STATISTICAL ANALYSIS PLAN
STUDY PART 2

VERSION: 2.0

DATE OF PLAN:

18-SEPTEMBER-2019

BASED ON:

Protocol Version 9.0 (August 17, 2017)

STUDY DRUG:

RTA 408, OMAVELOXOLONE

PROTOCOL NUMBER:

408-C-1402, Part 2

STUDY TITLE:

A Phase 2 Study of the Safety, Efficacy, and Pharmacodynamics of RTA 408 in the Treatment of Friedreich's Ataxia

SPONSOR:

Reata Pharmaceuticals, Inc.

2801 Gateway Drive, Suite 150

Irving, Texas 75063

(972) 865-2219

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

SIGNATURE PAGE

This document has been prepared and/or reviewed by:

[redacted]

[redacted]
Signature

18 Sep 2019
Date

[redacted]

[redacted]
Signature

17 Sep 2019
Date

This document has been reviewed and accepted by:

[redacted]

[redacted]
Signature

18 Sept 2019
Date

[redacted]

[redacted]
Signature

18 Sep 2019
Date

REVISION HISTORY

Version	Revision Description	Date
2.0	• Specified that patients in the primary efficacy analysis population must have at least one post-baseline measurement • Changed analysis of the two key secondary endpoints to analysis of covariance (ANCOVA) and specified how to handle missing responses • Added tipping point and control-based multiple imputation as sensitivity analyses for the primary and key secondary endpoints to assess the impact of missing data • Updated power calculations to account for dropouts in sample size planning • Added "site" as a covariate for analysis of the primary and key secondary endpoints • Added analysis of efficacy using the "All Randomized" and "Pes Cavus" populations for descriptive purposes. Analyses of the "All Randomized" population includes a covariate for "pes cavus" • [redacted] [redacted] • Clarifications throughout	17 September 2019
1.1	Corrected minor typographical errors throughout the document.	04 June 2019
1.0	New document.	16 May 2019

TECHNICAL SUMMARY REPORT (TSR)

Name of Sponsor/Company **Reata Pharmaceuticals**	**Individual Study Table Referring to Part of the Dossier:** **Volume:**	***(For National Authority Use Only)*:**
Name of Finished Product: Omaveloxolone capsules	**Page:**	
Name of Active Ingredient: Omaveloxolone		

Title of Study: A Phase 2 Study of the Safety, Efficacy, and Pharmacodynamics of RTA 408 in the Treatment of Friedreich's Ataxia

Investigators: [redacted]

Study Center(s): 11

Studied period (years): 2	**Phase of development:** 2

Objectives:

Primary:

- To evaluate the change in the modified Friedreich's ataxia rating scale (mFARS) score at Week 48
- To evaluate the safety and tolerability of omaveloxolone

Key Secondary:

- To evaluate the Patient Global Impression of Change (PGIC) at Week 48
- To evaluate the Clinical Global Impression of Change (CGIC) at Week 48

Secondary:

- To evaluate the change in nine-hole peg test (9-HPT) at Week 48
- To evaluate the change in timed 25-foot timed walk test (T25-FWT) at Week 48
- To evaluate the frequency of falls
- To evaluate the change in peak work during maximal exercise testing at Week 48
- To evaluate the change in Activities of Daily Living (ADL) at Week 48

Exploratory:

- To evaluate the change in raters' assessments of videos of normal walking at Week 48
- To evaluate the change in SF-36 at Week 48
- To characterize the pharmacokinetics of omaveloxolone and potential metabolites after oral administration of omaveloxolone capsules

Methodology:

Part 2 of the MOXIe Phase 2 study is a randomized, placebo-controlled, double-blind, parallel-group study to evaluate the safety and efficacy of omaveloxolone 150 mg in patients with Friedreich's ataxia. Patients enrolled in Part 2 are randomized 1:1 to receive omaveloxolone 150 mg, or matching placebo. Randomization is stratified by pes cavus status (pes cavus vs. no pes cavus).

Patients with pes cavus have a musculoskeletal foot deformity and may represent a different subtype of FA, having different pathophysiology and clinical phenotype. Analysis of Part 1 data showed that treatment with omaveloxolone did not statistically improve studied endpoints (i.e., mFARS, exercise testing) in patients with pes cavus. Although the study part's small sample size limited its ability to detect a treatment effect, the presence of pes cavus also likely interferes with the ability to perform assessments that require standing or pedaling. Two of the four subsections that comprise the study's primary endpoint, mFARS, include assessments of lower limb coordination and upright stability. As a result, the primary analysis of efficacy is based on the stratum of patients enrolled without pes cavus. Because pes cavus is common in patients with FA, patients with pes cavus are also included in the study but will not comprise more than 20% of patients enrolled in Part 2. Additional secondary endpoints that are less likely to be affected by pes cavus are also assessed in the study to determine whether a therapeutic benefit can be detected in these patients. Efficacy endpoints are summarized descriptively for patients enrolled in the stratum with pes cavus. Safety is assessed for all patients enrolled in both strata.

Selection of the omaveloxolone dose of 150 mg for Part 2 was based on Data Safety Monitoring Board (DSMB) and Sponsor review of available data from Part 1, including safety, efficacy, and pharmacodynamic (PD) data. Following randomization on Day 1, patients self-administer study treatment once daily for 48 weeks. A follow-up visit for safety occurs at Week 52 (4 weeks after the last dose). The DSMB performs quarterly reviews of unblinded data for safety throughout Part 2.

Number of Subjects (planned):

Planned: 100 (with pes cavus n=20; without pes cavus n=80)

Diagnosis and main criteria for inclusion (see protocol section 8.1):

1. Have genetically confirmed Friedreich's ataxia
2. Have a mFARS score ≥ 20 and ≤ 80. The average of the two mFARS values collected at Screening and Day 1 visits must fall within the allowable range, and they must be within 4.5 points of each other
3. Be male or female and ≥ 16 years of age and ≤ 40 years of age
4. Have no changes to their exercise regimen within 30 days prior to Study Day 1 and be willing to remain on the same exercise regimen during the study period
5. Have the ability to complete maximal exercise testing
6. Have adequate kidney function defined as an estimated glomerular filtration rate (eGFR) ≥ 60 mL/min/1.73 m^2 using the Modification of Diet in Renal Disease (MDRD) 4 variable formula
7. Have a left ventricular ejection fraction ≥ 40% (based on echocardiogram performed at Screening Visit or within 90 days prior to Screening Visit)
8. Be able to swallow capsules
9. Be willing and able to cooperate with all aspects of the protocol
10. Be willing to practice medically acceptable methods of birth control
11. Provide written informed consent for study participation, approved by the appropriate Institutional Review Board (IRB)

Test product, dose and mode of administration:

Capsules containing omaveloxolone at the 50-mg strength are used in this study. Treatment kits containing three bottles are provided to patients for self-administration. Patients are instructed to self-administer 3 capsules of study drug orally once daily in the morning on an empty stomach (approximately 1 hour before or 2 hours after eating).

Duration of treatment: 48 weeks

Reference therapy, dose and mode of administration: Matching capsules containing placebo are used in this study. Treatment kits containing three bottles are provided to patients for self-administration. Patients are instructed to self-administer 3 capsules orally once daily in the morning on an empty stomach (approximately 1 hour before or 2 hours after eating).

Criteria for evaluation (see protocol section 11.2, 12.0):

Efficacy: mFARS; parameters collected during maximal exercise testing (including peak work), 9-HPT, T25-FWT, ADL, SF-36, PGIC, CGIC, fall diary, and videos of normal walking.

Safety: Results of echocardiogram, electrocardiogram, vital sign measurements, weight, body mass index, physical examinations, adverse events, serious adverse events, concomitant medications, and laboratory test results (clinical chemistry, hematology, urinalysis, microscopy, and pregnancy tests [as indicated]).

Statistical methods:

The primary analyses of efficacy in Part 2 of the study are based on the Full Analysis Set (FAS) that includes all patients enrolled without pes cavus. Analyses compare omaveloxolone patients to placebo patients. Mixed models repeated measures (MMRM) analysis is used to compare the difference in the change from baseline in the mFARS and peak work data at Week 48. The distribution of the Patient Global Impression of Change and Clinical Global Impression of Change is evaluated using analysis of covariance (ANCOVA) to compare treatment groups at Week 48. The body of this SAP describes the methods to be used for the other secondary and exploratory outcomes.

TABLE OF CONTENTS

LIST OF TABLES

1. LIST OF ABBREVIATIONS

Table 1: List of Abbreviations

Abbreviation	Term
μ_{OMAV}	Omaveloxolone average
$\mu_{PLACEBO}$	Placebo average
T25-FWT	25-Foot Timed Walk Test
9-HPT	9-hole peg test
ADL	Activities of Daily Living
AE	Adverse Event
ALT	alanine aminotransferase
ALP	Alkaline phosphatase
ANCOVA	Analysis of covariance
ARP	All randomized population
AST	aspartate aminotransferase
ATC	Anatomical/Therapeutic/Chemical
BMI	body mass index
BNP	B-type natriuretic peptide
CGIC	Clinical Global Impression of Change
CI	Confidence interval
CK	creatinine kinase
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
cm	Centimeters
CPK	Creatine phosphokinase
CRF	Case Report Form
CSR	Clinical Study Report
ECG	Electrocardiogram
DSMB	Data Safety Monitoring Board
eGFR	estimated glomerular filtration rate
f	truncation fraction
FA	Friedreich's Ataxia
FARS	Friedreich's Ataxia Rating Scale
FAS	Full analysis set
FDA	Food and Drug Administration

GGT	Gamma-Glutamyl Transferase
HDL-C	High-density lipoprotein cholesterol
HIV	Human immunodeficiency virus
HLT	high level term
ICH E9	International Conference on Harmonisation Tripartite Guideline for Good Clinical Practice E9
IRB	Institutional Review Board
IWRS	Interactive Web Response System
kg	Kilogram
LDH	Lactate dehydrogenase
LDL-C	Low-density lipoprotein cholesterol
LLD	Lower Limit of Detection
LS	least squares
LVMI	left ventricular mass index
MAR	missing at random
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
MCV	Mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities Terminology
MMRM	mixed model repeated measures
MNAR	Missing not at random
mFARS	modified Friedreich's Ataxia Rating Scale
MDRD	Modification of Diet in Renal Disease
NT-pro-BNP	N-terminal prohormone of B-type natriuretic peptide
PCP	Pes Cavus population
PD	Pharmacodynamic
PGIC	Patient Global Impression of Change
pH	potential of hydrogen
PK	Pharmacokinetic
PT	Preferred term
Q1	First quartile
Q3	Third quartile
QTc	corrected QT interval

QTcF	Fridericia corrected QT interval
RBC	Red blood cell
RTA 408	Omaveloxolone
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SD	Standard deviation
SE	Standard error
SF-36	SF-36® Health Survey Update
SOC	System Organ Class
TBL	Total bilirubin
TEAE	treatment emergent adverse events
TLF	Tables, listings, and figures
ULD	Upper Limit of Detection
ULN	Upper Limit of Normal
VLDL-C	Very-low-density lipoprotein cholesterol
WBC	White blood cell
WHO	World Health Organization

2. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and data displays to be included in the Clinical Study Report (CSR) for Protocol 408-C-1402, Part 2.

Protocol Revision Chronology:		
Protocol version 1	23-July-2014	Original
Amendment 1 (Protocol version 2)	11-Sept-2014	• The primary endpoint to evaluate peak workload (watts) was changed to evaluate peak work (watts/kg) • Part 1 study design was modified based on FDA feedback. The design was changed to a randomized, placebo-controlled, double-blind, dose-escalation study to evaluate the safety of omaveloxolone at 2.5 mg, 5 mg, and 10 mg
Amendment 2 (Protocol version 3)	13-Oct-2014	• Per FDA recommendation, PK characterization was added as an objective of the study
Amendment 3 (Protocol version 4)	12-May-2015	• Part 1 of the study was modified to allow for up to 2 additional cohorts of 8 patients each
Amendment 4 (Protocol version 5)	29-Sept-2015	• Part 1 of the study was modified to include 8 cohorts for dose ranging analysis
Amendment 5 (Protocol version 6)	23-May-2016	• Part 1 of the study included 9 cohorts and enrollment was increased to 108 patients. • The maximum permitted dose of omaveloxolone was raised to 300 mg
Amendment 6 (Protocol version 7)	13-Apr-2017	• Part 2 enrollment was increased to 100 patients • Part 2 study design was modified based on FDA feedback
Amendment 7 (Protocol version 8)	29-July-2017	• Patient Global Impression of Change and Clinical Global Impression of Change were added as secondary endpoints
Amendment 8 (Protocol version 9)	17-Aug-2017	• Part 2 study treatment duration was extended to 48 weeks

This SAP was developed in accordance with ICH E9 guidelines. All decisions regarding final analysis, as defined in this SAP document, were made prior to Database Lock (unblinding) of the study data. Further information can be found in the protocol. Changes to protocol specified analyses, including analysis populations, can be found in Section 12.

The statistical analysis plan (SAP) is based on:

- Protocol No. 408-C-1402, Version 9, dated August 17, 2017
- ICH guidelines E4 and E9 (Statistical Principles for Clinical Trials)
- Feedback received from FDA throughout clinical development, including the following communications:
  - FDA Meeting Minutes [redacted]
  - FDA teleconference with Reata on [redacted]
  - FDA Study May Proceed letter [redacted]
  - Email communication from the Division on [redacted]

This SAP describes the study populations, how variables are derived, how missing data are handled, and details concerning the statistical methods to be used to analyze the safety and efficacy data in Part 2 of study 408-C-1402. Should the SAP and the protocol be inconsistent with respect to the planned analyses, the language of the SAP is governing.

The SAP is finalized, approved by the Sponsor, and placed on file before the database is locked. This version of the SAP describes the analyses planned prior to the database lock. Unless otherwise specified, these analyses are summarized in the clinical study report (CSR). Any substantive changes made to the SAP after the database lock are clearly identified, and any analyses in addition to those specified in the SAP prior to the database lock are considered ad hoc. The CSR will describe any deviations from the planned analyses.

3. STUDY OBJECTIVES AND ENDPOINTS

3.1. Study Objectives

In patients with Friedreich's ataxia, the study will compare those receiving omaveloxolone versus those receiving placebo with respect to several objectives.

3.1.1. Primary Objectives

- To evaluate the change in the modified Friedreich's ataxia rating scale (mFARS) score at Week 48
- To evaluate the safety and tolerability of omaveloxolone

3.1.2. Key Secondary Objectives

- To evaluate the Patient Global Impression of Change (PGIC) at Week 48
- To evaluate the Clinical Global Impression of Change (CGIC) at Week 48

3.1.3. Secondary Objectives

- To evaluate the change in performance on a 9-hole peg test (9-HPT) at Week 48
- To evaluate the change in performance on a timed 25-foot timed walk test (T25-FWT) at Week 48
- To evaluate the frequency of falls over 48 weeks
- To evaluate the change in peak work during maximal exercise testing at Week 48
- To evaluate the change in the Activities of Daily Living (ADL) score at Week 48

3.1.4. Exploratory Objectives

- To evaluate the change in raters' assessments of videos of normal walking at Week 48
- To evaluate the change in SF-36® Health Survey Update (SF-36) score at Week 48
- To characterize the pharmacokinetics of omaveloxolone and potential metabolites after oral administration of omaveloxolone capsules

4. STUDY DESIGN

4.1. Summary of Study Design

Part 2 of the 408-C-1402 study is a randomized, placebo-controlled, double-blind, parallel-group study to evaluate the safety and efficacy of omaveloxolone 150 mg in patients with Friedreich's ataxia. Patients enrolled in Part 2 are randomized 1:1 to receive omaveloxolone 150 mg, or placebo. Randomization is stratified by pes cavus status: pes cavus and no pes cavus.

Patients with pes cavus have a musculoskeletal foot deformity and may represent a different subtype of FA, having different pathophysiology and clinical phenotype. Analysis of Part 1 data showed that treatment with omaveloxolone did not statistically improve studied endpoints (i.e., mFARs, exercise testing) in patients with pes cavus. Although the small sample size in Part 1 of the study limited the ability to detect a treatment effect, the presence of pes cavus also likely interferes with the ability to perform assessments that require standing or pedaling. Two of the four subsections that comprise the study's primary endpoint, mFARS, include assessments of lower limb coordination and upright stability. As a result, the primary analysis population for Part 2 efficacy is based on the stratum of patients enrolled without pes cavus (Section 6.3.1). Because pes cavus is common in patients with FA, patients with pes cavus are also included in the study but do not comprise more than 20% of patients enrolled in Part 2, and randomization was stratified by pes cavus status (with pes cavus vs without pes cavus). Secondary endpoints that are less likely to be affected by pes cavus are also assessed in the study to determine whether a therapeutic benefit can be detected in patients with pes cavus. All efficacy endpoints are summarized descriptively for patients enrolled in the stratum with pes cavus. Safety is assessed for all patients enrolled in both strata.

Selection of the omaveloxolone dose of 150 mg for Part 2 was based on Data Safety Monitoring Board (DSMB) and Sponsor review of available data from Part 1, including safety, efficacy, and pharmacodynamic (PD) data. Following randomization on Day 1, patients self-administer study treatment once daily for 48 weeks. A follow-up visit for safety occurs at Week 52 (4 weeks after the last dose). The DSMB performs quarterly reviews of unblinded data for safety throughout Part 2.

4.2. Definition of Study Drugs

Capsules containing omaveloxolone at the 50 mg strength, or the corresponding placebos, are used in this study.

4.3. Sample Size Considerations

4.3.1. Patients Without Pes Cavus

With 80 patients in the stratum without pes cavus, the Part 2 portion of the study has approximately 85% power to test the difference between the two treatment groups with respect to change from baseline in mFARS at Week 48. [redacted]

- [redacted]
- Two-sided Type I error rate of 0.05
- A difference of 2.0 points between the change in mFARS in the omaveloxolone and placebo groups
- Standard deviation of change in mFARS of 3.5 points
- [redacted]

4.3.2. Patients with Pes Cavus

[redacted]

4.4. Randomization

Patients enrolled in Part 2 are randomized 1:1 to either omaveloxolone or placebo. Randomization is stratified by pes cavus status (pes cavus vs. no pes cavus). Patients in the pes cavus stratum do not comprise more than 20% of patients enrolled in MOXIe Part 2. Randomization is generated using a centralized IWRS.

4.5. Clinical Assessments

All patients in Part 2 of the study follow the same visit and assessment schedule. Table 2 lists the overall schedule of assessments for the study. Following the first dose of study drug on Day 1, patients are scheduled for in-person visits to be assessed during treatment at Weeks 2, 4, 12, 18, 24, 36, and 48, as well as by telephone contact on Days 7, 56, 210, and 294. Patients are assessed at an in-person follow-up visit for safety at Week 52 (4 weeks after last dose).

Table 2: Schedule of Assessments

Visit Number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13	Visit 14
Study Day/Week	Screening	Day 1	Week 1 (telephone)	Week 2	Week 4	Week 8 (telephone)	Week 12	Week 18	Week 24	Week 30 (telephone)	Week 36	Week 42 (telephone)	Week 48 End of Treatment	Week 52 [a] End of Study/ 4-week follow-up
Day Relative to First Dose	-60 to -1 [b]	1 [c]	7 (±3 days)	14 (±3 days)	28 (±3 days)	56 (±3 days)	84 (±3 days)	126 (±3 days)	168 (±3 days)	210 (±3 days)	252 (±3 days)	294 (±3 days)	336 (±3 days)	192 (±3 days)
Informed consent	X													
Inclusion/Exclusion criteria assessment	X	X												
Demographics and baseline disease characteristics	X													
Pes cavus assessment	X													
Foot X-ray	X[d]													
Prior and concomitant medication assessment	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Medical history	X													
Height	X													
Echocardiogram	X[e]								X				X	
Electrocardiogram	X[e]			X	X		X	X	X		X		X	X
Vital sign measurements	X	X		X	X		X	X	X		X		X	X
Weight and BMI	X	X		X	X		X	X	X		X		X	X
Physical examination	X								X				X	X
Adverse event collection		X	X	X	X	X	X	X	X	X	X	X	X	X
Clinical chemistry	X[f]	X		X	X		X	X	X		X		X	X
Hematology	X[f]	X		X	X		X	X	X		X		X	X
Urinalysis and microscopy	X[f]	X		X	X		X	X	X		X		X	X
BNP and NT-proBNP[g]	X[f]	X		X	X		X	X	X		X		X	X
Hepatitis B and C and HIV[h]	X[f,h]													
Pregnancy test WOCBP[i]	X[f,i]	X		X	X		X	X	X		X		X	X
Exercise regimen reporting	X	X												
Randomization		X												
Study drug dispensation		X			X		X		X		X			
Study drug return and pill count / diary[j]					X		X		X		X		X	
Study drug administration[k]		←	----------	----------	----------	----------	----------	----------	X	----------	----------	----------	----------	→

Visit Number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13	Visit 14
Study Day/Week	Screening	Day 1	Week 1 (telephone)	Week 2	Week 4	Week 8 (telephone)	Week 12	Week 18	Week 24	Week 30 (telephone)	Week 36	Week 42 (telephone)	Week 48 End of Treatment	Week 52[a] End of Study/ 4-week follow-up
Day Relative to First Dose	-60 to -1[b]	1[c]	7 (±3 days)	14 (±3 days)	28 (±3 days)	56 (±3 days)	84 (±3 days)	126 (±3 days)	168 (±3 days)	210 (±3 days)	252 (±3 days)	294 (±3 days)	336 (±3 days)	192 (±3 days)
Maximal exercise test[l]	X	X			X		X	X	X		X		X	
Neurologic FARS (practice)	X[m]													
Neurologic FARS[n]	X[o]	X[o]			X		X	X	X		X		X	
9-hole peg test	X	X							X				X	
25-foot timed walk test	X	X[p]							X[p]				X[p]	
Video Collection of Normal Walking [q]		X							X				X	
SF-36 Health Survey Update		X							X				X	
Activities of Daily Living	X								X		X		X	
Patient Global Impression of Change							X[r]		X[r]		X[r]		X[r]	
Clinical Global Impression of Change							X[r]		X[r]		X[r]		X[r]	
Falls Diary	X[s]	←					X						→	
PK analysis				X[t]			X[u]		X[u]				X[u]	

[a] These procedures should also be performed in the event of early termination.
[b] Study Day -1 is the day prior to first dose of study drug.
[c] All Day 1 procedures should be performed prior to administration of first dose of study drug.
[d] Only 1 x-ray is collected during the study. Only x-rays of the right foot are completed for the foot x-ray assessment. Patients should complete the x-ray foot assessment at Screening; however, if the patient was not able to complete the x-ray foot assessment at Screening, it may be completed during any visit while the patient is participating in the study.
[e] For patients with echocardiograms and electrocardiograms collected within 90 days prior to the Screening Visit, the most recent echocardiogram and electrocardiogram can be used to assess patient eligibility.
[f] A home health nurse visit may be used to collect all lab samples required at the Screening Visit.
[g] Patients must be allowed to rest for a minimum period of 1 hour following maximal exercise test before this blood sample is collected (at the same time as all central lab blood draws). This sample should be taken with the patient in the same position (e.g., sitting or semi-recumbent) at all appropriate visits.
[h] Blood samples should be collected for hepatitis B and C and HIV antibodies only in patients lacking evidence of a negative titer in the past year.
[i] Negative serum pregnancy test results are required at the Screening Visit before study enrollment, and negative urine pregnancy test results are required at all other times indicated for continued participation in the study.
[j] A dosing diary check must be performed at study drug return and pill count.
[k] Study drug should be administered in the presence of study staff in the clinic on Day 1 after all Day 1 assessments have been completed. Study drug should also be administered in the clinic on Day 14 (Visit 4), Day 84 (Visit 7), Day 168 (Visit 9), and Day 336 (Visit 13) after the blood collection for predose PK analysis. All other doses can be administered at home. Study drug should be administered once daily through Day 336 (Visit 13).

[l] On study days where multiple assessments are to be completed, the maximal exercise test is the first functional assessment performed.
[m] The practice FARS assessment is not required for patients having completed a FARS assessment within 90 days prior to Screening. The practice FARS assessment must be performed prior to the screening maximal exercise test.
[n] Other than the practice FARS, all FARS assessments must always be performed after the maximal exercise test.
[o] Both Screening and Day 1 mFARS assessments are required to determine patient eligibility.
[p] Video of the 25-foot timed walk test is optional and is collected only for those patients who provide consent to video.
[q] Patients who consent to video collection of the 25-foot timed walk test also has video collection during normal walking with a 25-foot (un-timed) walk using his/her normal gait. The normal walk must be performed after the 25-foot timed walk test.
[r] Patients and investigators must complete the patient global impression of change and clinical global impression of change following completion of the neurological FARS exam.
[s] Patients are provided the fall diary at Screening and must record fall incidents between Screening and Week 48.
[t] Blood samples for PK analysis at Day 14 (Visit 4) should be collected prior to study drug administration as well as 1, 2, 4, and 8 hours after study drug administration.
[u] A single blood sample for PK analysis at Day 84 (Visit 7), Day 168 (Visit 9), and Day 336 (Visit 13) should be collected prior to study drug administration.

Abbreviations: BMI=body mass index; BNP= B-type natriuretic peptide; FARS=Friedreich's ataxia rating scale; HIV=human immunodeficiency virus; NT-proBNP= N-terminal prohormone of B-type natriuretic peptide; PK=pharmacokinetic; WOCBP=women of childbearing potential.

5. PLANNED ANALYSES

5.1. Final Analyses

The final analyses of efficacy are based on locked data and performed after all enrolled patients have completed all efficacy assessments (i.e., Week 48). The final analyses of safety are based on locked data and performed after all enrolled patients have completed all safety assessments (i.e., Week 52). The database lock plan describes details of the database lock.

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

The efficacy and safety analyses use the analysis sets defined in Section 6.3. Patient listings of all analysis data that support summary tables and figures are provided along with their source data. Measurements from patients excluded from the pre-defined analysis sets or extra measurements (such as unscheduled or repeat assessments) are not included in summary tables unless otherwise specified, but they are included in the patient listings. Missing data are not imputed, unless otherwise specified. In general, patient listings are sorted by patient number and assessment date (time and parameter, as applicable).

6.1. General Summary Table and Individual Subject Data Listing Considerations

Results of statistical analyses are reported using summary tables, listings, and figures (TLFs). All TLFs will use ICH numbering conventions. The following conventions are used:

- Unless otherwise noted, all statistical testing is two-sided and is performed at the 0.05 significance level.
- Tests are declared statistically significant if the calculated p-value is <0.05.

All analyses and summaries are produced using SAS® version 9.3 (or higher).

6.2. Data Presentation Conventions

Unless otherwise specified, descriptive statistics for continuous variables include the number of patients with data (N), mean, standard deviation (SD), quartiles (i.e., median, 25th, and 75th percentiles), minimum, and maximum. The same number of decimal places as in the observed value are presented when reporting minimum and maximum; 1 more decimal place than in the observed value is generally presented when reporting mean and quartiles; and 2 more decimal places than in the observed value are presented when reporting SD. When calculating the minimum and maximum of average baseline values, the same number of decimal places as in the observed value will be presented when reporting minimum and maximum. For imputed values and the average of baseline values, the average values will be rounded to the same number of decimal places as the source value.

Categorical (qualitative) data are presented using frequency counts and percentages. All percentages are rounded to 1 decimal place, unless otherwise specified. Percentages equal to 100 are presented as 100% and no percentages are presented for zero frequencies. Where individual variable values are missing, summaries of categorical data are based on reduced denominators (i.e., only patients with available data are included in the denominators) and the number of missing values is presented. For summaries of AEs and concomitant medications (CM), the percentages are based on the number of patients who received study drug.

6.3. Analysis Populations

Analysis populations defined in this section pertain to patients enrolled in Part 2 of the trial.

6.3.1. Full Analysis Set (FAS)

The primary analysis of efficacy is based on patients without pes cavus. The Full Analysis Set (FAS) includes patients enrolled without pes cavus who have at least one post-baseline measurement, categorized by their randomized treatment group (whether or not they received study drug).

6.3.2. All Randomized Population (ARP)

The All Randomized Population (ARP) includes all patients randomized, categorized by their randomized treatment group (whether or not they received study drug). Only descriptive analyses of efficacy are performed using the ARP.

6.3.3. Pes Cavus Population (PCP)

The Pes Cavus Population (PCP) includes all patients in the pes cavus stratum categorized by their randomized treatment group (whether or not they received study drug). Only descriptive analyses of efficacy will be performed using the PCP.

6.3.4. Safety Population

Safety analyses are based on all enrolled patients. The safety population includes all patients who received at least 1 dose of randomized study drug. The safety population is used for evaluation of safety variables. Patients who receive at least one dose of omaveloxolone are classified in the omaveloxolone group. Patients who receive at least one dose of placebo and no dose of omaveloxolone are classified in the placebo group.

6.3.5. Per-Protocol (PP) Population

A sensitivity analysis exploring the robustness of the primary FAS findings is based on the patients in the per-protocol population. The per-protocol population is defined as patients without pes cavus who:

- Received study drug through Week 48; and
- Had no major protocol deviation that could potentially affect the efficacy assessments.

6.4. Baseline Definition

Baseline values are defined as the last non-missing assessment prior to the first study drug administration, unless otherwise specified below. If the first study drug administration occurs after the date of randomization, the last measurement prior to the first study drug administration is considered the Day 1 measurement for the calculation of baseline.

6.4.1. Efficacy Assessments

6.4.1.1. mFARS

The protocol specifies that the Screening and Day 1 mFARS (Section 6.5.9) must be within 4.5 points to meet inclusion criteria for enrollment. If mFARS scores for Screening and Day 1 exceed the protocol allowed difference of 4.5 points for enrollment by at least two-fold (i.e., at

least 9 points different), then the Screening mFARS assessment is considered invalid, and the mFARS collected closest to randomization (i.e., Day 1) is used as the baseline.

Otherwise, the mean of Screening and Day 1 assessments is used as the baseline mFARS.

6.4.1.2. Other Efficacy Assessments

The mean of Screening and Day 1 assessments is used as baseline for continuous efficacy parameters: peak work, T25-FWT, 9HPT, and ADL.

6.4.2. Safety Assessments

Baseline for continuous safety assessments (i.e., vital sign assessments, weight, BMI, and laboratory measurements) is defined as the average value of measurements collected up through, but prior to, first study drug administration.

6.5. Derived and Transformed Data

6.5.1. Baseline Age

Subject's age in years is defined as the age at consent (Screen A).

- Age (year) = Floor ((date of consent – date of birth)/365.25)

6.5.2. Study Day

Study day is the day relative to the date of randomization. Day 1 is defined as the date of randomization, unless otherwise specified for the calculation of baseline (Section 6.4).

Assessments that occur after randomization but before the first study drug administration are considered to occur on study Day 1. Assessments collected on the same date as the first date of study drug administration will be considered to occur before the first dose of study drug administration.

For visits (or events) after randomization, day is calculated as:

- Study day = visit (or event) date - date of randomization + 1

For visits (or events) before randomization, day is calculated as:

- Study day = visit (or event) date - date of randomization

For listings (such as for adverse events) the quantity 'days since first (or last dose)' is defined as:

- days since first (or last dose) = event date – date of first (or last dose) + 1

For summaries that present distribution of time expressed in weeks and months, weeks will be defined as days divided by seven and months as days divided by 30.4.

6.5.3. Change from Baseline

Change from baseline is calculated using the baseline value (Section 6.4) and the value closest to the target study day, using the rules defined in Section 6.5.4.

6.5.4. Visit Windows

Because clinical visits may occur outside protocol-specified windows, instead of relying on visit labels in the clinical database, analysis visits and their windows are defined using derived study day (Section 6.5.2). Study day is calculated using the actual date for each scheduled and unscheduled assessment and compared to the target study day for each analysis visit. Data analysis and summaries are based on the collection date that is closest to the protocol scheduled target study day (Table 3).

Table 3: Analysis Visits

Analysis Visit	Label	Target Study Day	Analysis Window
2	Week 2	14	7 ≤ Study Day ≤ 21
4	Week 4	28	22 ≤ Study Day ≤ 55
12	Week 12	84	56 ≤ Study Day ≤ 104
18	Week 18	126	105 ≤ Study Day ≤ 146
24	Week 24	168	147 ≤ Study Day ≤ 210
36	Week 36	252	211 ≤ Study Day ≤ 294
48	Week 48	336	295 ≤ Study Day ≤ 350

The safety follow-up (Week 52) is based on days since last dose. If more than one assessment exists during the 4-week follow-up after last dose, the one closest to 28 days following the date of the last study drug administration is used for analysis and summary.

If a parameter is assessed or measured more than once within a visit window, the one that is closest to the protocol-scheduled time point (or target study day) is used for the purposes of data analysis and summary. If two assessments are equidistant from a target study day, the earlier assessment is used. If the visit used for analysis includes two assessments on the same day, the average of the two measurements will be used.

Records from visits not closest to the target study day, and therefore not used in analyses, are presented in by-subject data listings.

6.5.4.1. Off-Treatment Visit Windows

Off-treatment values for clinical laboratory evaluations (Section 10.4), vital signs (Section 10.5), and electrocardiogram (Section 10.8) will be summarized based on the date of last dose. 'Off-treatment' measurements are those that occur after the last dose date. The last assessment while on study drug is defined as the on-treatment result with the latest date. Table 4 indicates the windows used to attribute results to post-treatment timepoints.

Table 4: Analysis Visits for Post-Treatment Safety Analysis

Analysis Visit	Label	Target Study Day (Reference Day is Days After Last Dose)	Analysis Window
40	4-weeks – Off Treatment	28	15 ≤ Study Day ≤ 43
80	8-weeks – Off Treatment	56	44 ≤ Study Day ≤ 71
120	12-weeks– Off Treatment	84	72 ≤ Study Day ≤ 99
240	24-weeks– Off Treatment	168	100 ≤ Study Day ≤ 252
480	48-weeks– Off Treatment	336	253 ≤ Study Day ≤ 350

6.5.5. Years Since FA Onset

The years since FA onset are calculated as the difference between Baseline Age (Section 6.5.1) and Age at Onset.

6.5.6. GAA1 and GAA2 Repeat Length

Where GAA1 and GAA2 repeat length are entered as ranges, the value at the lowest end of the range is used. All GAA1 and GAA2 values are listed, including the range of GAA1 and GAA2 values when entered in the CRF.

6.5.7. Clinical Laboratory Results Outside the Limit of Quantification

Laboratory results reported as less than the lower limit of detection (i.e., <LLD) are imputed as LLD/2. Laboratory results reported as greater than the upper limit of detection (i.e., >ULD) are imputed as the ULD.

6.5.8. Estimated Glomerular Filtration Rate

The estimated glomerular filtration rate (eGFR) is calculated using the formula specified below.

For patients consented at age 18 years and older, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation is used:

- $eGFR\ (mL/min/1.73\ m^2) = 141 \times \min(S_{cr}/\kappa, 1)^{\alpha} \times \max(S_{cr}/\kappa, 1)^{-1.209} \times 0.993^{Age} \times 1.018$ [if female] × 1.159 [if black]

For patients consented at <18 years old, the Bedside Schwartz equation is used:

- $eGFR\ (mL/min/1.73\ m^2) = (0.41 \times \text{Height in cm}) / S_{cr}$

Where S_{cr} is serum creatinine (mg/dL), κ is 0.7 for females or 0.9 for males, and α is -0.329 for females or -0.411 for males. Min indicates the minimum of S_{cr}/κ and 1 and max indicates the maximum of S_{cr}/κ and 1. Age and height at screening are used for eGFR calculations.

6.5.9. mFARS

The Friedreich Ataxia Rating Scale (FARS) is a neurological-exam-based rating scale with five sections: Bulbar (section A), Upper Limb Coordination (section B), Lower Limb Coordination (section C), Peripheral Nervous System (section D), and Upright Stability (section E). Assessments in each section are added according to the logic in Table 5 as specified in the FARS manual (Lynch 2006). mFARS is the sum of sections A, B, C, and E.

Table 5: Scoring the FARS Neurologic Exam

Component Score of Neurologic FARS	Logic	Handling of Missing Data
A	Sum of all of the section A responses	If question is left blank but FARS assessment was completed, use question response from last visit
B	Sum of all of the section B responses	
C	Sum of all of the section C responses	
D	Sum of all of the section D responses, except 5a and 5b. For Question 5, only the results of Questions 5c1 and 5c2 is included in the section D sum.	
E	Average for items 2-5 item in the database; Reports sums for questions 1, 6, and 7 and sums of averages for questions 2-5	If questions 1-5 are left blank but FARS assessment was completed, take average of existing responses (e.g., average of 3a.1 and 3a.2). If no existing responses exist at the current visit for questions 1-5, use the question response from the last visit. For questions 6-7, use the response from the last visit.

6.5.10. 9-Hole Peg Test (9-HPT)

Both the dominant and non-dominant hands are tested twice at each assessment. The times recorded for the two trials for each hand are averaged, and the average value for each hand is used in analyses.

The average time value for each hand is transformed for analysis to allow for a more normal distribution of the data using the following methods:

- The number of pegs per second (pegs/second) for each hand is calculated based on nine pegs placed compared to the time needed to complete the test: 9 / (average time)
- The reciprocal of the average time is calculated as 1 / (average time) (Lynch 2005).

6.5.11. Timed 25-Foot Walk Test (T25-FWT)

Patients are instructed to attempt two T25-FWT trials at each visit. If both walk trials are completed, walk times are averaged for use in analyses. Otherwise, only the completed walk from each visit is used for analyses. The walk time is transformed for analysis to allow for a more normal distribution of the data using the following methods:

- The number of feet per second (feet/second) for each walk is calculated based on the distance (25 feet) and the total time needed to complete the test: 25 / (average time)
- The reciprocal of the walk time is calculated as 1 / (average time) (Lynch 2005).

6.5.12. Electrocardiogram Intervals

Electrocardiogram QTcF intervals are calculated from QTc and RR intervals using the following formula:

$$QTcF = QT/\sqrt[3]{RR}$$

where RR = 60 / (Heart Rate).

6.6. Handling of Missing Data

6.6.1. Missing Efficacy Endpoint Data

The primary analysis of efficacy is based on an assumption of missing at random (MAR). Missing mFARS data are not imputed for the MMRM analysis of the primary endpoint. Missing PGIC and CGIC data are imputed for the primary ANCOVA analysis of the key secondary endpoints using multiple imputation based on an assumption of MAR. A set of sensitivity analyses is included to assess the robustness of conclusions to the MAR assumption (see Section 8.5.2 and Section 8.6.3) for the analysis of the primary and key secondary endpoints.

6.6.2. Missing Start and Stop Dates for Concomitant Medication

Missing start dates for concomitant medications are not imputed.

Concomitant medications with incomplete end dates are considered concomitant medications if:

- Day and month are missing, and the year is equal to or after the year of the first date of study drug administration;
- Day is missing and the year is after the year of the first date of study drug administration;
- Day is missing and the year is equal to the year of the first date of study drug administration and the month is equal to or after the month of the first date of study drug administration; or
- Year is missing.

6.6.3. Missing Start and Stop Dates for Adverse Events

Treatment-emergent adverse events (TEAEs) are events that either:

- Had a date of onset on or after the date of the date of first dose of study drug and not more than 30 days after the date of the last dose of study drug, or
- Had no recorded date of onset with a stop date after the first dose of study drug, or
- Had no recorded date of onset or stop date.

Adverse events with incomplete start dates are considered after the date of first dose if:

- Day and month are missing, and the year is equal to or after the year of the first date of study drug administration;
- Day is missing and the year is after the year of the first date of study drug administration;
- Day is missing and the year is equal to the year of the first date of study drug administration and the month is equal to or after the month of the first date of study drug administration; or
- Year is missing.

Adverse events with incomplete start dates are considered on or within 30 days of last dose, if:

- Day and month are missing, and the year is equal to or before the year of the date of last dose of study drug plus 30 days;
- Day is missing and the year is equal to or before the year of the date of last dose of study drug plus 30 days, and month is equal to or before the month of the date of last dose of study drug plus 30 days;
- Year is missing.

7. STUDY POPULATIONS

Data are summarized for the Full Analysis Set (FAS), the All Randomized Population (ARP), and the Pes Cavus Population (PCP).

7.1. Subject Disposition

A disposition summary includes the number and percentage of patients in the following categories:

- Full Analysis Set (FAS)
- All Randomized Population (ARP)
- Pes Cavus Population (PCP)
- Safety population
- Per-protocol population
- Completed treatment through Week 48
- Discontinued treatment prior to Week 48
    - Reason for discontinuing treatment
- Completed study follow-up through Week 52
    - on treatment
    - discontinued treatment early but completed study visits
- Terminated early from the study
    - Reason for terminating study

A listing of disposition is provided for all enrolled patients.

7.2. Screen Failures

Screen failures are not summarized.

7.3. Protocol Deviations

Prior to database lock, a blinded team will identify all deviations, including major protocol deviations that could potentially affect the efficacy or safety conclusions of the study. Protocol deviations for excluded medications, patients who entered the study even though they did not satisfy the entry criteria, patients who received the wrong treatment or incorrect dose, and major protocol deviations are listed.

7.4. Demographic and Baseline Characteristics

Summaries of demographic and other baseline characteristics summaries are presented by treatment group for all analysis populations.

Demographic and other baseline characteristics include the following:

- Baseline Age, Age category (<18, ≥18)
- Sex, Race, Ethnicity
- Weight (kg), Height (cm), BMI (kg/m^2)
- Diastolic and systolic blood pressure (mmHg), Heart rate (bpm)
- Peak work (watts/kg), mFARS, each FARS subsection (i.e., A, B, C, D, E), ADL
- Age at FA onset, Years since FA onset
- Average number of hours exercised per week
- GAA1 repeat length, GAA2 repeat length, GAA1 repeat length ≥ 675 kbps
- X-ray measurements (calcaneal pitch, naviculocuboid overlap, talo calcaneal, talo first metatarsal, tibia calcaneal, tibia talar)
- Ambulatory status, Ambulatory assistive devices
- History of cardiomyopathy, areflexia, foot surgery, sensory neuropathy, swallowing difficulties, scoliosis, scoliosis surgery.

7.5. Listing of Subject Inclusion and Exclusion Criteria

A listing of enrolled patients who did not meet inclusion or exclusion criteria is generated.

7.6. Medical History

Medical history is summarized by treatment. Medical history is coded using MedDRA (Medical Dictionary for Regulatory Activities) version 14.1. Medical history items are summarized by MedDRA SOC and PT. Patient listings are also provided.

8. PRIMARY EFFICACY

Analyses of efficacy described in this section are the primary analyses of the efficacy endpoints, and pertain to the Full Analysis Set (i.e., patients without pes cavus). Analyses of efficacy described in this section will also be performed using the All Randomized and Pes Cavus Populations as exploratory analyses of efficacy for descriptive purposes. Analyses of efficacy using the All Randomized Population will include pes cavus (yes/no) as an additional covariate.

8.1. General Considerations

Analyses are performed for all omaveloxolone patients in comparison with all placebo patients. Summary statistics for observed values, change from baseline, and percent change from baseline (including 95% CI and quartiles) are presented by randomized treatment group.

8.2. Statement of the Null and Alternate Hypotheses

All efficacy endpoints compare patients randomized to omaveloxolone (μ_{OMAV}) to patients randomized to placebo ($\mu_{placebo}$). The primary and key secondary efficacy objectives will be evaluated according to the following hypotheses:

- H_0: $(\mu_{OMAV}) - (\mu_{placebo}) = 0$
- H_1: $(\mu_{OMAV}) - (\mu_{placebo}) \neq 0$

Because mFARS scores increase with disease progression, a decrease in mFARS score is considered evidence of benefit. Because PGIC and CGIC are assessed using a 7-point scale ranging from 1 (very much improved) to 7 (very much worse), a decrease in PGIC and CGIC is considered evidence of benefit.

8.3. Subgroup Analyses

The primary efficacy endpoint, change in mFARS score at Week 48, will be analyzed for subgroups of interest to assess homogeneity of the effect size for those subgroups that are sufficiently large and have an appropriate number of subjects to warrant these analyses. Key secondary (PGIC and CGIC) and secondary outcomes (9-HPT, T25-FWT, falls, peak work, and ADL) will also be analyzed for these subgroups. Additionally, safety analyses will be performed for these subgroups.

The following subgroup analyses will be tabulated:

- Age: <18; ≥18
- Sex: female; male
- Geographic location: US; Other
- Ethnicity: Non-Hispanic/Latino; Hispanic/Latino
- Race (White, Non-White)
- GAA1 Repeat Length ≥675: Yes, No

8.4. Multiple Comparisons and Multiplicity

This study has a single primary efficacy outcome: change in mFARS score at Week 48. If the study shows statistically significant evidence of benefit for this endpoint (change in mFARS score at Week 48), then the key secondary and secondary endpoints are analyzed using a hierarchical approach to maintain the family-wise overall Type I error rate of 0.05.

Each endpoint is tested at the 0.05 significance level. Formal testing in the hierarchy may proceed so long as statistically significant evidence of benefit continues to be shown. At such point in the hierarchy that one endpoint does not show statistically significant evidence of benefit, formal statistical testing of subsequent endpoints will not occur. Key secondary and secondary endpoints are tested in the following order:

1. PGIC at Week 48 (key secondary)
2. CGIC at Week 48 (key secondary)
3. Change in 9-HPT (reciprocal time measure of the non-dominant hand) at Week 48
4. Change in T25-FWT (reciprocal time measure) at Week 48
5. Frequency of falls over 48 weeks
6. Change in peak work during maximal exercise testing at Week 48
7. Change in ADL score at Week 48

All other objectives listed as "Exploratory" are presented with nominal significance levels for descriptive purposes only, including visits other than Week 48.

8.5. Primary Analysis of the Primary Efficacy Endpoint

8.5.1. Primary Efficacy Analysis

The primary endpoint of this study is the change in mFARS score (Section 6.5.9) at Week 48.

The mFARS scores for patients treated with omaveloxolone are compared with placebo at Week 48 using mixed models repeated measures (MMRM) analysis, with site (SITE) and baseline mFARS (BASE_MFARS) as covariates and the following fixed factors: treatment group (TRT), time (VISITNUM), the interaction between treatment and time (TRT*VISITNUM), the interaction between baseline and time (BASE_MFARS*VISITNUM) (code provided in Section 14.3). The analysis will use analysis visits 4, 12, 18, 24, 36, and 48 (Section 6.5.4). An unstructured covariance matrix is assumed.

In the event the MMRM model with an unstructured covariance structure does not converge, the following covariance structures are substituted, in the order listed. Each subsequent covariance structure is used only if each previous covariance structure is used and no previous model converged.

1. Heterogeneous Toeplitz covariance structure (assuming different variances at each time point and that measurements taken closer together in time are more highly correlated than those taken farther apart).

2. Toeplitz covariance structure (assuming measurements taken closer together in time are more highly correlated than those taken farther apart).
3. First order auto-regressive [AR(1)] covariance structure (assuming measurements taken closer together in time are more highly correlated than those taken farther apart, but more constrained than the Toeplitz structure).
4. Compound symmetry covariance structure (assuming equal correlation for measurements from a patient, regardless of how far apart in time when they were taken). [redacted]

The difference between omaveloxolone and placebo in change from baseline of mFARS will be estimated along with the 95% confidence interval at Week 48 for the primary analysis. The primary analysis of omaveloxolone vs placebo on mFARS score is at Week 48.

8.5.2. Sensitivity Analyses of the Primary Efficacy Results

Sensitivity analyses are included to assess the robustness of conclusions to the primary analysis of the primary efficacy endpoints (i.e., change in mFARS at Week 48), and to assess the assumption of MAR. The following sets of sensitivity analyses are summarized in tables and forest plots: tipping point with multiple imputation, treatment-based multiple imputation, control-based multiple imputation, and the MMRM model fit to the per-protocol population. Additional sensitivity analyses may be performed as appropriate. Sensitivity analyses are only performed in the FAS population (Section 6.3.1).

8.5.2.1. Tipping Point

The following tipping point sensitivity analysis will be performed to assess how severe departures from MAR must be in order to overturn conclusions from the primary analysis:

1. The PROC MI procedure in SAS will be used to generate 100 datasets satisfying the assumption of monotone missingness, imputing any missing mFARS values using multiple imputation. Imputation will be based on the non-missing observations within each treatment group.
2. The PROC MI procedure in SAS will be applied to each of the 100 datasets and all missing values in the placebo group will be imputed. These 100 datasets will then be used for the tipping point analysis.
3. For the tipping point analysis with a shift parameter, patients in the omaveloxolone group with a missing value will be assigned a shift parameter in the imputation procedure for progressively worse (higher) scores to find the point at which statistical significance is lost. Specifically, if the outcome of the hypothesis test favors omaveloxolone over the placebo group (P-value <0.05), then one point will be added as a shift parameter in the imputation for the omaveloxolone group, whilst the data for the placebo group will not be changed.
4. A mixed model repeated measures (MMRM) model (see Section 8.5.1 and Section 14.3) will be used to analyze each of the 100 completed data sets and PROC MIANALYZE will be used to combine these results to obtain the final estimates for the given value of the shift parameter.

5. This "mFARS shifting", as outlined in steps 3 and 4, will be repeated with one shift point at a time until the hypothesis test no longer rejects the null hypothesis in favor of the omaveloxolone group over the placebo group (i.e. when the P-value becomes greater than 0.05). Additional increments may be used to locate the tipping point. If the result is marginally significant, progressively lower scores will be added to find the point at which statistical significance is attained.

Further details of the tipping point analysis and associated code are provided in the Appendix (see Section 14.5).

8.5.2.2. Treatment-Based Multiple Imputation

As a sensitivity analysis, all missing Week 48 mFARS values are imputed with multiple imputation using the Week 48 data based on the randomized treatment group. The treatment-based sensitivity analysis is performed using the MMRM statistical model defined in Section 8.5.1.

8.5.2.3. Control-Based Multiple Imputation

As a sensitivity analysis, all missing Week 48 mFARS values are imputed with multiple imputation using the Week 48 data from the placebo group. The control-based sensitivity analysis is performed using the MMRM statistical model defined in Section 8.5.1.

8.5.2.4. Per-Protocol Population Analysis

The per-protocol population analysis will be performed in the per-protocol population using the MMRM statistical model defined in Section 8.5.1.

8.6. Primary Analysis of the Two Key Secondary Efficacy Endpoints

The following sections describe the primary analysis of the key secondary endpoints. Sensitivity analyses of key secondary endpoints may be performed as appropriate.

8.6.1. Patient Global Impression of Change

The PGIC is a 7-point scale that requires the patient to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of an intervention (Guy 1976). The PGIC is assessed by completing the following statement "since I began trial treatment, my overall status is: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), very much worse (7)." All PGIC values collected at each visit using the appropriate analysis windows (Section 6.5.4) will be used; the number of patients in each category is summarized at each analysis visit by treatment group.

The PGIC responses for patients treated with omaveloxolone are compared to placebo at Week 48 using analysis of covariance (ANCOVA), with treatment group and site as covariates (Section 14.8) and Week 48 value of PGIC as the outcome. Missing Week 48 data are imputed as defined in Section 14.9 using multiple imputation based on the treatment group to which the subject is assigned.

8.6.2. Clinical Global Impression of Change

The CGIC is a 7-point scale that requires the clinician to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of an intervention (Guy 1976). The CGIC is assessed by completing the following statement "Compared to the patient's condition at the start of the trial, this patient's overall status is: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), very much worse (7)." The number of patients in each category are summed at each analysis visit by treatment group.

CGIC is analyzed using the same methods specified for PGIC in Section 8.6.1 with the primary analysis at Week 48.

8.6.3. Sensitivity Analyses of the Key Secondary Efficacy Results

Sensitivity analyses are included to assess the robustness of conclusions to the primary analysis of the key secondary efficacy endpoints (i.e., PGIC and CGIC at Week 48), and to assess the assumption of MAR. The following sets of sensitivity analyses are summarized in tables and forest plots: tipping point with multiple imputation, control-based multiple imputation, and the ANCOVA model fit to the per-protocol population. Additional sensitivity analyses may be performed as appropriate. Sensitivity analyses are only performed in the FAS population (Section 6.3.1).

8.6.3.1. Tipping Point

The following tipping point sensitivity analysis will be performed to assess how severe missing data results must be in order to overturn conclusions from the key secondary analysis:

1. The PROC MI procedure in SAS will be used to generate 100 datasets imputing any missing values using multiple imputation
2. For the tipping point analysis with a shift parameter in the omaveloxolone group, patients in the omaveloxolone group with a missing value will be assigned a shift parameter in the imputation procedure for progressively worse (higher) scores to find the point at which statistical significance is lost. Specifically, if the outcome of the hypothesis test favors omaveloxolone over the placebo group (P-value <0.05), then one point will be added as a shift parameter in the imputation for the omaveloxolone group, whilst the data for the placebo group will not be changed.
3. A analysis of covariance (ANCOVA) model (see Section 14.8) will be used to analyze each of the 100 completed data sets and PROC MIANALYZE will be used to combine these results to obtain the final estimates for the given value of the shift parameter.
4. This "shifting", as outlined in steps 2 and 3, will be repeated with one shift point at a time until the hypothesis test no longer rejects the null hypothesis in favor of the omaveloxolone group over the placebo group (i.e. when the P-value becomes greater than 0.05). Additional increments may be used to locate the tipping point. If the result is marginally significant, progressively lower scores will be added to find the point at which statistical significance is attained.

Further details of the tipping point analysis and associated code are provided in the Appendix (see Section 14.10).

8.6.3.2. Control-Based Multiple Imputation

As a sensitivity analysis, all missing Week 48 PGIC and CGIC values are imputed with multiple imputation using the Week 48 data from the placebo group. The control-based sensitivity analysis is performed using the ANCOVA statistical model defined in Section 8.6.1.

8.6.3.3. Per-Protocol Population Analysis

The per-protocol population analysis will be performed in the per-protocol population using the ANCOVA statistical model defined in Section 8.6.1.

8.7. Primary Analysis of the Secondary Efficacy Endpoints

The following sections describe the primary analyses of the secondary endpoints. Sensitivity analyses of secondary endpoints may be performed as appropriate.

8.7.1. 9-Hole Peg Test (9-HPT)

The 9-HPT is a brief, standardized, quantitative test of upper extremity function. Two trials are attempted using each hand, and the time for the dominant and the non-dominant hand are analyzed separately. The average observed time for each hand is summarized descriptively at baseline, Week 24, and Week 48. Change from baseline by treatment is summarized for each hand at Weeks 24 and 48. Longer test times reflect more impairment on the patient's upper extremity function, thus a decrease from baseline suggests an improvement. Reciprocal time values and pegs/second values (Section 6.5.10) are also summarized at each time point along with change from baseline in these parameters. The primary analysis compares the change from baseline in the reciprocal time measure of the non-dominant hand for patients treated with omaveloxolone to patients treated with placebo at Week 48 with MMRM analysis using the statistical model defined in Section 8.5.1. The comparison of omaveloxolone with placebo is estimated using the difference in adjusted means and 95% CI for the difference in changes from baseline to Week 48. All other analyses of 9-HPT are considered exploratory and the nominal p-value will be reported for these analyses.

8.7.2. Timed 25-Foot Walk Test (T25-FWT)

The T25-FWT is a quantitative mobility and leg function performance test based on time in seconds to complete a 25-foot walk. The average observed time is summarized descriptively at baseline, Week 24, and Week 48. Change from baseline by treatment is summarized at Weeks 24 and 48. Longer test times reflect more impairment on the patient's ability to walk, thus a decrease from baseline suggests an improvement. Reciprocal time values and feet/second values (Section 6.5.11) are also summarized at each time point along with change from baseline in these parameters. The primary analysis is the change from baseline in the reciprocal time measure for patients treated with omaveloxolone compared to patients treated with placebo at Week 48 using MMRM analysis using the statistical model defined in Section 8.5.1. All other analyses of T25-FWT are considered exploratory and the nominal p-value will be reported for these analyses.

8.7.3. Frequency of Falls

Throughout Part 2 of the study, patients are instructed to record any instances of falls in a fall diary. A fall is defined as "the patient unintentionally coming to rest on the ground or at a lower level." Items included on the falls diary may include, but are not limited to, the date and time of each fall, the location of the fall, the preceding activity prior to the fall, the perceived cause of the fall, and if an injury was sustained after the fall. Patients are provided the fall diary at Screening and must record fall incidents between Screening and Week 48. The total number of falls before and after study drug administration is reported. The mean, quartiles (i.e., median, 25th, and 75th percentiles), and confidence interval of the total number of falls per patient recorded are summarized descriptively by treatment group. The number of falls by 12-week exposure interval (i.e., Day 1 to Week 12, Week 12 to Week 24, Week 24 to Week 36, and Week 36 to Week 48) overall and normalized by the number of patients per group, will also be summarized descriptively.

The primary analysis is the comparison of the total number of falls after randomization, summarized by computing the incidence rate of falls using a Poisson model fit with PROC GENMOD in SAS (Section 14.4). The comparison with placebo will be estimated from the Poisson model with the natural logarithm of time on study (days) included as an offset term; 95% confidence intervals of the incidence rates and the difference between those rates will be computed.

8.7.4. Peak Work

The change in peak work in (W/kg) is analyzed using an MMRM approach analogous to the method outlined for the primary efficacy endpoint (baseline peak work will be substituted for baseline mFARS as covariate and change from baseline in peak work will be the outcome) as described in Section 8.5.1. The comparison of omaveloxolone with placebo is estimated using the difference in adjusted means and 95% CI for the changes from baseline to Week 48.

8.7.5. Activities of Daily Living

The Activities of Daily Living scale examines the ability of patients to complete everyday tasks (e.g., holding a fork, dressing). The Activities of Daily Living is a 9-question assessment, with the total score being the sum of 9 questions. A score of 0 is considered normal, therefore, a decrease from baseline would be considered an improvement. If a question result is unanswered, then the total Activities of Daily Living score is considered missing. The number of patients with an unanswered ADL question will be presented. In addition, summaries of each Activities of Daily Living question are generated.

A mixed model (MMRM) will be fit with the change from baseline in ADL as the outcome and baseline ADL, treatment group, visit, treatment group by visit interaction, and baseline ADL by visit interaction as covariates. The model uses analysis visits 24, 36, and 48 weeks. The SAS code for this analysis can be found in Section 14.3. The comparison of omaveloxolone with placebo is estimated using the difference in adjusted means and 95% CI for the difference in changes from baseline to Week 48.

8.8. Analysis of the Exploratory Efficacy Endpoints

Summary statistics and 95% confidence intervals for treatment differences are provided for all exploratory endpoints as described below. P-values are provided for descriptive purposes only. Sensitivity analyses of exploratory endpoints may be performed as appropriate.

8.8.1. Videos of Normal Walking

Patients have the option to consent to have videos of normal walking recorded. Patients who consent to video collection will have video collection during normal walking with a 25-foot (un-timed) walk using his/her normal gait. The normal walk is not part of the T25-FWT, and must be performed after the 25-foot timed walk test.

Paired videos of normal walking at baseline to Week 24, and baseline to Week 48 (i.e., end-of-treatment) will be assessed by a movement specialist using the 7-point Likert scale to determine the clinical global impression of change (CGIC-w). CGIC-w values from the video analysis will be analyzed using the same methods specified in Section 8.6.1.

8.8.2. SF-36 Health Survey Update

The SF-36 is a multi-purpose, short-form health survey with 36 questions. It yields a profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures and a preference-based health utility index. The SF-36 is a generic measure, as opposed to one that targets a specific age, disease, or treatment group.

There are eight domains of the SF-36 as follows: Physical Functioning (items 3a-3j), Role-Physical (items 4a-4d), Role-Emotional (items 5a-5c), Social Functioning (items 6, 10), Bodily Pain (items 7, 8), Mental Health (items 9b, 9c, 9d, 9f, 9h), Vitality (items 9a, 9e, 9g, 9i) and General Health Perceptions (items 1, 11a-11d). Change in Health (item 2) is an overall health status assessment that compares to one year ago. The score is converted to a 0 - 100 scale, with 0 indicating poor health and 100 indicating good health (Section 14.14).

The MCS (mental component scale) and the PCS (physical component scale) as defined by the SF-36 manual are summarized at baseline and Week 48, along with the change from baseline by treatment. Lower scores reflect poor quality of life thus, a positive change from baseline suggests an improvement.

There are two assessments of SF-36 post-baseline, one at Week 24 and one at Week 48. A MMRM will be fit with change from baseline as the outcome and baseline MCS/PCS, treatment group, visit, treatment group by visit interaction, and baseline MCS/PCS by visit interaction as covariates. The SAS code for this analysis is similar to that provided in Section 14.3. The comparison of omaveloxolone with placebo is estimated using the difference in adjusted means and 95% CI for the difference in changes from baseline to Week 24 and baseline to Week 48.

9. EXPLORATORY EFFICACY

9.1. Exploratory Analyses of Efficacy Using the Full Analysis Set

Analyses of efficacy described in this section are exploratory analyses of the efficacy endpoints and are examined in all populations.

9.1.1. mFARS

Changes in mFARS for the omaveloxolone group versus the placebo group will be evaluated at Weeks 4, 12, 18, 24, and 36 using the statistical model defined in Section 8.5.1.

Changes in each FARS subsection score (Section 6.5.9) for the omaveloxolone group versus the placebo group will be evaluated at Weeks 4, 12, 18, 24, 36, and 48 using the statistical model defined in Section 8.5.1.

The proportion of patients at Week 48 with an mFARS improvement from baseline at or less than -1 at 1-point increments (e.g., ≤-1, ≤-2, ≤-3, ≤-4, ≤-5, ≤-6, ≤-7, ≤-8) at Week 48 are summarized by treatment group. Similarly, the proportion of patients at Week 48 with an mFARS worsening from baseline at or above 1 at 1-point increments (e.g., ≥1, ≥2, ≥3, ≥4, ≥5, ≥6, ≥7, ≥8) at Week 48 are summarized by treatment group.

[redacted]

9.1.2. Patient Global Impression of Change

[redacted]

To describe the PGIC over time, the proportion of patients having a PGIC response of "very much improved" or "much improved" (combined categories) are summarized at each visit. The treatments are compared descriptively by calculating the differences in proportions along with the 95% confidence intervals for those differences.

The correlation of PGIC response and change from baseline in mFARS at Week 48 will be calculated to describe the relationship between PGIC and the primary endpoint.

9.1.3. Clinical Global Impression of Change

Exploratory analyses specified for PGIC (Section 9.1.2) will be performed for CGIC.

9.1.4. 9-Hole Peg Test (9-HPT)

Changes in reciprocal 9-HPT time measure for the omaveloxolone group versus the placebo group will be evaluated for the non-dominant hand at Week 24 using the statistical model defined in Section 8.5.1.

[redacted]

9.1.5. Timed 25-Foot Walk Test (T25-FWT)

Changes in the reciprocal T25-FWT time measure for the omaveloxolone group versus the placebo group will be evaluated at Week 24 using the statistical model defined in Section 8.5.1.

[redacted]

9.1.6. Frequency of Falls

No exploratory statistical comparisons are planned for frequency of falls.

9.1.7. Maximal Exercise Test

Change from baseline in peak work (W/kg) for the omaveloxolone group versus the placebo group will be evaluated at Weeks 4, 12, 18, 24, and 36 using the statistical model defined in Section 8.5.1.

Change from baseline in other parameters collected during the maximal exercise test (Section 14.12) for the omaveloxolone group versus the placebo group will be summarized descriptively. For parameters measured at resting and at the end of the maximal exercise test (Table 27), the change from resting is also analyzed at each visit and for change from baseline.

[redacted]

9.1.8. Activities of Daily Living

Change from baseline in ADL for the omaveloxolone group versus the placebo group will be evaluated at Weeks 24 and 36 using the statistical model defined in Section 8.7.5.

[redacted]

9.2. Exploratory Analyses of Efficacy Using Other Analysis Populations

The primary analysis of efficacy is based on the full analysis set, which does not include patients with pes cavus. Because pes cavus is common in patients with FA, patients with pes cavus are also included in the study but do not comprise more than 20% of patients enrolled in Part 2. Secondary endpoints that are less likely to be affected by pes cavus are also assessed in the study to determine whether a therapeutic benefit can be detected in patients with pes cavus.

As indicated in Section 8, all primary, key secondary, secondary, and exploratory efficacy endpoints listed in Section 3.1 are summarized descriptively at each visit and change from baseline using the all randomized population (ARP) and pes cavus population (PCP). General considerations for data analysis (Section 6) are followed. Inferential testing performed using the ARP or PCP will be for descriptive purposes only.

10. SAFETY AND TOLERABILITY

Safety data (including AEs, laboratory data, vital signs, electrocardiogram [ECG] data, echocardiogram data, and physical examinations) are listed and summarized for patients in the safety population (Section 6.3.2). Safety data are summarized by treatment both overall and by pes cavus stratum.

The analysis visit windows (Section 6.5.4) are used for all safety analyses. Unscheduled visits that are outside the visit windows are listed and reflected in summaries of change to worst post-baseline measures where appropriate.

10.1. Adverse Event Preferred Term and Body/Organ System Summary Tables

AEs are summarized by treatment as defined by the safety analysis set. General considerations for AE summaries and calculations are:

- Multiple events by preferred term (PT) and system organ class (SOC) are counted once only per patient for each treatment.
- For summaries by severity, only the most severe event is counted per patient for each treatment.
- For summaries by relationship, for frequency counts by patient, only the most related event is counted per patient for each treatment.
- An AE with a missing resolution date or incomplete date that is not identified as continuing is assumed to be continuing and no duration is calculated.
- Only treatment-emergent adverse events (TEAEs) are included in summaries.

AEs are coded using MedDRA (Medical Dictionary for Regulatory Activities) version 14.1. In MedDRA, each verbatim term is mapped to a preferred term and high level term (HLT), which is then mapped to a system organ class. Tables and listings present data at the SOC and PT level.

Treatment-emergent adverse events (TEAEs) are events that either:

- Had a date of onset on or after the date of the date of first dose and not more than 30 days after the date of the last dose of study drug, or
- Had no recorded date of onset with a stop date after the first dose of study drug, or
- Had no recorded date of onset or stop date.

In addition, adverse events which occurred >30 days after the date of last dose of study drug are summarized as late-onset AEs (LOAEs). These are adverse events that had a date of onset more than 30 days after the date of the last dose of study drug.

The investigator grades the severity of the AEs as mild, moderate, or severe as defined in the study protocol.

Association or relatedness to the study medication is graded by the investigator according to criteria specified in Section 12.4 of the study protocol.

As defined in the protocol and captured on the CRF, a serious adverse event (SAE) is an adverse event that results in any of the following:

- Death;
- A life-threatening adverse drug experience;
- Inpatient hospitalization or prolongation of existing hospitalization;
- Results in persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions;
- Results in a congenital anomaly or birth defect in an offspring of a patient taking study drug;
- Is an important medical event.

10.1.1. Missing and Partial AE Onset Dates

Rules for handling missing partial AE Onset Dates are included in Section 6.6.3.

10.1.2. Summaries of Adverse Events for Serious Adverse Events (SAE), Adverse Event Dropouts, and Death

Treatment-emergent AEs are summarized by treatment at onset of the AEs. For each treatment, SOC, and PT, the number and percentage of patients reporting an event are calculated. In summary tables, SOC are presented alphabetically and events within SOC are presented by decreasing frequency count.

Summary tables (number and percentage of patients) of AEs (by SOC and PT) are provided by treatment as follows:

- All treatment-emergent AEs
- All treatment-emergent related AEs (definitely, probably, or possibly related)
- All treatment-emergent AEs by severity
- All treatment-emergent serious adverse events (including deaths)
- All treatment-emergent adverse events leading to discontinuation of study drug.

Listings are provided showing:

- All AEs
- Serious adverse events (including deaths)
- AEs leading to discontinuation of study drug.

10.2. Exposure and Compliance

The duration of study drug exposure is defined as the number of days on treatment from the first dose of study drug until the last dose of study drug (last dose – first dose + 1). Study drug exposure is summarized by descriptive statistics. Summaries include the total dose (mg) received (based on the number of pills returned), study drug compliance, the number and

percentage of patients receiving study drug by scheduled dispense visit, and duration (days) of exposure during the study treatment period.

Total number of doses dispensed and total dose (mg) dispensed are calculated from total number of kits (bottles) recorded on the Study Drug Dispensation eCRF. Total number of doses received is calculated from information on the eCRF of Study Drug Return and Study Drug Dispensation, as the total number of doses dispensed – total number of doses returned. Study drug compliance (%) is calculated as 100 × (total number of doses received) / (total number of doses dispensed). The proportion of patients who have ≥80% compliance are summarized.

In addition, the following exposure categories are summarized by number and percentage of patients:

- ≤16 Weeks
- >16 Weeks
- >24 Weeks

10.3. Concomitant and Other Medications

Concomitant medications are coded using the World Health Organization (WHO) drug dictionary (Enhanced version, September 2011, B2 format) for anatomical therapeutic chemical classification (ATC) and preferred drug name. A patient who used multiple medications is counted only once for each ATC and preferred drug name. ATC and preferred drug name within each ATC are sorted alphabetically. Coded concomitant medications are summarized by treatment. Percentages are based on the number of patients in the safety analysis set.

A concomitant medication is any medication taken at the time of first study treatment or a medication that was started after the start of study drug dosing. Specifically, concomitant medications are medications

- that are continued from screening and continued after the first study drug dosing, or
- that have start dates or stop dates within the treatment period.

Concomitant medications include those with an end date after the first study drug administration as well as medications without an end date. Medications with an end date on the date of first study drug administration are not to be considered concomitant medications.

In addition, patients who take excluded medications (defined in the Protocol Section 9.2.1) during the study are listed.

10.3.1. Missing and Partial Concomitant and Other Medication Start and Stop Dates

Missing and partial concomitant medication start and stop dates are detailed in Section 6.6.2.

10.4. Clinical Laboratory Evaluations

Laboratory data are summarized at baseline and at each time point by treatment. Only values obtained within analysis study windows (Section 6.5.4 and Section 6.5.4.1) are included in by-visit summary statistics. On-treatment values will be summarized according to the analysis study

windows in Section 6.5.4 and off-treatment values will be summarized according to the analysis study windows in Section 6.5.4.1.

10.4.1. Summaries of Laboratory Results

Selected laboratory evaluations and change from baseline are summarized by treatment, laboratory category (hematology, chemistry), test, and study visit using continuous statistics. Laboratory tests to be summarized are provided in Section 14.13. The estimated glomerular filtration rate (eGFR) results are calculated using formulas described in Section 6.5.8.

Line graphs of change from baseline are generated for selected laboratory tests, such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), ferritin, and creatine kinase (CK). Line graphs include mean ± SE over time for change from baseline.

Due to the nature of urinalysis parameters, summaries of continuous statistics are not provided. Urinalysis results (other than pH and specific gravity) of ketones, protein, blood, glucose, clarity, color, leukocytes, nitrite, bilirubin, and microscopic examinations (if indicated based on laboratory results), urine microscopic findings, and pregnancy test results are not summarized.

The number and percentage of patients with laboratory normality and abnormality categories at any time during the study (Normal, Low, High) as well as the treatment-emergent shift are summarized by treatment, laboratory category (hematology, chemistry, and urinalysis), and laboratory test. Shift tables summarizing Baseline, Last post-baseline, and shift to last post-baseline value are presented for lab parameters as appropriate.

10.4.2. Transaminases and Total Bilirubin

To assess the potential for drug induced liver injury, the number of patients meeting the follow criteria is summarized:

- ALT more than 3x, 5x, 10x or 20x ULN
- AST more than 3x, 5x, 10x or 20x ULN
- either ALT or AST more than 3x, 5x, 10x or 20x ULN
- either ALT or AST more than 5x ULN for more than two consecutive weeks
- alkaline phosphatase (ALP) >1.5x ULN
- TBL >2x ULN
- AST or ALT >3x ULN with an associated TBL>1.5x ULN
- AST or ALT >3xULN with an associated TBL>2x ULN.

A summary table that includes frequencies and percentages of patients that meet any of the above criteria at any time during the study are provided. A listing of subjects with abnormal ALT, AST, or TBL is also provided.

Moreover, the following shift tables of ALT, AST, and TBL are generated:

- Baseline to highest on-treatment;

- Baseline to follow-up (i.e., off-treatment); and
- Highest on-treatment to follow-up.

In addition, evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plots are generated for ALT and AST versus total bilirubin (TBL).

10.5. Vital Signs

Vital signs assessments include systolic blood pressure (SBP, mmHg) and diastolic blood pressure (DBP, mmHg), body temperature (°C), heart rate (HR, bpm), height (cm), weight (kg), and BMI (kg/m^2). Only values obtained within analysis study windows (Section 6.5.4 and Section 6.5.4.1) are included in by-visit summary statistics. On-treatment values will be summarized according to the analysis study windows in Section 6.5.4 and off-treatment values will be summarized according to the analysis study windows in Section 6.5.4.1.

Vital signs are summarized at baseline and at each time point along with the change from baseline by treatment. All data are listed.

10.6. Physical Exam

Physical exam results are listed.

10.7. Echocardiogram

Cardiac function [ejection fraction (%), left and right wall thickness (mm), septum thickness (mm) and left ventricular mass index (LVMI) (g/m^2)] measurements captured on echocardiogram are summarized at baseline and each time point along with the change from baseline.

10.8. Electrocardiogram

Electrocardiogram (ECG) data, such as clinical interpretation of ECGs, ventricular rate and interval assessments of PR, QRS, and QT are collected on the eCRF. QTcF is calculated (Section 6.5.12). Descriptive statistics for observed values and change from baseline at each time point are presented for these 12-lead ECG interval assessments. Only values obtained within analysis study windows (Section 6.5.4 and Section 6.5.4.1) are included in by-visit summary statistics. On-treatment values will be summarized according to the analysis study windows in Section 6.5.4 and off-treatment values will be summarized according to the analysis study windows in Section 6.5.4.1.

In addition, number and percentage of patients with any abnormal values at any time during the study (i.e., above a pre-specified threshold) are summarized by time point and overall while on study drug. The pre-specified levels of ECG QTc thresholds are consistent with FDA guidance.

Table 6: Pre-Specified Threshold Levels for ECG Parameters

ECG Parameter	Pre-Specified Level
PR	>200 msec

ECG Parameter	Pre-Specified Level
QTcF	>450, >480 or >500 msec, >30 or >60 msec increase from baseline
Heart rate	<40, >100 beats/min

10.9. Pregnancy

A listing is provided for serum and urine pregnancy results of all on-study pregnancies.

11. PHARMACOKINETICS

A separate document contains details regarding the planned pharmacokinetic analyses.

12. CHANGES FROM PROTOCOL-SPECIFIED ANALYSIS

This analysis plan specifies (Section 6.3) that the primary analysis of efficacy is based on the Full Analysis Set, which includes only patients without pes cavus. The protocol (Section 14.3.1) indicates that all enrolled patients are included in the ITT population for analysis of efficacy. The rationale for this change is based on insights from Part 1 analyses. Analysis of Part 1 data showed that treatment with omaveloxolone did not statistically improve studied endpoints (i.e., mFARs, exercise testing) in patients with pes cavus. Although the small sample size in Part 1 of the study limited the ability to detect a treatment effect, the presence of pes cavus also likely interferes with the ability to perform assessments that require standing or pedaling. Two of the four subsections that comprise the study's primary endpoint, mFARS, include assessments of lower limb coordination and upright stability. As a result, the primary analysis population for Part 2 efficacy is based on the stratum of patients enrolled without pes cavus (Section 6.3.1). Because pes cavus is common in patients with FA, patients with pes cavus are also included in the study but do not comprise more than 20% of patients enrolled in Part 2, and randomization was stratified by pes cavus status (with pes cavus vs without pes cavus). Secondary endpoints that are less likely to be affected by pes cavus are also assessed in the study to determine whether a therapeutic benefit can be detected in patients with pes cavus. This analysis plan specifies that patients having pes cavus are assessed descriptively for efficacy in the Pes Cavus Stratum.

The protocol sample size calculations (Section 14.1) were based on 100 patients (total enrollment) [redacted]. With the primary analysis of efficacy specified as being based on only those patients without pes cavus, this analysis plan includes power calculations for the 80 patients in the primary ITT population (Section 4.3.1). [redacted]

Additionally, this SAP changes the defined order of some efficacy endpoints. This analysis plan elevated PGIC and CGIC from secondary endpoints in the protocol to key secondary endpoints. Select protocol-specified exploratory endpoints (9-HPT, T25-FWT, frequency of falls, and ADL) are elevated to secondary endpoints in this analysis plan. The protocol did not specify change in raters' assessments of videos of normal walking as an endpoint, but this analysis plan added it as an exploratory endpoint.

The protocol includes only a high-level description of the analysis of the primary endpoint. No change is being made to that (a mixed model for repeated measurements), but this document sets out all the important details of how that analysis is done, including issues around handling missing data and planned sensitivity analyses.

This SAP specifies CKD-EPI and Bedside Schwartz as the eGFR formulas to be used for analyses of eGFR (Section 6.5.8). The protocol specified the Modification of Diet in Renal Disease (MDRD) 4 variable formula to calculate eGFR. This change aligns eGFR calculations across the sponsor's development programs.

13. REFERENCES

Guy, W. ECDEU Assessment Manual For Psychopharmacology, DHEW Publication No. ADM 76–338, US Government Printing Office, Washington, DC, USA, 1976.

Lynch DR, Farmer JM, Wilson RL, Balcer LJ.Performance measures in Friedreich ataxia: potential utility as clinical outcome tools. Mov Disord. 2005 Jul;20(7):777-82.

Lynch DR, Farmer JM, Tsou AY, Perlman S, Subramony SH, Gomez CM, Ashizawa T, Wilmot GR, Wilson RB, Balcer LJ. Measuring Friedreich ataxia: complementary features of examination and performance measures. Neurology. 2006 Jun 13;66(11):1711-6.

14. APPENDIX

14.1. Programming Specifications

The example code for programming specifications represent best approximations based on existing data structures and content. However, the specific SAS coding conventions may change depending on the structure and content of the data.

Continuous data are listed corresponding to the precision measured or calculated. Measures of central tendency are presented using one more decimal place than the precision of the data. Summaries of variability are presented using two significant digits more than the precision of the underlying data. Quartiles, the minimum, and the maximum are presented using the precision of the data. Rounding for derived values should be one additional decimal place to which the raw data were entered. For instance, if the raw data are rounded to the 0.001 decimal place, the derived values must be rounded to the 0.0001 decimal place.

All percentages are to be expressed as integers with one decimal place. The convention for rounding percentages is as follows:

- Values greater than or equal to x.x5% are rounded up
- Values between 0 and x.x5% are rounded down

Table 7: Parameter Names Used in Analysis

Description	Parameter
Baseline mFARS	BASE_MFARS
Change from Baseline mFARS	DELT_MFARS
Total Number of Falls	FALLS
Patient Number	PATNO
Pes Cavus (Yes/No)	PES_CAVUS
PGIC/CGIC Values	PGIC
Site	SITE
Treatment Group	TRT
Visit	VISIT
Visit Number	VISITNUM

14.2. Rationale for Programming Conventions

Assumption of No Monotone Missing for ANCOVA Analysis of PGIC/CGIC at Week 48

Since PGIC/CGIC results by visit week are not highly correlated within a patient, and because past subjective PGIC/CGIC results are not indicative of future results, no monotone missing datasets are generated to compute the Week 48 PGIC/CGIC results. In essence, past PGIC/CGIC results have no influence on the imputation of Week 48 PGIC/CGIC results, since PGIC/CGIC is a highly subjective measure.

Specification of A Single Seed Per Analysis (Table 26)

In order to both aid in validation and pre-define the seed values for multiple imputation analyses, a single seed is defined for each analysis. For tipping point analyses, only a single seed is used.

14.3. Mixed Model Repeated Measures Code

Sample SAS code for the primary analysis will be as follows:

Table 8: SAS code for MMRM analysis

```
proc mixed data=OUT2 method=reml alpha=0.05 covtest;
   class  VISITNUM(ref=LAST)
          TRT PATNO SITE;
   model CHG_MFARS =  BASE_MFARS TRT VISITNUM
         TRT*VISITNUM BASE_MFARS*VISITNUM
          SITE / solution ddfm=kr;
   repeated VISITNUM / subject=PATNO type=un r;
   lsmeans TRT*VISITNUM / diff om cl e;
   estimate 'estimate at Week 48' TRT 1 -1 VISITNUM*TRTGRP 0 0 0
    0 0 0 0 0 0 0 1 -1
        / CL;
   ods output estimates=SI_ESTIMATES;
run;
```

The pes cavus covariate is only utilized when analyzing results for the all randomized population (Section 6.3.2), as shown in Table 9.

[redacted]

The primary result will be obtained from the estimate statement and will be checked against the lsmeans statement to ensure that the correct result is obtained. This code will also be used for additional MMRM analyses such as that for peak work, SF-36, and ADL. The estimate statement will be adjusted to obtain results for Week 24 in models where the outcome is measured at baseline, Week 24 and Week 48.

14.4. Code for Frequency of Falls

The following SAS code serves as sample code to provide the estimated differences, confidence intervals, and p-value. A Poisson model is fit to the data with the natural logarithm of time (days) on study included as an offset term and an option to account for potential over dispersion (dscale). The test of the statistical significance of the coefficient associated with treatment group membership will test the null hypothesis of no differences in the number of falls between the two groups.

The pes cavus covariate is only utilized when analyzing results for the all randomized population (Section 6.3.2).

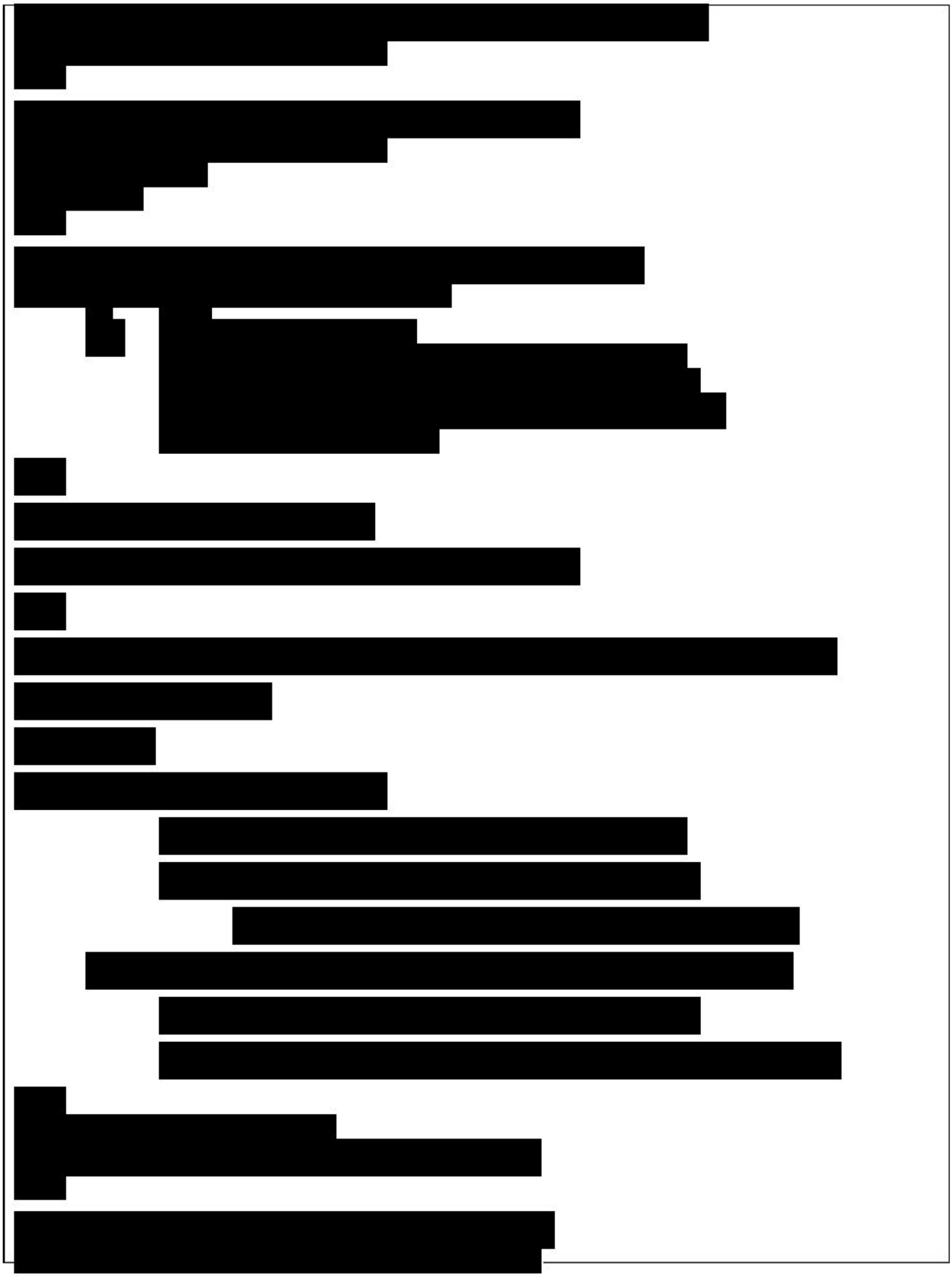
14.5. Treatment-based imputation and analysis of the primary endpoint

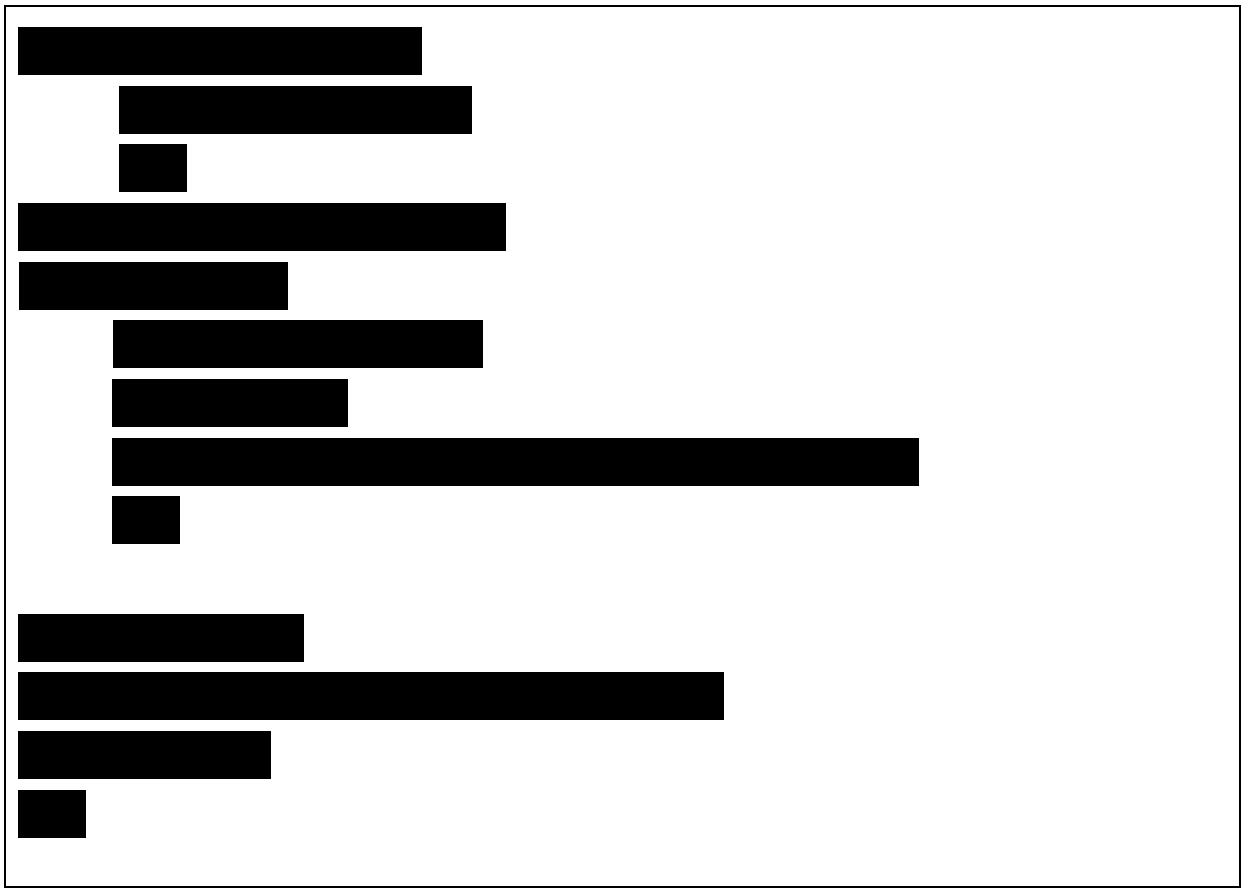
This section provides details and sample SAS programming code for the treatment-based imputation of the primary endpoint analysis outline in Section 8.5.2.

The treatment-based imputation of the primary endpoint will use a single seed from Section 14.11.

The dataset DATA_IMP is the transposed input dataset, with columns for changes in mFARS at Weeks 1-48. The OUT_IMP_CONT dataset and the code described in Table 10 creates a monotone missing dataset with columns for change from baseline by analysis study week.

Table 10: SAS code for treatment-based imputation of the primary endpoint





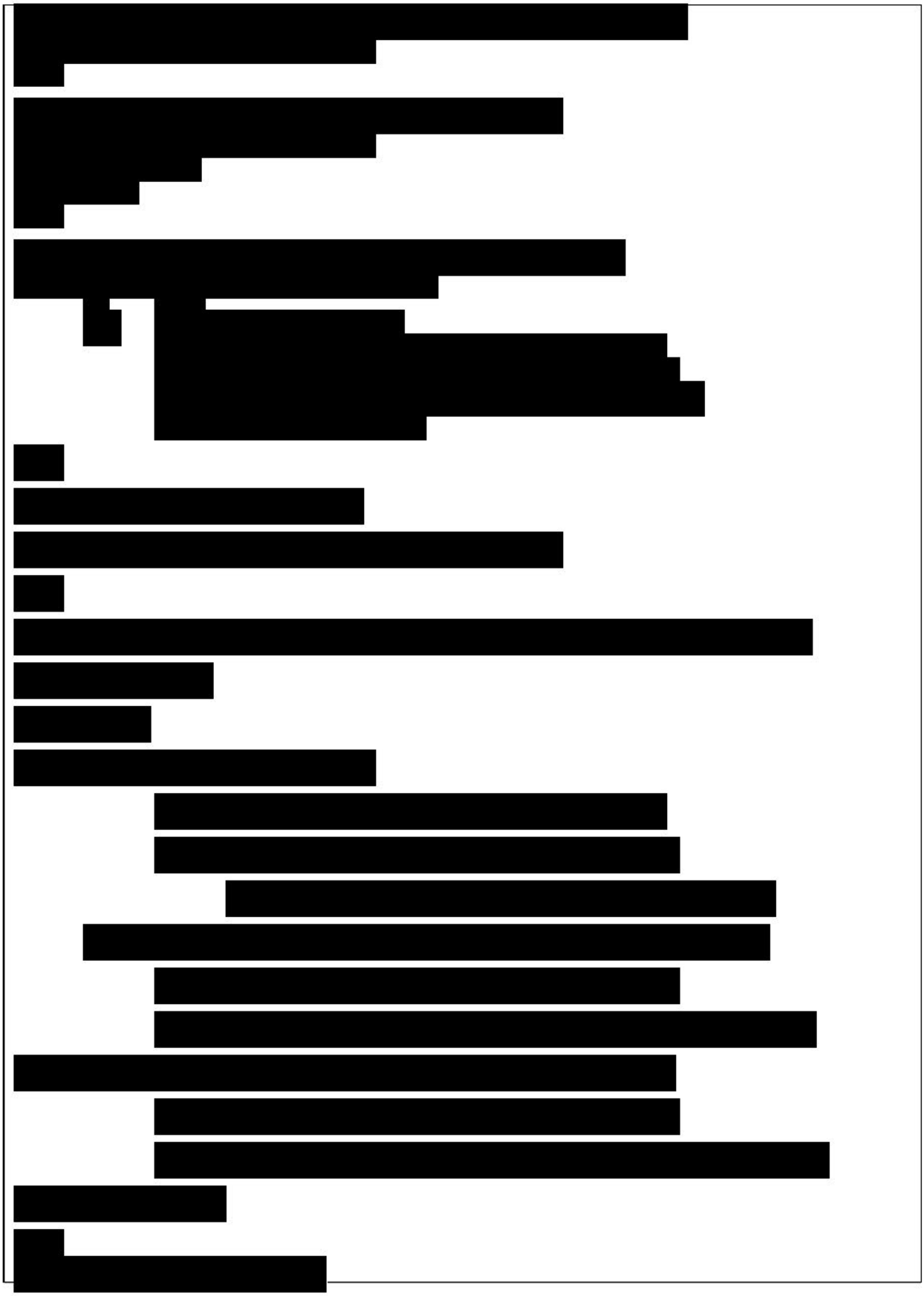
14.6. Control-based imputation and analysis of the primary endpoint

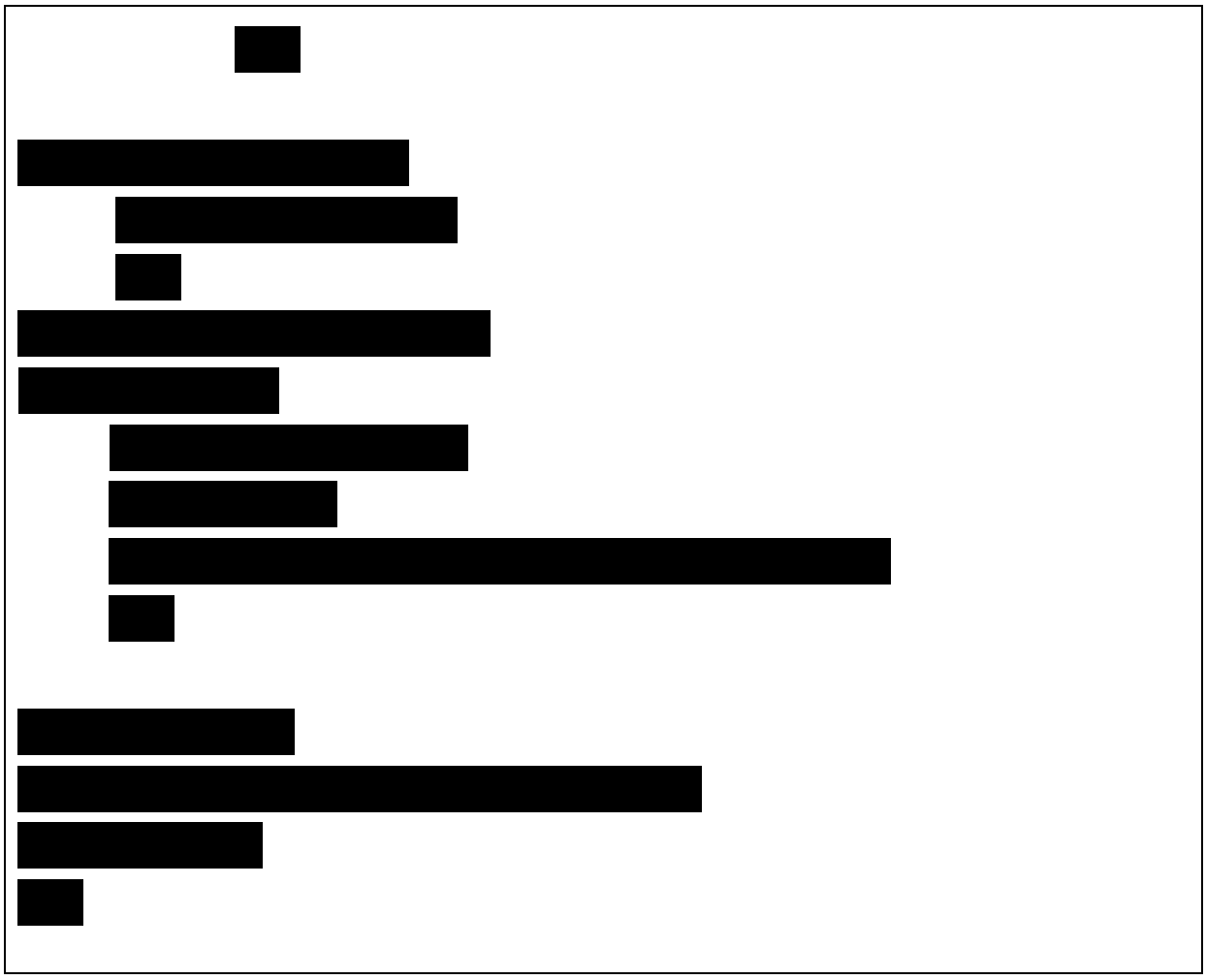
This section provides details and sample SAS programming code for the control-based imputation of the primary endpoint analysis outline in Section 8.5.2.

The control-based imputation of the primary endpoint will use a single seed from Section 14.11.

The dataset DATA_IMP is the transposed input dataset, with columns for changes in mFARS at Weeks 1-48. The OUT_IMP_CONT dataset and the code described in Table 11 creates a monotone missing dataset with columns for change from baseline by analysis study week.

Table 11: SAS code for control-based imputation of the primary endpoint





14.7. Tipping point analysis of primary endpoint

This section provides details and sample SAS programming code for the tipping point analysis outline in Section 8.5.2.1 of the SAP.

The tipping point analysis assesses the effect of potential deviations from the assumption of MAR and explores the consequences of assuming data in the omaveloxolone arm are missing not at random (MNAR) (i.e., subjects in the omaveloxolone arm with missing outcome are assumed to have lower mFARS values than subjects in the placebo arm). This analysis is based on the assumptions of monotone missingness using the regression approach and that the trajectory of mFARS changes once the subject stops use of omaveloxolone. The specific steps are as follows:

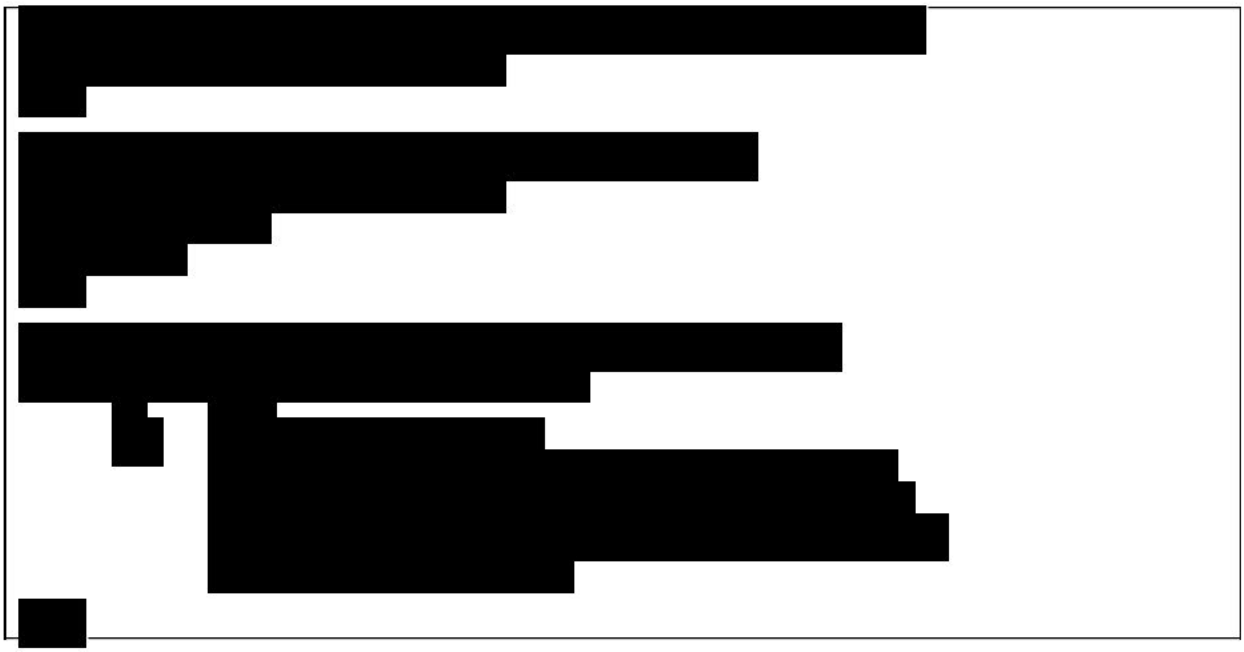
1. The missing information on mFARS prior to subject discontinuation will be imputed using the monotone method with missing data imputed based on their assigned treatment group. This step will create 100 data sets that satisfy the assumption of monotone missingness. The code for this step is provided in Table 12.

2. The missing information on mFARS will be imputed in the placebo group only for all remaining values. This is the data set that will be used for Step 3 in all analyses.
3. The remaining missing data for mFARS in the omaveloxolone arm will be multiply imputed using a shift parameter S, where S can change in both directions progressively to find the point at which statistical significance is lost or achieved. The sample SAS code for performing multiple imputation with a shift S in the omaveloxolone arm is presented in Table 13. Missing data in the placebo arm are imputed assuming MAR, using the code in Table 13 subset to the placebo subjects.
4. The multiply-imputed data will be analysed using Rubin's rule to combine the results and to obtain an estimated treatment effect and its associated significance level.
5. Steps 3 and 4 will then be repeated. The shift parameter S starts from 0, which corresponds to the primary efficacy analysis with no shift effect, and the effect potentially increased in both a negative and positive direction by a certain amount in each step until the analysis reaches the "tipping point", the point at which the effect of omaveloxolone is no longer superior to that of placebo. The more the tipping point diverges from the observed data, the more robust the conclusion based on primary efficacy analysis.

The tipping point analysis will use a single seed from Section 14.11.

The dataset DATA_IMP is the transposed input dataset, with columns for changes in mFARS at Weeks 1-48. The OUT_IMP_CONT dataset and the code described in Table 12 creates a monotone missing dataset with columns for change from baseline by analysis study week.

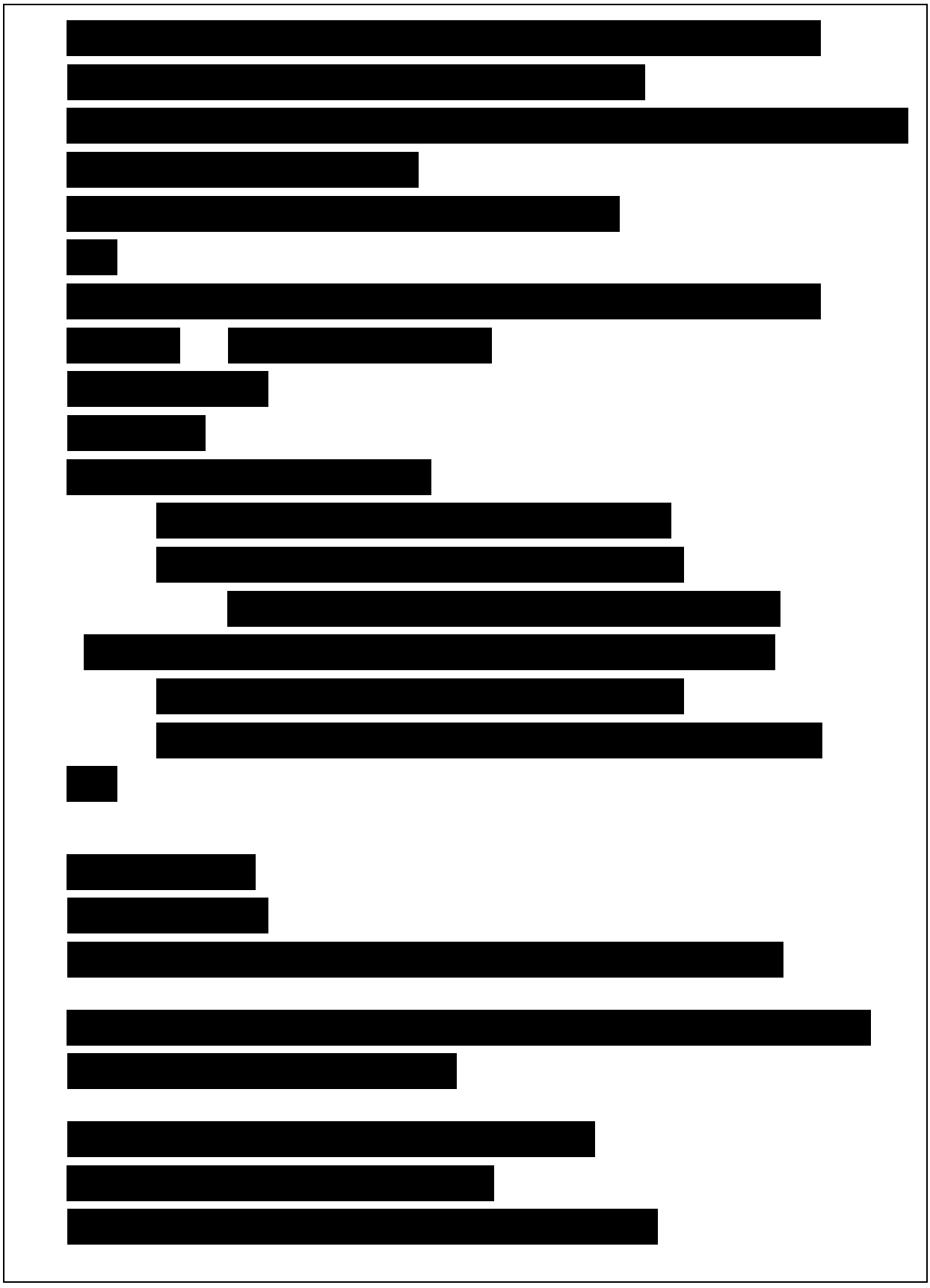
Table 12: SAS code to create datasets satisfying monotone missing condition for continuous measures

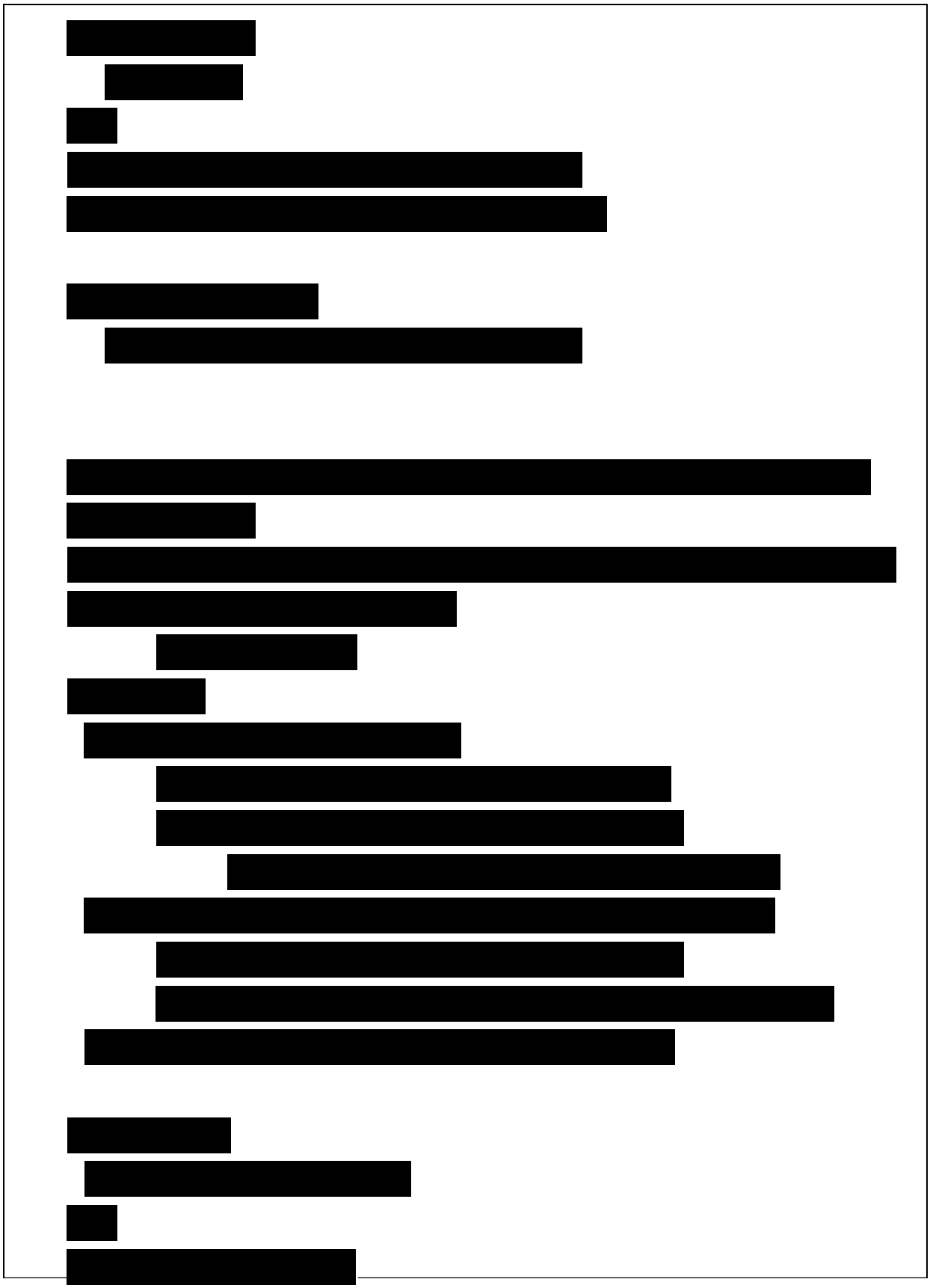


The code to impute the data sets for the tipping point analysis is in Table 13 below. Note that a single group of 100 datasets is created from this analysis. This set of 100 datasets of imputed values for the placebo data is used for the full tipping point analysis. Note that the code used to impute the placebo data and create the dataset IMPDAT_PBO only needs to be executed once. After each shift dataset is created, each of these sets of imputed datasets are then analysed using PROC MIXED and combined using PROC MIANALYZE.

For each shift value, the code in Table 14 and Table 15 is also executed within the %do loop and macro shown inTable 13. For simplicity, Table 14 and Table 15 show the mixed model code and analysis of results for a single shift parameter. Once the code from Table 15 is executed within the %do loop in Table 13 for a single shift parameter, the results for that parameter are output using the &outparms variable for a particular shift value.

Table 13: SAS code for tipping point analysis



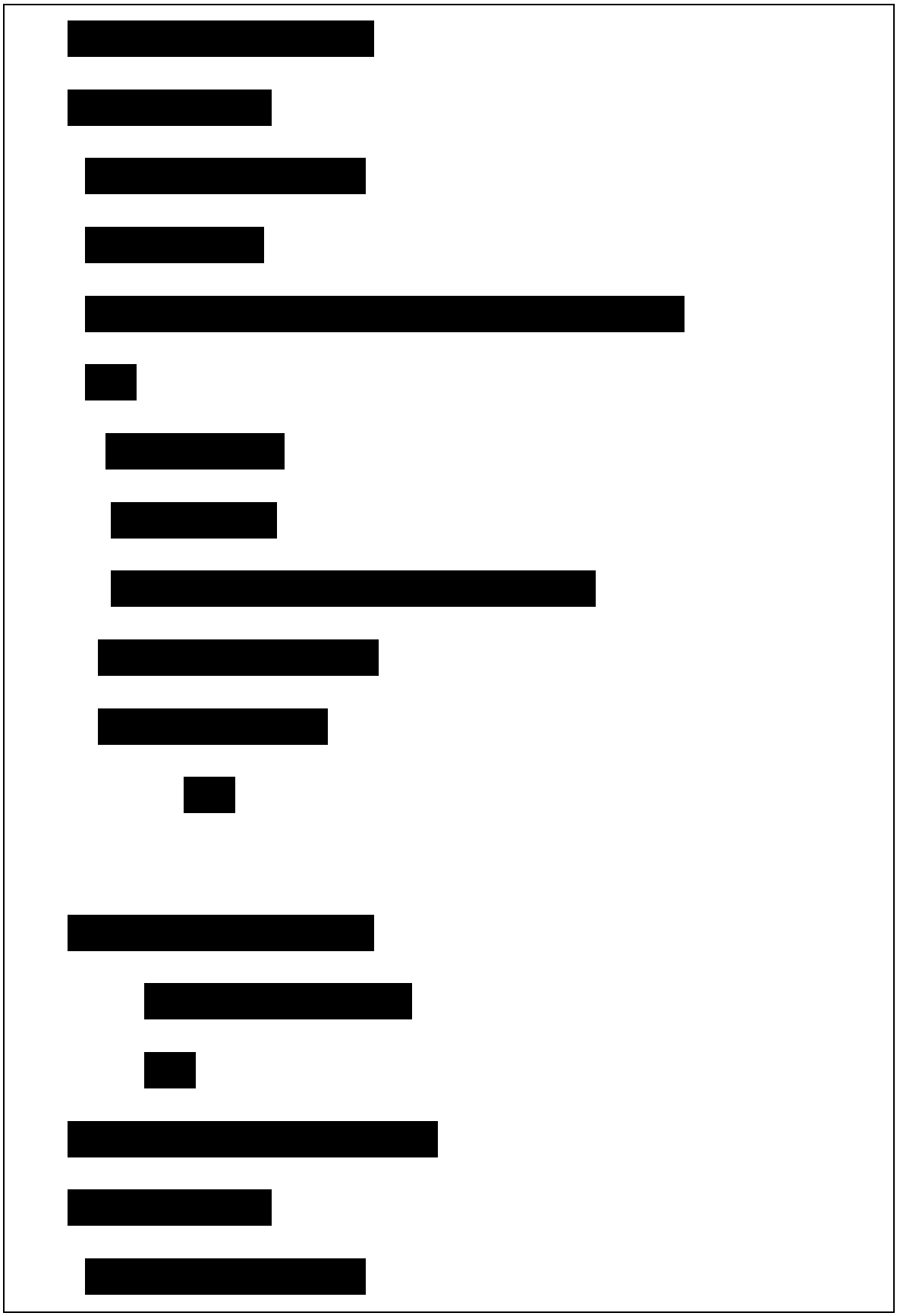


The example code below fits a mixed model to each of the sets of 100 imputed datasets created from the code in Table 13. Note that the number of datasets to be analyzed depends on the number of sensitivity parameters used.

Table 14: SAS code for MMRM analysis

The input data set "COMBINE"" includes the imputed data for the treatment and placebo groups from Table 13. This dataset is then input into PROC MIANALYZE to obtain the final result for each sensitivity parameter.

Table 15: SAS code to obtain the overall estimate and standard error

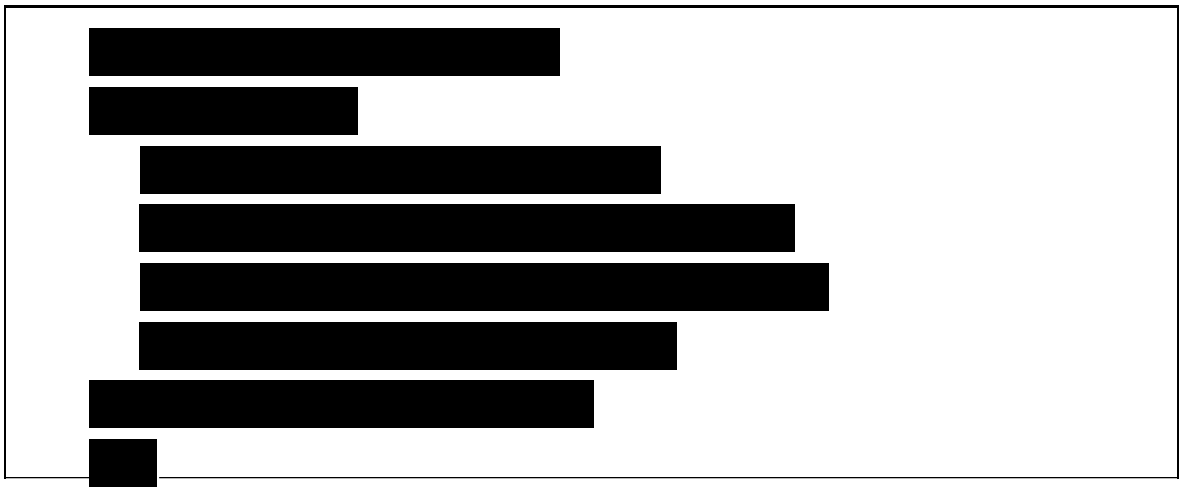


14.8. ANCOVA analysis of the secondary endpoint

The key secondary endpoint of change from baseline in PGIC/CGIC at Week 48 is based on an ANCOVA model. The code for this model is provided in: Table 16.

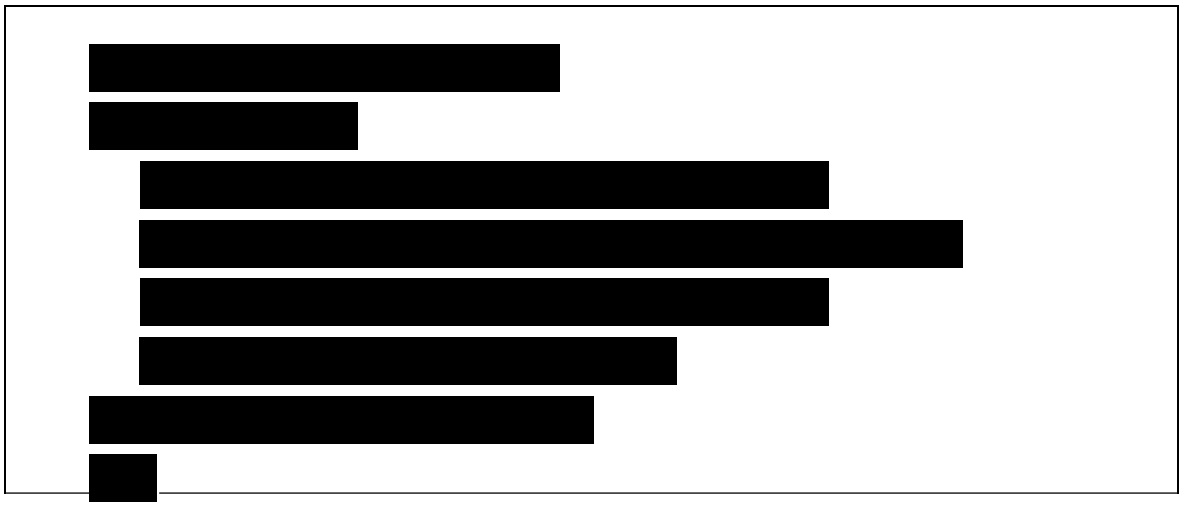
The test of the null hypothesis of no difference between groups in the change from baseline at Week 48 is obtained from the lsmeans statement.

Table 16: SAS code for PGIC/GCIC ANCOVA



The pes cavus covariate is only utilized when analyzing results for the all randomized population (Section 6.3.2), as shown in Table 17.

Table 17: SAS code for PGIC/GCIC ANCOVA in the All Randomized Population

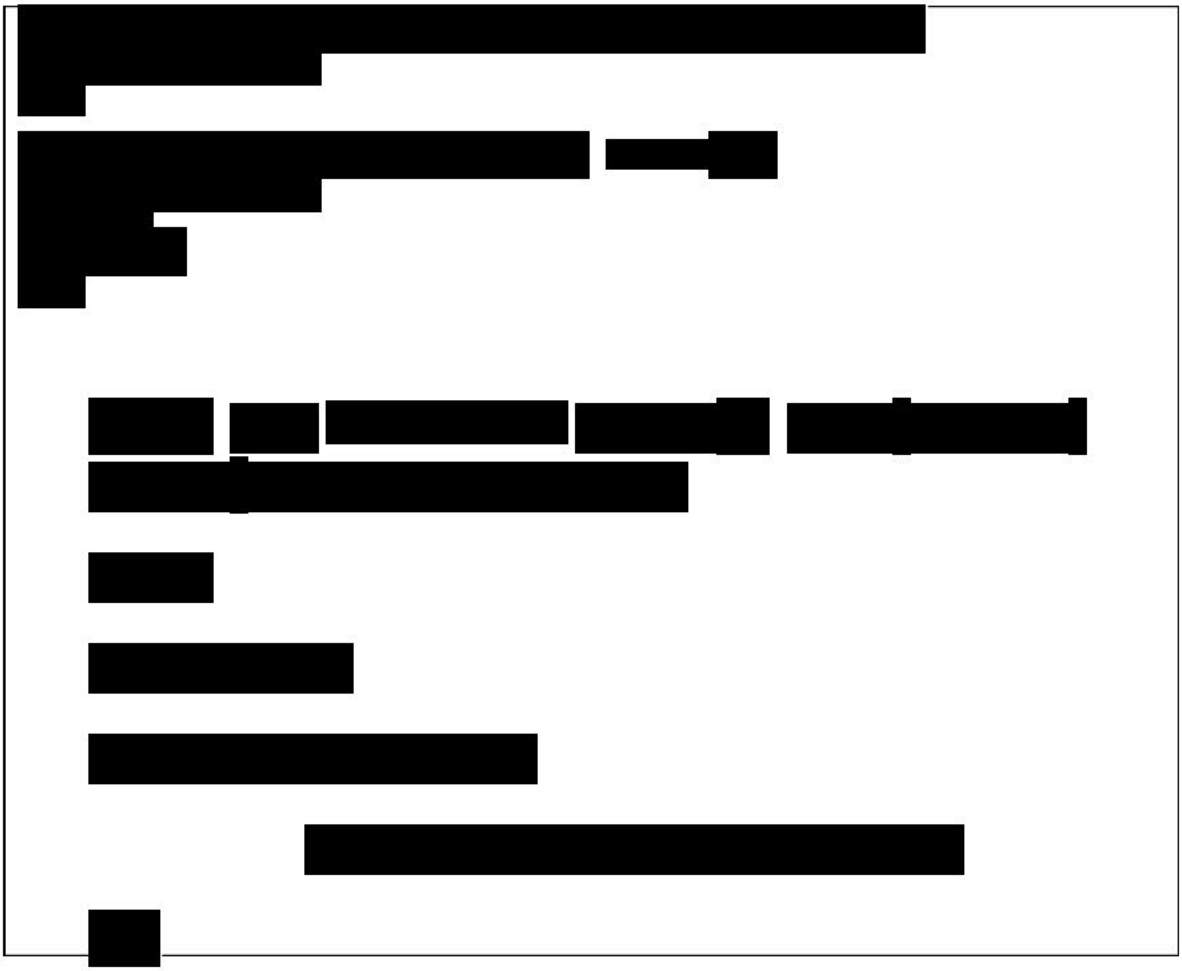


14.9. Missing data approaches for secondary endpoint analysis

The ANCOVA model presented in Table 16 includes the Week 48 timepoints only. All missing values are imputed based on the treatment group to which the patients were randomized. The code for where missing values are imputed based on the treatment group the subject was randomized to, is provided in Table 18.

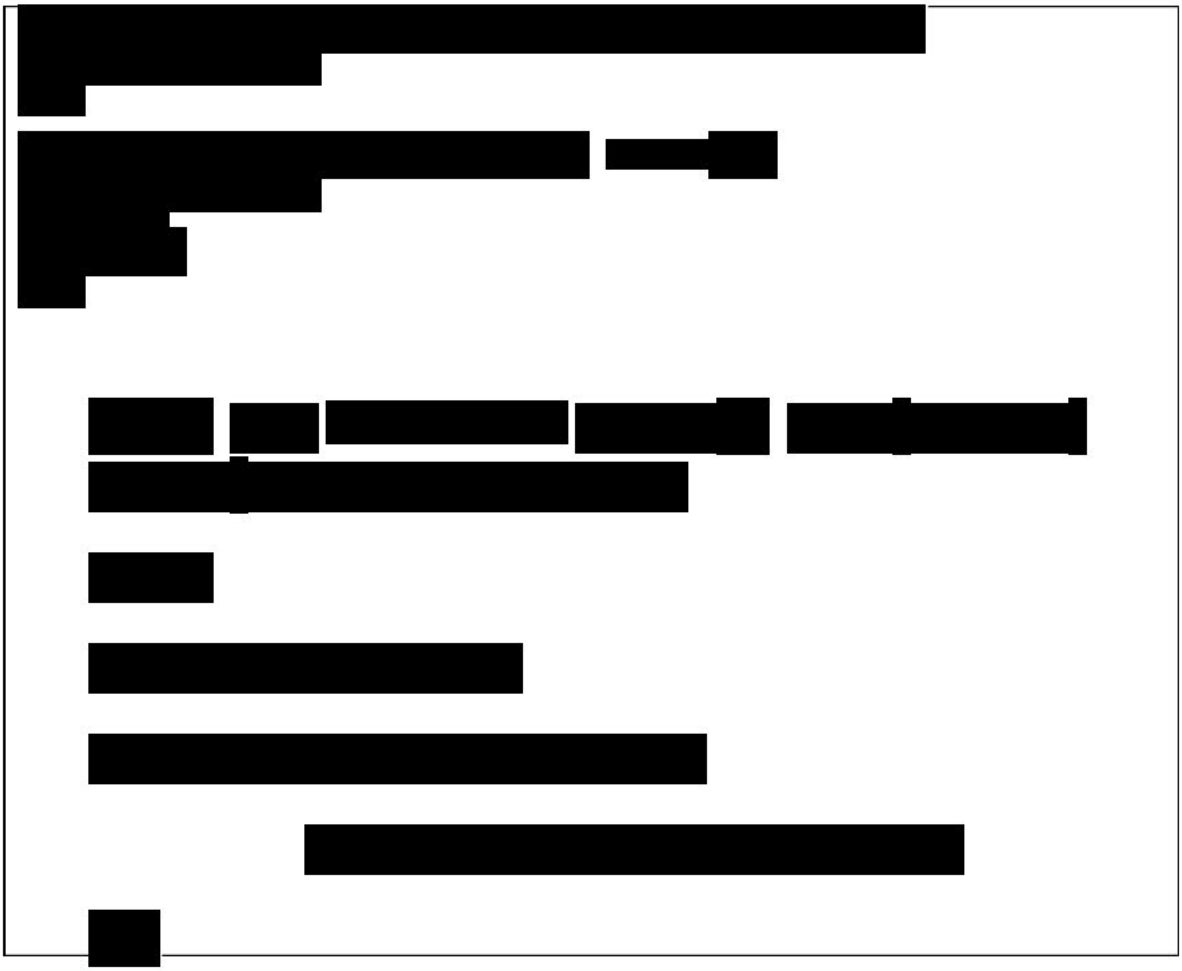
The OUT_IMP_CONT is the transposed input dataset, with columns for PGIC/CGIC at Week 48.

Table 18: SAS code to generate imputation of missing data



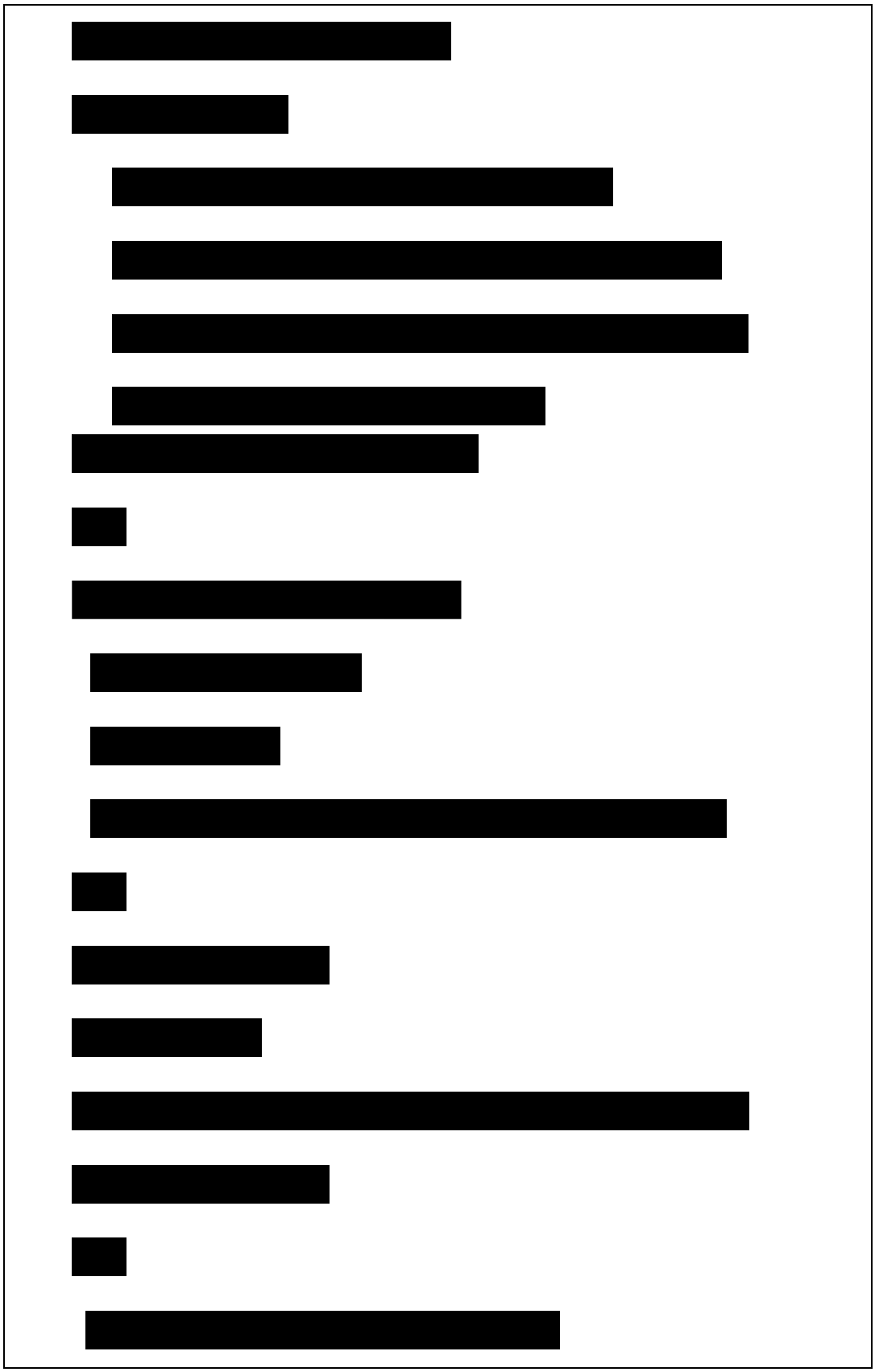
The pes cavus covariate is only utilized when analyzing results for the all randomized population (Section 6.3.2), as shown in Table 19.

Table 19: SAS code to generate imputation of missing data in the All Randomized Population



The code for the remainder of the multiple imputation process is outlined in Table 20 below, and in Table 21 for the All Randomized Population. The final result is obtained from the output of PROC MIANALYZE.

Table 20: SAS code for multiple imputation of the secondary endpoint



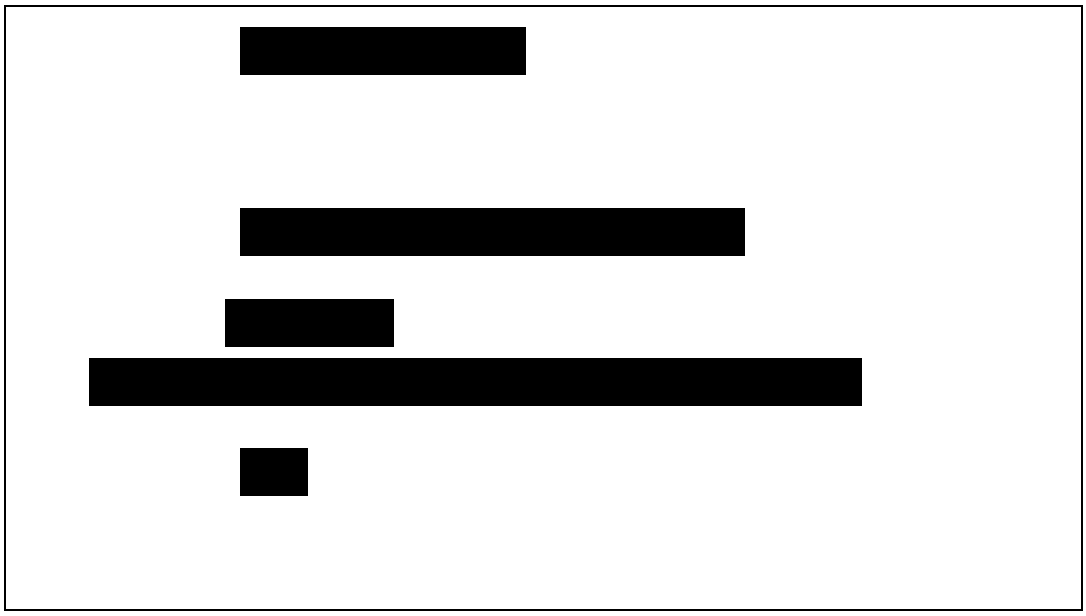
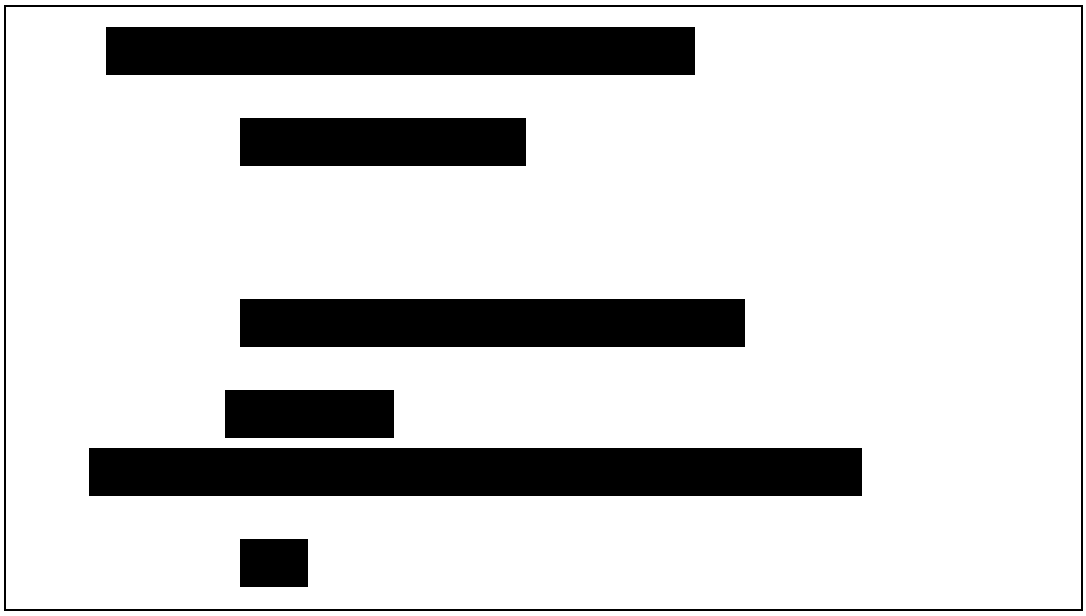
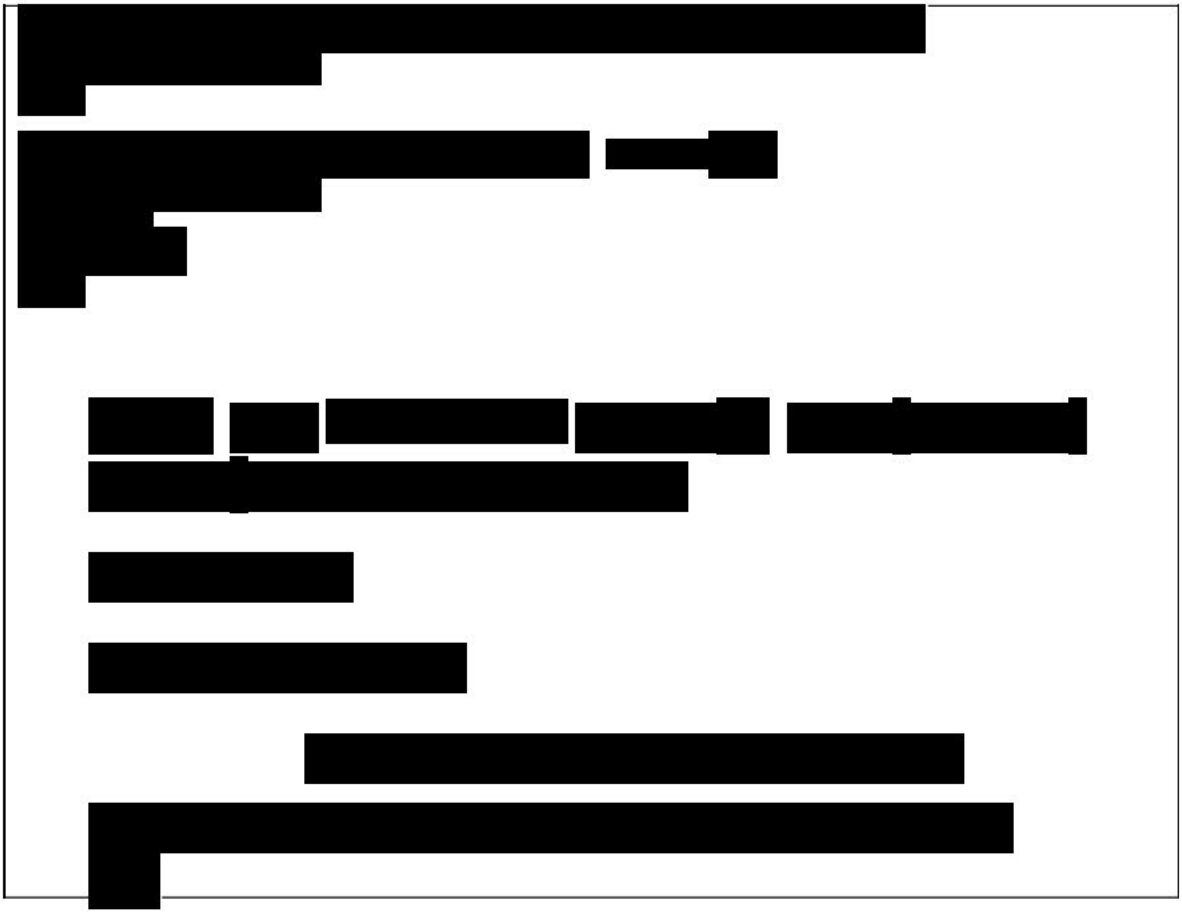


Table 21: SAS code for multiple imputation of the secondary endpoint in the All Randomized Population



Control-based imputation assumes that all data are imputed based on the placebo data. The data for the imputation is subsetted to all placebo subjects and any treated subjects with missing data at Week 48. The code for this analysis is provided in Table 22.

Table 22: SAS code to generate imputation of missing data in placebo subjects



Analysis of the results follows the code in Table 20.

14.10. Tipping point analysis of the secondary endpoint

The tipping point analysis is only performed for the Full Analysis Set (Section 6.3.1). The specific steps are as follows:

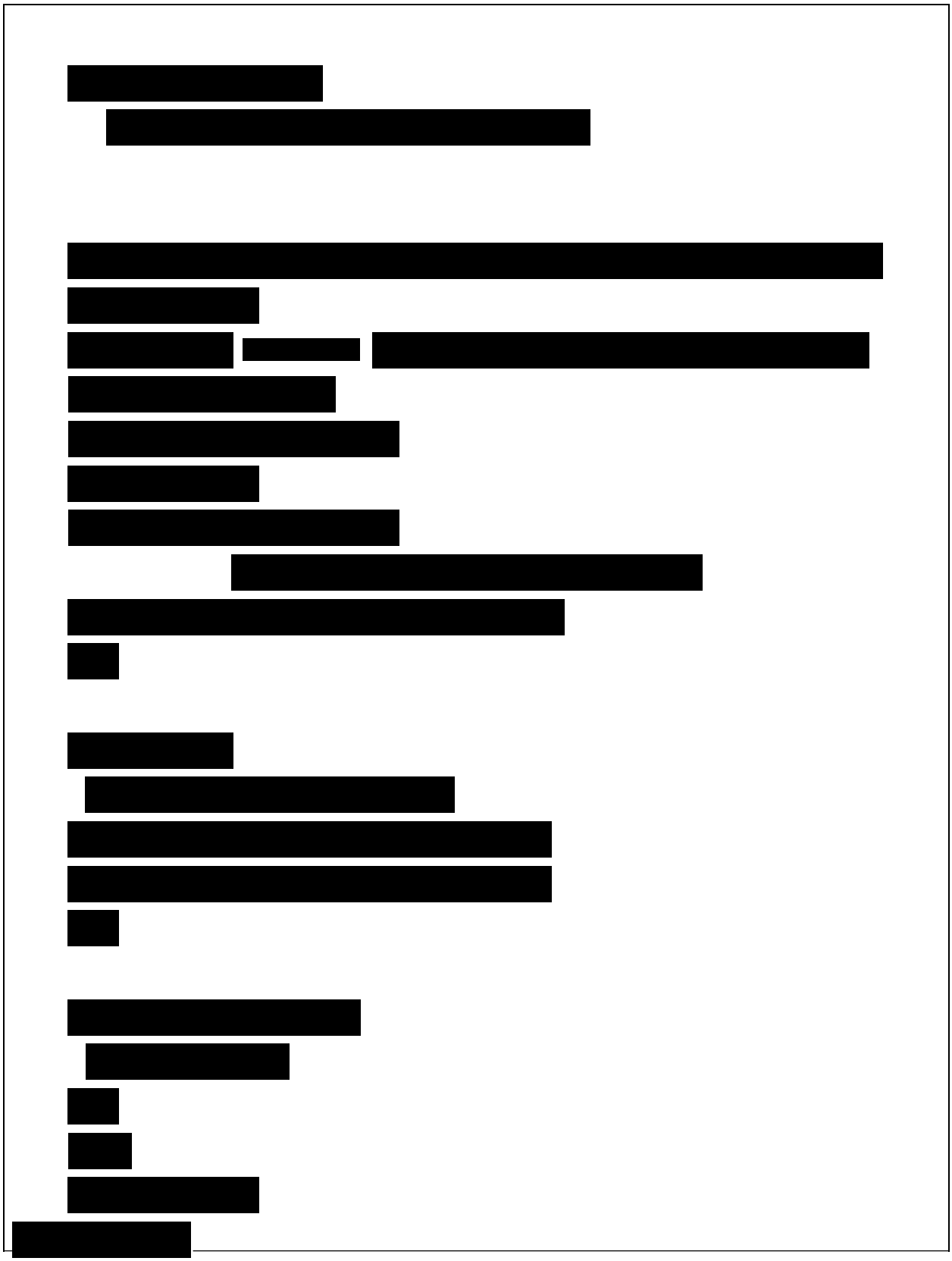
1. The missing information on PGIC/CGIC will be imputed in the placebo group only for all remaining values. This is the data set that will be used for Step 2 in all analyses.
2. The remaining missing data for PGIC/CGIC in the omaveloxolone arm will be imputed using a shift parameter S in the omaveloxolone group, where S can change in both directions progressively to find the point at which statistical significance is lost or achieved. The sample SAS code for performing multiple imputation with a shift S in the omaveloxolone arm is presented in Table 23. Missing data in the placebo arm are imputed assuming MAR, using the code in Table 23 subset to the placebo subjects.
3. The -imputed data will be analysed using standard multiple imputation combining rules to obtain an estimated treatment effect and its associated significance level.

4. Steps 3 and 4 will then be repeated. The shift parameter S starts from 0, which corresponds to the primary efficacy analysis with no shift effect, and the effect potentially increased in both a negative and positive direction by a certain amount in each step until the analysis reaches the "tipping point", the point at which the effect of omaveloxolone is no longer superior to that of placebo. The more the tipping point diverges from the observed data, the more robust the conclusion based on primary efficacy analysis.

The tipping point analysis will use a single seed from Section 14.11.

The PGIC is a transposed dataset with columns for the Week 48 PGIC values.

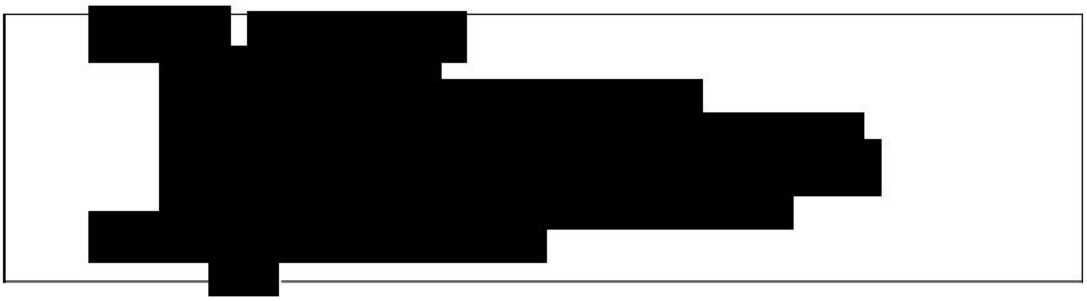
Table 23: SAS Code for Tipping Point Analysis

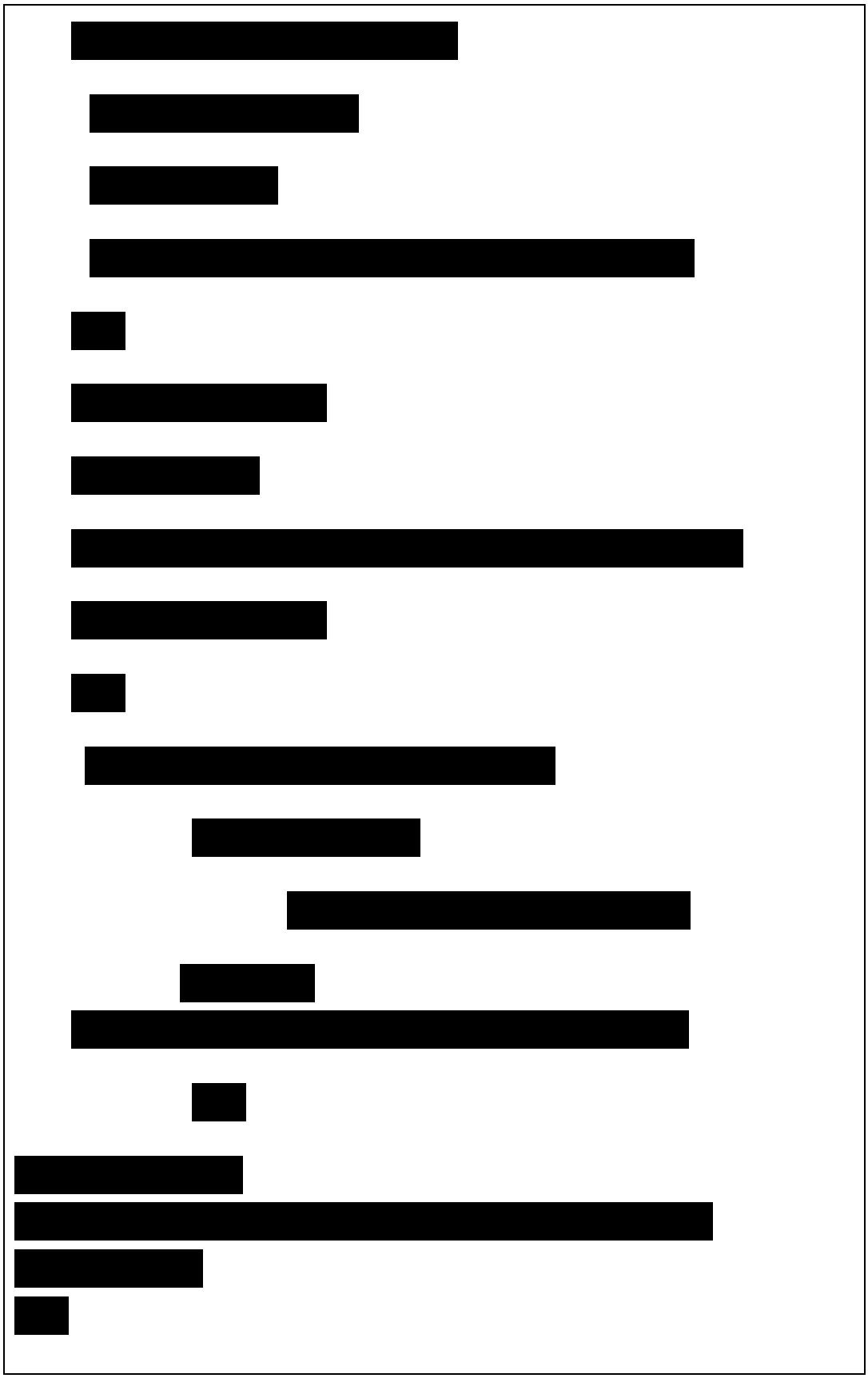


The code for the remainder of the multiple imputation process is outlined in Table 24 below. The final result is obtained from the output of PROC MIANALYZE.

For each shift value, the code in Table 24 and is also executed within the %do loop and macro shown in Table 16. For simplicity, Table 24 displays the model code and analysis of results for a single shift parameter. Once the code from Table 24 is executed within the %do loop in Table 23 for a single shift parameter, the results for that parameter are output using the &outparms variable for a particular shift value.

Table 24: SAS code for analysis of the PGIC/CGIC tipping point analysis





14.11. Random seed specification

The example code in Table 25 will be used to generate the random seeds for the analyses outlined above. The MASTER seed will be 406191150. The NUMSEED will be 13, and the seeds will be used in the order in which they are generated for the analyses.

Table 25: SAS code for random seed generation

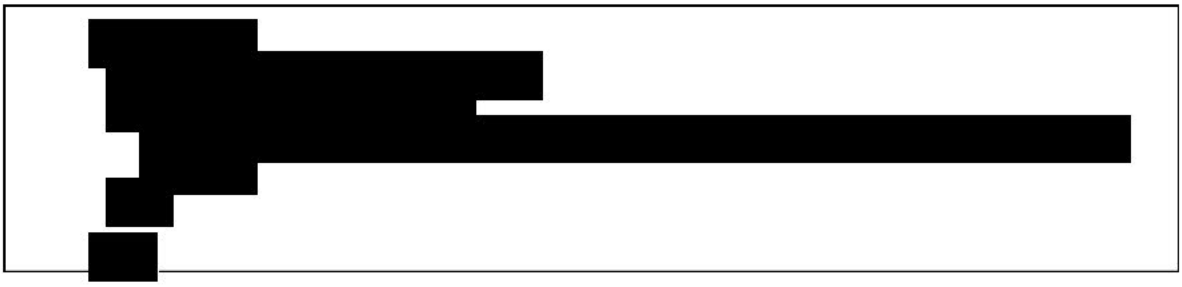


Table 26: Pre-Specified Seed Values and Analyses

Number	Seed	Description
1	1583283832	Tipping point analysis of the primary endpoint: FAS population
2	1495344738	Treatment-based imputation of the primary endpoint: FAS population
3	188497609	Control-based imputation of the primary endpoint: FAS population
4	16733020	ANCOVA analysis of the secondary endpoint: PGIC in the FAS population
5	1664555052	Tipping point analysis of the secondary endpoint: PGIC in the FAS population
6	294012164	Control-based imputation of the secondary endpoint: PGIC in the FAS population
7	284685512	ANCOVA analysis of the secondary endpoint: CGIC in the FAS population
8	987560035	Tipping point analysis of the secondary endpoint: CGIC in the FAS population
9	1955031032	Control-based imputation of the secondary endpoint: CGIC in the FAS population
10	656058139	ANCOVA analysis of the secondary endpoint: PGIC in the PCP population
11	960440053	ANCOVA analysis of the secondary endpoint: PGIC in the ARP population
12	2120233908	ANCOVA analysis of the secondary endpoint: CGIC in the PCP population
13	1207777846	ANCOVA analysis of the secondary endpoint: CGIC in the ARP population

14.12. Other Maximal Exercise Testing Parameters

Table 27: Summary of Other Maximal Exercise Testing Parameters

	At Rest (Prior to Test)	At Maximal Work (End of Test)	Change from Resting[c]
Heart rate (bpm)	X	X	X
VO_2[a] (mL/kg/min)	X	X	X
O_2 pulse[b] (mL/beat)	X	X	X
Systolic blood pressure (mmHg)	X	X	X
Diastolic blood pressure (mmHg)	X	X	X
Oxygen saturation (SpO_2) (%)	X	X	X
Forced vital capacity (FVC) (L)	X		
Forced expiration volume within one second (FEV1) (L)	X		
Maximum voluntary ventilation (MVV) (L/min)	X		
%Predicted FVC (%)	X		
VCO_2[a] (mL/kg/min)		X	
Anaerobic threshold (L/min)		X	
Time to reach maximal work (Min:Sec)		X	
Maximum rating of perceived exertion (BORG; 0-20 scale)		X	
Maximum minute ventilation (VE) BTPS (L/min)		X	
Tidal Volume (L)		X	
Maximum respiratory rate (breaths/min)		X	
Breathing Reserve (%)		X	

[a] Analysis of maximal VO2, VCO2, and resting VO2 are performed using adjustments for baseline weight:
- Maximal_VO_2_adj=Maximal_VO_2*1000/Baseline weight (kg);
- Maximal_VCO2_adj=Maximal_VCO2*1000/Baseline weight (kg);
- Resting_VO_2_adj=Resting_VO_2*1000/Baseline weight (kg);

[b] O_2 pulse are calculated for analysis: O_2 pulse = (Maximal_VO_2 (in units of mL/min)*1000)/Maximal_HR

[c] The change from resting (delta) for parameters measured at resting and at the end of the maximal exercise test is defined as:
- Delta = Parameter at maximal work - Parameter at resting

The change from baseline is defined for parameters measured at resting and at the end of the maximal exercise test is defined as:

- Change from baseline in Delta =
  [Parameter at maximal work(At Visit X) - Parameter at resting(At Visit X)] -
  [Parameter at maximal work(At Baseline) - Parameter at resting(At Baseline)]

14.13. List of Laboratory Tests

Blood samples are collected throughout the study for hematology, chemistry, and urinalysis for clinical laboratory evaluation. Test panels include the following:

Table 28: List of Laboratory Tests

Hematology	Chemistry	Urinalysis
Hematocrit	Blood urea nitrogen (BUN)	Specific gravity
Hemoglobin	Creatinine	Ketones
HbA1C	Total bilirubin	pH
Red blood cell (RBC) count	Alanine aminotransferase (ALT)	Protein
White blood cell (WBC) count	Aspartate aminotransferase (AST)	Blood
Neutrophils	Alkaline phosphatase (ALP)	Glucose
Bands (if detected)	Ferritin	Urobilinogen
Lymphocytes	Sodium	Bilirubin
Monocytes	Potassium	
Basophils (if detected)	Calcium	
Eosinophils (if detected)	Inorganic phosphorus	
Absolute platelet count	Magnesium	
Mean corpuscular hemoglobin (MCH)	Chloride	
Mean corpuscular volume (MCV)	Bicarbonate	
Mean corpuscular hemoglobin concentration (MCHC)	Uric acid	
Reticulocyte count	Cholesterol	
	Total protein	
	Glucose	
	Triglycerides	
	Albumin	
	Creatine phosphokinase (CPK)	
	Lactate dehydrogenase (LDH)	
	High-density lipoprotein cholesterol (HDL-C)	
	Low-density lipoprotein cholesterol (LDL-C)	
	Very-low-density lipoprotein cholesterol (VLDL-C)	
	Gama-glutamyl transpeptidase (GGT)	
	Estimated glomerular filtration rate (eGFR) using the CKD-EPI / Schwartz Bedside Equation	

14.14. SF-36 Scoring

This study also uses the SF-36®, a 36-item health survey to evaluate the patient's physical, social and mental well-being. It comprises 35 items that provide the generic core and an overall heath rating item.

For the SF-36®, the raw scale scores will be calculated for each of the eight dimensions (Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, Mental Health) from 0 to 100 with 0 representing poor health and 100 good health as described in the SF-36® Health Survey Manual.

According to the SF-36® Health Survey Manual, the pre-coded item values of the SF-36® should first be recoded as follows to generate the final item value:

Table 29: SF-36® final item values

Recoding of item			Final item value
3a, 3b, 3c, 3d, 3e, 3f, 3g, 3h, 3i, 3j, 4a, 4b, 4c, 4d, 5a, 5b, 5c, 9b, 9c, 9f, 9g, 9i, 10, 11a, 11c			= precoded item value
6, 11b, 11d			= 6 – precoded item value
9a, 9d, 9e, 9h			= 7 – precoded item value
1:	precoded item value =	1	= 5.0
		2	= 4.4
		3	= 3.4
		4	= 2.0
		5	= 1.0
7:	precoded item value =	1	= 6.0
		2	= 5.4
		3	= 4.2
		4	= 3.1
		5	= 2.2
		6	= 1.0
8 if Item 7 is answered:		1	= 6.0 (if Item 7 = 1)
	precoded item value =	1	= 5.0 (if Item 7 ≠ 1)
		2	= 4.0
		3	= 3.0
		4	= 2.0
		5	= 1.0
8 if Item 7 is not answered:		1	= 6.0
	precoded item value =	2	= 4.75
		3	= 3.5
		4	= 2.25
		5	= 1.0

The raw scale scores will then be calculated for each of the eight dimensions as the sum of the corresponding final item values as follows:

Table 30: SF-36® raw scale scores

Dimension	Sum of final item values	Possible raw score		
		Lowest	Highest	Range
Physical Functioning (PF)	3a, 3b, 3c, 3d, 3e, 3f, 3g, 3h, 3i, 3j	10	30	20
Role-Physical (RP)	4a, 4b, 4c, 4d	4	20	16
Bodily Pain (BP)	7, 8	2	12	10
General Health (GH)	1, 11a, 11b, 11c, 11d	5	25	20
Vitality (VT)	9a, 9e, 9g, 9i	6	22	16
Social Functioning (SF)	6, 10	2	10	8
Role-Emotional (RE)	5a, 5b, 5c	3	15	12
Mental Health (MH)	9b, 9c, 9d, 9f, 9h	7	27	20

A dimension score will only be calculated if at least half of the items in the scale are not missing; otherwise the raw scale score will be set to missing. If fewer than half the items are missing, the missing items will be imputed as the mean across completed items in the same dimension scale.

The following formula will be used to convert thee raw scale scores into transformed scale scores:

$$\text{Transformed scale} = \frac{\text{(Actual raw score – lowest possible score)}}{\text{Possible raw score range}} \times 100$$

Additionally, two summary scale scores, based on weighted combinations of the eight dimension scores, will be computed: the Physical Component Summary (PCS) and the Mental Component Summary (MCS). The algorithm for calculating PCS and MCS, which consists of three steps, is based on Table 31.

Table 31: SF-36® summary scores

	US population characteristics		Factor score coefficients	
SF-36® dimension	Mean	Standard Deviation	Coefficients for the aggregate physical score	Coefficients for the aggregate mental score
PF	84.52404	22.89490	0.42402	-0.22999
RP	81.19907	33.79729	0.35119	-0.12329
BP	75.49196	23.55879	0.31754	-0.09731
GH	72.21316	20.16964	0.24954	-0.01571
VT	61.05453	20.86942	0.02877	0.23534
SF	83.59753	22.37642	-0.00753	0.26876
RE	81.29467	33.02717	-0.19206	0.43407
MH	74.84212	18.01189	-0.22069	0.48581

Step 1: Calculate the z-score standardizations of SF-36® subscales in this study by subtracting the US population mean and dividing by the US population SD for each dimension. For instance, the z-score for PF is given by Z = (PF-84.52404)/22.89490).

Step 2: Calculate the aggregate physical and mental component scores (AGG_PHYS and AGG_MENT) by summing up all the z-scores weighted with the factor score coefficients corresponding to PCS and MCS respectively. The PCS and MCS will only be calculated if at least half of the dimension scores are not missing.

Aggregate Physical Component Score = AGG_PHYS. The score, a linear combination of standardized SF-36 subscale scores, is given in the following formula:

$$\begin{aligned}\text{AGG\_PHYS} = {} & 0.42402 \times (\text{PF} - 84.52404)/22.89490 \\ & + 0.35119 \times (\text{RP} - 81.19907)/33.79729 \\ & + 0.31754 \times (\text{BP} - 75.49196)/23.55879 \\ & + 0.24954 \times (\text{GH} - 72.21316)/20.16964 \\ & + 0.02877 \times (\text{VT} - 61.05453)/20.86942 \\ & - 0.00753 \times (\text{SF} - 83.59753)/22.37642 \\ & - 0.19206 \times (\text{RE} - 81.29467)/33.02717 \\ & - 0.22069 \times (\text{MH} - 74.84212)/18.01189\end{aligned}$$

Aggregate Mental Component Score = AGG_MENT. The score, a linear combination of standardized SF-36 subscale scores, is given in the following formula:

$$\begin{aligned}\text{AGG\_MENT} = {} & -0.22999 \times (\text{PF} - 84.52404)/22.89490 \\ & - 0.12329 \times (\text{RP} - 81.19907)/33.79729 \\ & - 0.09731 \times (\text{BP} - 75.49196)/23.55879 \\ & - 0.01571 \times (\text{GH} - 72.21316)/20.16964 \\ & + 0.23534 \times (\text{VT} - 61.05453)/20.86942 \\ & + 0.26876 \times (\text{SF} - 83.59753)/22.37642 \\ & + 0.43407 \times (\text{RE} - 81.29467)/33.02717 \\ & + 0.48581 \times (\text{MH} - 74.84212)/18.01189\end{aligned}$$

Step 3: Calculate the T-score transformation of the aggregate component scores:

$$\text{PCS} = 50 + (\text{AGG\_PHYS} \times 10)$$

$$\text{MCS} = 50 + (\text{AGG\_MENT} \times 10)$$