8: Analyses under missing-not-at-random assumptions. In Clinical Trials with Missing Data: a Guide for Practitioners. O'Kelly, M and Ratitch B. Wiley. 2014.

Ratitch, B. Chapter 6: Multiple Imputation. In Clinical Trials with Missing Data: a Guide for Practitioners. O'Kelly, M and Ratitch B. Wiley. 2014.

Stokes M, Davis C, Koch G Categorical Data Analysis using the SAS system. SAS Institute, 2000.

14 LIST OF POTENTIAL DISPLAYS

List of displays and table shells are in a separate document. Data displays (tables, figures, and listings) may be added, deleted, rearranged or the structure may be modified after finalization of the SAP. Such changes require no amendment to the SAP as long as the change does not contradict the text of the SAP.

15 APPENDIX 1 EPISTAXIS SEVERITY SCORE SCORING

The epistaxis severity score (ESS) was developed by Hoag et al and is a clinically-validated score for quantifying epistaxis in HHT which correlates inversely with QOL. Questions comprising the ESS are depicted in below.

Epistaxis Severity Score

Please answer each of the following questions as they pertain to your TYPICAL (usual or most common) symptoms during the past 4 weeks.

1. How often did you TYPICALLY have nose bleeding during the past 4 weeks?
   - ☐ Less than monthly [0]
   - ☐ Once per month [1]
   - ☐ Once per week [2]
   - ☐ Several per week [3]
   - ☐ Once per day [4]
   - ☐ Several per day [5]

2. How long did your nose bleeding episodes TYPICALLY last during the past 4 weeks?
   - ☐ Less than 1 minute [0]
   - ☐ 1 to 5 minutes [1]
   - ☐ 6 to 15 minutes [2]
   - ☐ 16 to 30 minutes [3]
   - ☐ More than 30 minutes [4]

3. How would you describe your TYPICAL nose bleeding intensity during the past 4 weeks?
   - ☐ Not Typically Gushing or Pouring [0]
   - ☐ Typically Gushing or Pouring [1]

4. Have you sought medical attention for your nose bleeding during the past 4 weeks?
   - ☐ No [0]
   - ☐ Yes [1]

5. Are you anemic (low blood counts) currently?
   - ☐ No [0]
   - ☐ Yes [1]
   - ☐ I don't know [2]

6. Have you received a red blood cell transfusion SPECIFICALLY for nose bleeding during the past 4 weeks?
   - ☐ No [0]
   - ☐ Yes [1]

Clinical characteristics of the bleeding frequency and severity are collected and assigned a numerical score, which is added and multiplied by specific coefficients to determine the final ESS. The maximum ESS score is 10, with patients in the 7-10 range being considered to have severe epistaxis.

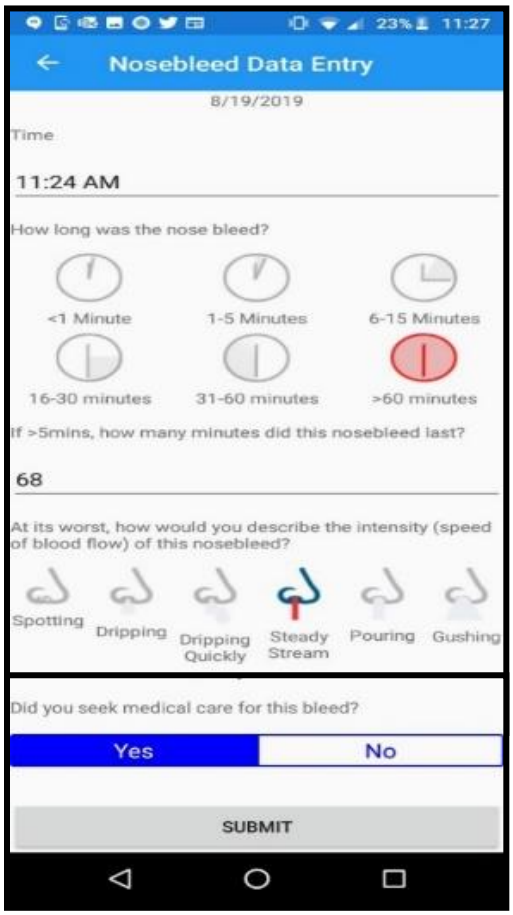
Calculations for determining the ESS are depicted below, and an automated determination of the ESS is programmed to be automatically calculated in the EDC system. The presence of missing data will result in no score being calculated.

Calculation of the Epistaxis Severity Score

Question	Response	Multiplied by		Coefficient	Result
1	Less than monthly	0	X	0.14 (0.70 Den)	
	Once per month	1			
	Once per week	2			
	Several per week	3			
	Once per day	4			
	Several per day	5			
2	< 1 minute	0	X	0.25 (1.00 Den)	
	1-5 minutes	1			
	6-15 minutes	2			
	16-30 minutes	3			
	>30 minutes	4			
3	No	0	X	0.25 (0.25 Den)	
	Yes	1			
4	No	0	X	0.30 (0.30 Den)	
	Yes	1			
5	No	0	X	0.20 (0.20 Den)	
	Yes	1			
	I don't know	Drop		0	
6	No	0	X	0.31 (0.31 Den)	
	Yes	1			
Total				**Denominator (Sum Den)**	**Raw Score**

Normalized HHT-EES = [Raw Score / Denominator (2.76)] X **10**

16 APPENDIX 2 SCREEN SHOT OF EPISTAXIS DIARY



17 APPENDIX 2 INTERIM ANALYSIS AT 50%

The purpose of the 50% interim analysis is to assess for early efficacy of the primary efficacy outcome, the change from baseline in the ESS totals score at 24 weeks. Futility is not assessed at the 50% interim.

As stated in the protocol, the significance level of the primary outcome at the interim and final analysis will be based on Lan-DeMets α-spending functions with O'Brien-Fleming boundaries in order to maintain the study-wise α level at 0.05. The α level will be 0.0031 at 50%, 0.0183 at 75% and 0.0440 at the final analysis. The exact α level at the interim will be determined using Lan-DeMets α-spending functions based on the actual percent of data available at the interim.

The interim analysis will be reviewed by the DSMB at the meeting on October 19, 2022, based on data from a database snapshot date of **September 6, 2022**.

If the significance level of the primary outcome is less than the pre-specified alpha level, then the DSMB may recommend that study enrollment be halted, with currently active patients completing their treatment and follow-up, and the database be cleaned and locked for the final statistical analysis.

Study Population, Sample size and Alpha level

The interim efficacy analysis will be conducted on all randomized and treated participants (the modified Intent to Treat (MITT) population that will be used for the final statistical analysis as specified in the SAP) that have either completed the 24-week (C6) visit or have discontinued from the study. Participants that have completed C6 but have not yet completed the 4-week post-treatment follow-up are included.

Randomized participants that are ongoing in the trial and have not yet reached the C6 visit will not be included in the interim analysis. While their data collected as of the data snapshot is fully applicable, the repeated measures model used for the ESS 24-week analysis assumes that the data not yet collected for ongoing participants is missing at random and would be assumed to have a similar correlation with prior visits as participants that have complete data. While this assumption is to be evaluated for missing data in the final analysis (as specified in the SAP), we do not want to include cases with missing data that is yet to be collected in the interim analysis where a decision might be made about stopping the trial.

As of the database snapshot, there were **86** participants that met the criteria for inclusion in the interim efficacy analysis, which is **54%** of the total 159 planned sample size. Using Lan-DeMets α-spending functions with O'Brien-Fleming boundaries in order to maintain the study-wise α level at 0.05 based on the 2 planned interim analyses, the revised alpha level for the 54% initial interim is **0.0046**. The alpha level at the other 2 timepoints is now 0.0179 at 75% and 0.0439 at the final analysis.

Pooled Site

The statistical efficacy analysis adjusts for pooled site. For the interim analysis, sites with fewer than 10 participants will be pooled together with another site such that the new pooled site is not larger than the largest non-pooled single site. Sites with the smallest enrollment numbers will all be pooled together, and sites with moderately small numbers will be pooled with sites that are geographically similar.

The pooled sites for the interim analysis are as follows:

- Pooled small sites (n=14): Cleveland Clinic (n=4), Johns Hopkins (n=4), UCSF (n=2), Utah (n=4)
- Pooled moderately small sites (Midwest) (n=19): Wisconsin (n=7) pooled with Mayo (n=12)
- Non-pooled sites: MGH (n=28), UNC (n=15), Penn (n=10).

Tables and Figures

The efficacy interim analysis will be submitted to the DSMB with a full regular safety DSMB report. The safety DSMB report tables and figures include all available data for all randomized participants. Because this analysis population is different, a few DSMB displays will be repeated for the interim analysis using the interim analysis MITT population.

The following displays will be included in the interim efficacy analysis based on the interim analysis MITT population:

- CONSORT diagram
- Demographics table
- ESS table, showing change from baseline at each visit C1-C6, the average of C4, C5, and C6 and at the post-treatment follow-up visit. Model estimates, confidence intervals and p-values of the differences between treatment groups for the change from baseline at each timepoint will be provided. The model will be fitted based on the model provided in SAP Section 9.3. The primary p-value to be assessed for early study stopping (relative to the significance level of 0.0046) is the comparison between treatment groups at the c6 (week 24) visit. P-values for all other timepoints are purely descriptive.
- Table of the key secondary outcomes, showing descriptive statistics by treatment group but no statistical testing for the following:
    - Amount of parenteral iron administered (in mg) during the 24-week treatment period (pro-rated to 24 weeks in patients who discontinue early).
    - Amount of packed red blood cell transfusions (in units) during the 24-week treatment period (pro-rated to 24 weeks in patients who discontinue early).
    - Change in Neuro-QoL Satisfaction with Social Roles and Activities Short Form (V1.1) T-score from baseline to weeks 12 and 24 (key timepoint), and the post-treatment follow-up visit.
    - Change in the HHT-specific QOL total score from baseline to weeks 12 and 24 (key timepoint), and the post-treatment follow-up visit.
    - Change in average weekly epistaxis duration from the four-week screening period prior to baseline to weeks 8-12 and to weeks 20-24 (key timepoint), and to weeks 1-4 post-treatment.

18 APPENDIX 3 INTERIM ANALYSIS AT 75%

The purpose of the 75% interim analysis is to assess for both futility and early efficacy of the primary efficacy outcome, the change from baseline in the ESS totals score at 24 weeks.

As stated in the protocol, the significance level of the primary outcome at the interim and final analysis will be based on Lan-DeMets α-spending functions with O'Brien-Fleming boundaries in order to maintain the study-wise α level at 0.05. The α level will be 0.0031 at 50%, 0.0183 at 75% and 0.0440 at the final analysis. The exact α level at the interim will be determined using Lan-DeMets α-spending functions based on the actual percent of data available at the interim.

The interim analysis will be reviewed by the DSMB in June 2023, based on data from a database snapshot date of **May 30, 2023**.

If the significance level of the primary outcome is less than the pre-specified alpha level, then the DSMB may recommend that study enrollment be halted, with currently active patients completing their treatment and follow-up, and the database being cleaned and locked for the final statistical analysis.

Study Population, Sample size and Alpha level

The interim efficacy analysis will be conducted on all randomized and treated participants (the modified Intent to Treat (MITT) population that will be used for the final statistical analysis as specified in the SAP) that have either completed the 24-week (C6) visit or have discontinued from the study. Participants that have completed C6 but have not yet completed the 4-week post-treatment follow-up are included.

Randomized participants that are ongoing in the trial and have not yet reached the C6 visit will not be included in the interim analysis. While their data collected as of the data snapshot is fully applicable, the repeated measures model used for the ESS 24-week analysis assumes that the data not yet collected for ongoing participants is missing at random and would be assumed to have a similar correlation with prior visits as participants that have complete data. While this assumption about missing data is to be evaluated in the final analysis (as specified in the SAP), we do not want to include cases with missing data that are yet to be collected in the interim analysis where a decision might be made about stopping the trial.

As of the database snapshot, there were **127 participants** that met the criteria for inclusion in the interim efficacy analysis, which is **80%** of the total 159 planned sample size. Using Lan-DeMets α-spending functions with O'Brien-Fleming boundaries in order to maintain the study-wise α level at 0.05 based on the 2 planned interim analyses, the revised alpha level for the 54% initial interim was 0.0046, and the revised alpha level is now **0.0230 at 80%** and **0.0425 at the final analysis**.

Futility

Futility for the primary efficacy outcome will be assessed by conditional power assuming the remaining outcome data to be collected will be similar to the data that has been collected thus far and not necessarily match the originally hypothesized difference. As stated in the protocol, the DSMB may recommend study stop due to futility if the upper limit of the 80% confidence interval for conditional power does not exceed 50%.

Conditional power will be calculated as follows:

$$ProbNorm\left\{\left[\sqrt{\frac{(1-P)}{P}}+\sqrt{\frac{P}{(1-P)}}\right]Z-\frac{Z_\alpha}{\sqrt{1-P}}\right\}$$

where $Z=\left(\hat{T}/SE\right)$

- probnorm = the **lower** 1-sided alpha value corresponding to the Z value of the formula in the curly brackets based on a normal distribution with mean of zero and standard deviation of 1, denoted N(0,1). We are using the lower tail because if the difference is calculated as placebo – pomalidomide, the Z will be negative if there is a treatment effect, corresponding to a greater decrease in the ESS score for pomalidomide than placebo
- p= the percentage of data available at the interim analysis **(=0.80**)
- Z_α = critical value for the one-sided alpha at the end of the study **(= -2.0286**, corresponding to Z for alpha = **0.0425/2**). If the difference is placebo – pomalidomide, the Z will be negative, corresponding to a greater decrease in the ESS score for pomalidomide than placebo
- $\hat{T}$ is the estimate of the treatment group difference (placebo-pomalidomide) from the MMRM model at C6
- SE is the standard error of the estimate of the treatment group difference from the MMRM model at C6

The 80% confidence interval for the conditional power is calculated as:

$$ProbNorm\left\{\sqrt{\frac{(1-P)}{P}}\times\left(Z\pm Z_\gamma\right)+\sqrt{\frac{P}{(1-P)}}\times Z-\frac{Z_\alpha}{\sqrt{(1-P)}}\right\}$$

Where Zγ is the Z value corresponding to the lower 1-sided tail of the N(0,1) distribution at the alpha level = (1-0.8)/2 = 0.10 = **-1.28155**

Pooled Site

The statistical efficacy analysis adjusts for pooled site. For the interim analysis, sites with fewer than 10 participants will be pooled together with another site such that the new pooled site is not larger than the largest non-pooled single site. Sites with the smallest enrollment numbers will all be pooled together, and sites with moderately small numbers will be pooled with sites that are geographically similar.

The pooled sites for the interim analysis are as follows:

- Pooled small sites (n=21): Cleveland Clinic (n=6), Johns Hopkins (n=6), UCSF (n=2), Utah (n=6), and UCSD (n=1)
- Non-pooled sites: MGH (n=43), UNC (n=20), Penn (n=15), Wisconsin (n=11), Mayo (n=17)

Tables and Figures

The efficacy interim analysis will be submitted to the DSMB with a full regular safety DSMB report. The safety DSMB report tables and figures include all available data for all randomized participants. Because this analysis population is different, a few DSMB displays will be repeated for the interim analysis using the interim analysis MITT population.

The following displays will be included in the interim efficacy analysis based on the interim analysis MITT population:

- CONSORT diagram
- Demographics table
- ESS table, showing change from baseline at each visit C1-C6, the average of C4, C5, and C6 and at the post-treatment follow-up visit. Model estimates, confidence intervals and p-values of the differences between treatment groups for the change from baseline at each timepoint will be provided. In addition, the conditional power and corresponding 80% confidence interval will be provided for the C6 timepoint. The model will be fitted based on the model provided in SAP Section 9.3.
    - The primary p-value to be assessed for early study stopping for efficacy (relative to the significance level of **0.0230**) is the comparison between treatment groups at the C6 (week 24) visit. P-values for all other timepoints are purely descriptive.
    - Stopping early for futility will be assessed using the upper bound of the 80% confidence interval for conditional power at the C6 (week 24) visit (relative to the outpoint level of 50%).
- Table of the key secondary outcomes, showing descriptive statistics by treatment group but no statistical testing for the following:
    - Amount of parenteral iron administered (in mg) during the 24-week treatment period (pro-rated to 24 weeks in patients who discontinue early).
    - Amount of packed red blood cell transfusions (in units) during the 24-week treatment period (pro-rated to 24 weeks in patients who discontinue early).
    - Change in Neuro-QoL Satisfaction with Social Roles and Activities Short Form (V1.1) T-score from baseline to weeks 12 and 24 (key timepoint), and the post-treatment follow-up visit.
    - Change in the HHT-specific QOL total score from baseline to weeks 12 and 24 (key timepoint), and the post-treatment follow-up visit.
    - Change in average weekly epistaxis duration from the four-week screening period prior to baseline to weeks 8-12 and to weeks 20-24 (key timepoint), and to weeks 1-4 post-treatment.