Protocol C0251002

A PHASE 2, DOUBLE-BLIND, RANDOMIZED, PLACEBO-CONTROLLED STUDY TO EVALUATE THE EFFICACY, SAFETY, AND TOLERABILITY OF PF-06823859 IN ADULT SUBJECTS WITH DERMATOMYOSITIS

Statistical Analysis Plan (SAP)

Version: 5.0

Date: 09-JUN-2022

TABLE OF CONTENTS

LIST OF TABLES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study C0251002 is based on the protocol dated 01 July 2020.

Table 1. Summary of Major Changes in SAP Amendments

SAP Amendment	Change	Rationale
1	Minor corrections • Pulse is measured in sitting position • RR interval is excluded from CSR/SAP • Serum (correct) replaces plasma in PK section	• Matches CRF • Matches Protocol • Corrects the language of the earlier version of the SAP.
2	• Slight widening of the acceptable analysis window in Table 4.	• Need to incorporate all available observations
3	• Slight widening of analysis window in Table 4 (week 12).	• Need to incorporate all available observations
4	This SAP amendment supports the addition of Stage 2, Amended Stage 2, and Stage 3 following protocol amendment 4. • Section 2.1.2, 2.1.3, 2.1.4: Added all objectives for newly added stages. • Section 2.2: Updated study design to reflect newly added stages. • Section 3: Add general definitions and conventions for endpoints in newly added stages.	• Align with new objectives and endpoints of newly added stages; provide context and update of design changes.

SAP Amendment	Change	Rationale
	• Section 3.2, 3.3, 3.4: Added all endpoints for newly added stages. • Section 4: Created separate subsections to define Full Analysis Sets, Per Protocol Analysis Sets, Safety Analysis Sets, and other analysis sets in newly added stages. • Section 5.1.2, 5.1.3: Added hypotheses and decision rules for newly added stages. • Section 6.1.1.1: %CFB adding to CDASI-A reporting. • Section 6.1.2: Added plots for secondary endpoint reporting. • Section 6.1.3.1, 6.1.3.2: Added population PK modeling description and additional PK plots. • Section 6.1.3.4: Clarified that post-hoc analyses may occur for PROs with established MCIDs. • Section 6.2, 6.3: Added all statistical analyses for newly added stages. • Section 6.6, 6.7: Update baseline and safety summaries to include newly added stages. • Section 7: Clarified that an interim analysis previously occurred and provided details	• New analysis sets required for analyses of newly added stages. • Clarify that all alpha is consumed in Stage 1 analyses. • Update reporting for Stage 1 analyses. • Describe reporting and analyses for newly added stages. Provide descriptions of new PRO endpoints and the TIS.

SAP Amendment	Change	Rationale
	for other potential interim analyses. • Sections 9.4.5-9.4.9: Added descriptions for PROs that were included in newly added stages. • Section 9.8: Provided summary of the TIS derivation.	
5	• Section 3: Reworded some descriptions for clarity. In this section and throughout the SAP, rephrased "CSMs of the TIS" to "components of the TIS" to emphasize analyses will be done on the native scale and not the CSM scale. • Section 4: Updated names of analysis sets for clarity. • Sections 6.1.1.2, 6.2.1.3: Added sensitivity analysis that addresses prohibited medication use. • Section 6.3.2.1.1: Added supportive and sensitivity analyses for the TIS. • Section 6.3.2.2: Clarified analyses for the components of the TIS. • Sections 6.3.2.3, 6.6.4: Added that additional plots may be provided. • Section 9.1: Updated visit windows.	• Clarify analyses and analysis set names. • Update supportive and sensitivity analyses. • Update visit windows.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study C0251002. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives

2.1.1. Study Objectives in Stage 1

2.1.1.1. Primary Objective in Stage 1

- *To evaluate the efficacy of PF-06823859 in adult subjects with moderate to severe DM.*

2.1.1.2. Secondary Objectives in Stage 1

- *To evaluate the efficacy of PF-06823859 over time.*
- *To determine the safety, and tolerability, of PF-06823859.*

2.1.1.3. Exploratory Objectives in Stage 1

- *To characterize pharmacokinetics (PK) of PF-06823859.*
- *To characterize pharmacodynamics (PD) effects of PF-06823859.*
- *To evaluate the effects of PF-06823859 on patient-reported outcomes (PROs) and physician global assessment (PhGA) over time.*
- *To evaluate the effects of PF-06823859 on CDASI sub-scores.*
- CCI [redacted]

CCI [redacted]

- [redacted]
- [redacted]

2.1.2. Study Objectives in Stage 2

2.1.2.1. Primary Objectives in Stage 2

- *To estimate the efficacy of PF-06823859 in adult participants with moderate to severe DM across two stages.*

2.1.2.2. Secondary Objectives in Stage 2

- *To estimate the efficacy of PF-06823859 across two Stages.*

- *To determine the safety, and tolerability, of PF-06823859 across two Stages.*

CCI

- *To estimate the effects of PF-06823859 on patient-reported outcomes (PROs) and physician global assessment (PhGA) over time across the two stages.*
- *To estimate the effects of PF-06823859 on CDASI sub-scores across the two stages.*

CCI

CCI

2.1.3. Study Objectives in Amended Stage 2

2.1.3.1. Primary Objectives in Amended Stage 2

- *To estimate the efficacy of PF-06823859 in adult participants with moderate to severe DM across Stage 1 and Stage 2.*

2.1.3.2. Secondary Objectives in Amended Stage 2

- *To estimate the efficacy of PF-06823859 over time across Stage 1 and Stage 2.*
- *To determine the safety, and tolerability, of PF-06823859 across Stage 1 and Stage 2.*

CCI

- *To estimate the effects of PF-06823859 on the Physician Global Assessment (PhGA, VAS) across Stage 1 and Stage 2.*
- *To estimate the effects of PF-06823859 on patient-reported outcomes (PROs) over time across Stage 1 and Stage 2.*
- *To estimate the effects of PF-06823859 on CDASI sub-scores across Stage 1 and Stage 2.*

CCI

CCI

CCI

CCI

2.1.4. Study Objectives in Stage 3

2.1.4.1. Primary Objectives in Stage 3

- *To evaluate the safety and tolerability of PF-06823859 in adult DM participants with moderate to severe active muscle disease.*

2.1.4.2. Secondary Objectives in Stage 3

- *To evaluate the efficacy of PF-06823859 over time in adult DM participants with moderate active muscle disease.*

CCI

- *To estimate the effects of PF-06823859 on the physician global assessment (PhGA, VAS) over time.*
- *To estimate the effects of PF-06823859 on patient-reported outcomes (PROs) over time.*

CCI

CCI

2.2. Study Design

Initially, this was a multi-center Phase 2A study planned to be conducted in approximately 30 adult participants with moderate to severe DM. Thirty two participants were randomized in order to collect data from 24 participants who completed the study. Each participant was to take part in the study for a period of approximately 9 months. After the 35 day screening period, participants who were eligible enrolled into the study and were randomized in a blinded manner, 2:1 (Active:Placebo) ratio. Investigational drug or placebo administration took place on Day 1, Week 4, and Week 8 of the study. Participants then entered a 5-month follow-up period.

For Stage 2, with the addition of 1 lower dose 150 mg arm, this study then became a Phase 2, dose-finding double-blind, randomized, placebo-controlled study. Approximately 20 additional participants who meet eligibility criteria will be enrolled and randomly assigned in a blinded manner to receive PF-06823859 600 mg, 150 mg, or placebo dosed every 4 weeks according to a randomization ratio. 5:11:4. Participants will receive study drug, (active or placebo) on Day 1, Week 4, and Week 8 of the study and when treatment is completed all participants will enter a 5-month follow-up period.

The study was further amended, and Stage 2 became "Amended Stage 2" changing the study to a fixed sequence design. In this design participants will be randomized to a treatment sequence at Day 1. The treatment sequence (Active -> Placebo or Placebo ->Active) will dictate whether the participant will receive active study drug during Day 1 to Week 12 or active study drug during Week 12 to Week 20. Although dosing is completed at Week 20, the treatment period goes through to Week 24. There is no washout in between the dose changes (Day 1 to Week 12 and Week 12 to Week 20) for this design. Another key feature is that this design will allow placebo subjects to receive active treatment at some timepoint during the study. No re-randomization will occur at Week 12. Participants who start the study on active study drug (600 mg or 150 mg) will have a prespecified dose change to placebo at Week 12, and participants who are on placebo will have a prespecified dose change to active study drug at Week 12. The placebo participants will receive prespecified doses of either 600 mg or 150 mg, in order to maintain equal distribution of these doses. There will be no unblinding or interruption of dosing during the treatment period. Each participant will receive 3 consecutive doses of active study drug and 3 consecutive doses of placebo during the trial. The dose change between active and placebo in this fixed sequence design will provide each participant 12 weeks of active drug. The participant, study doctor, or sponsor will not know what the participant is assigned to during the treatment period. The end of the treatment period is at Week 24, although the last dose is administered at the Week 20 visit (Visit 8). All participants will then enter a 4 month follow-up period (the interval between the last dose and the end of the follow-up period is 5 months).

Participants who entered Stage 2 prior to Protocol Amendment 4 will follow the design and SoA of Protocol Amendment 3, which is Stage 2. Statistical analyses including Stage 2 endpoints will involve data from the original Stage 2 of Amendment 3 and the Amended Stage 2 of Amendment 4 or later, as appropriate.

In Protocol Amendment 4, an additional DM cohort of 8 to 16 participants with active moderate muscle involvement (MMT-8 ≤136/150) will be enrolled. This cohort is referred to as Stage 3. Participants in this cohort are allowed to have any amount of skin involvement, as no specific CDASI score is required. All participants who meet eligibility criteria for (Stage 3) will be randomized to receive either PF-06823859 600 mg active study drug or placebo. This study design is similar to the Amended Stage 2 as all participants will receive their dosing assignment at Day 1 in a fixed sequence manner for the duration of the treatment period. The treatment period will consist of 6 dosing's occurring at Day 1, Week 4, (Visit 4), Week 8, (Visit 5), Week 12 (Visit 6), Week 16 (Visit 7), and Week 20, (Visit 8). Each participant will receive 3 consecutive doses of active study drug and 3 consecutive doses of placebo. At Week 12 (Visit 6), participants will be switched to the opposite treatment arm from what they were randomly assigned to at Day 1, Week 4 and Week 8. The study design permits all participants to receive active study drug in a blinded manner. The dose change between active and placebo will provide each participant 12 weeks of active drug. The end of the treatment period is at Week 24, although the last dose is administered at the Week 20 visit (Visit 8). All participants will then enter a 4 month follow-up period (the interval between the last dose and the end of the follow-up period is 5 months).

If a participant meets eligibility criteria for both Amended Stage 2 and Stage 3, the participant will be enrolled in Stage 3 unless the enrollment maximum for Stage 3 has been reached, at which point the participant would enroll in the Amended Stage 2.

The overall duration of the study will be approximately 69 months from first subject first visit, (FSFV) to last subject last visit, (LSLV) and will be conducted at approximately 35-40 investigative sites. Overall, approximately 76 participants will be randomized into this study.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

In Stage 1, Stage 2, and Amended Stage 2, the primary and secondary endpoints assess changes in the modified CDASI *activity* and *damage* scores. The details of the calculation of these scores are presented in Section 9.2. In Stage 3, the secondary endpoints assess changes in the modified CDASI activity and damage scores *and* changes in the TIS and the components of the TIS. The details of the calculation of the TIS are presented in Section 9.8. In each stage, baseline is defined as the measurement at Day 1 (prior to dosing). Change from baseline is defined as the value at a specific visit minus the value from baseline.

Efficacy endpoint analyses may use data from a single stage or pool data across multiple stages. Stage 1 analyses will include Stage 1 data only, and Stage 3 analyses will include Stage 3 data only. Pooled analyses of data from Stage 1, Stage 2, and Amended Stage 2 will pool data from the first 12 weeks of these stages. For CCI endpoints newly added to Amended Stage 2 that were not previously included in Stage 1 or Stage 2, analyses will occur among Amended Stage 2 participants only.

Primary analysis of the CDASI activity score in Stage 1, Stage 2, and Amended Stage 2 include data from baseline to Week 12. Sensitivity and supportive analyses in these stages may include data beyond Week 12.

Secondary analysis of the TIS, components of the TIS, CDASI activity, and CDASI damage scores in Stage 3 include data from baseline to Week 12. Sensitivity and supportive analyses in these stages may include data beyond Week 12.

3.1. Endpoints in Stage 1

3.1.1. Primary Endpoint in Stage 1

- *Change from baseline in CDASI activity score at Week 12*.

3.1.2. Secondary Endpoint(s) in Stage 1

- *Absolute values and change from baseline of CDASI activity and CDASI damage scores at all scheduled time points. Exception: The change from baseline in CDASI activity score at Week 12 is a primary endpoint*.

- *Incidence of adverse events (AEs), laboratory abnormalities, changes in vital signs, and electrocardiogram (ECG) findings.* The extended discussion of these endpoints is presented in Section 9.3.

CCI [redacted]

- *Absolute values and change from baseline in the values of selected biomarkers (Gene signature panel in blood and tissue; Human Myxovirus A (MXA), interferon gamma-induced protein 10 (IP-10), IFNbeta, high sensitive C-reactive protein, (hsCRP), muscle and bone blood biomarkers, autoantibody concentrations in serum/plasma) at all scheduled time points.*

- *Absolute values and change from baseline in the values of PROs (the short form (SF36) Health Survey, Dermatology and Life Quality Index (DLQI), 5D Itch Scale, and physician's evaluation of disease (PhGA)) at all scheduled time points.* The brief description of the scales is presented in Section 9.4.

- *Absolute values and change from baseline values of the sub-scores of CDASI activity and CDASI damage scores at all scheduled time points.* The description on the scores is presented in Section 9.2.

- CCI [redacted]

CCI [redacted]

[redacted].

CCI [redacted]

3.2. Endpoints in Stage 2

3.2.1. Primary Endpoint in Stage 2

- *Change from baseline in CDASI activity score at Week 12.*

3.2.2. Secondary Endpoints in Stage 2

- *Absolute values and change from baseline of CDASI activity and CDASI damage scores at all scheduled times points. (Exception: The change from baseline in CDASI activity score at Week 12 is a primary endpoint).*

- *Incidence of adverse events (AEs) laboratory abnormalities, changes in vital signs, and electrocardiogram (ECG) findings.* The extended discussion of these endpoints is presented in Section 9.3.

CCI [redacted]

- *Absolute values and change from baseline in the values of selected biomarkers (Gene signature panel in blood and tissue; Human Myxovirus A (MXA), interferon gamma-induced protein 10 (IP-10), IFNbeta, high sensitive C-reactive protein, (hsCRP), muscle and bone blood biomarkers, autoantibody concentrations in serum/plasma) at all scheduled time points.*

- *Absolute values and change from baseline in the values of PROs (the short form (SF36) Health Survey, Dermatology and Life Quality Index (DLQI), 5D Itch Scale, and physician's evaluation of disease (PhGA)) at all scheduled time points.* The brief description of the scales is presented in Section 9.4.

- *Absolute values and change from baseline values of the sub-scores of CDASI activity and CDASI damage scores at all scheduled time points.* The description on the scores is presented in Section 9.2.

CCI [redacted]

CCI [redacted]

CCI [redacted]

3.3. Endpoints in Amended Stage 2

3.3.1. Primary Endpoint in Amended Stage 2

- *Change from baseline in CDASI activity score at Week 12.*

3.3.2. Secondary Endpoints in Amended Stage 2

- *Absolute values and change from baseline of CDASI activity and CDASI damage scores at all scheduled time points. (Exception: The change from baseline in CDASI activity score at Week 12 is a primary endpoint).*

- *Incidence of adverse events (AEs) laboratory abnormalities, changes in vital signs, and electrocardiogram (ECG) findings.* The extended discussion of these endpoints is presented in Section 9.3.

CCI [redacted]

- *Absolute values and change from baseline in the values of selected biomarkers (Gene signature panel in blood and tissue;* CCI [redacted]

- *Absolute values and change from baseline in the values of the PhGA, VAS at all scheduled time points.*

- *Absolute values and change from baseline in the values of PROs including the Patient Global Assessment (PtGA), the Health Assessment Questionnaire - Disability Index (HAQ-DI), the 5-D Pruritus Scale,* CCI [redacted]*, the Short Form (SF-36 v2 Acute), the Dermatology and Life Quality Index (DLQI) and the European Quality of Life Five Dimension, Five Level Scale (EQ-5D-5L) & EQ Visual Analogue Scale (EQ-VAS) at all scheduled time points.* The brief description of the scales is presented in Section 9.4.

- *Absolute values and change from baseline values of the sub-scores of CDASI activity and CDASI damage scores at all scheduled time points.* The description on the scores is presented in Section 9.2.

CCI [redacted]

CCI [redacted]

- [redacted]
- [redacted]

3.4. Endpoints in Stage 3

3.4.1. Primary Endpoint in Stage 3

- *Incidence of adverse events (AEs) laboratory abnormalities, changes in vital signs, and electrocardiogram (ECG) findings.* The extended discussion of these endpoints is presented in Section 9.3.

3.4.2. Secondary Endpoints in Stage 3

- *Total Improvement Score (TIS) at Week 12 and intermediate scheduled time points.* The description of the TIS is presented in Section 9.8.
- *Change from baseline in the core Set Measures (CSM) of the TIS including PGA, PtGA, MMT, HAQ-DI, muscle enzymes, and extra-muscular activity, (MDAAT).*
- *Absolute values and change from baseline of CDASI activity and CDASI damage scores at all scheduled time points.* The description on the scores is presented in Section 9.2.

CCI [redacted]

- [redacted]
- *Absolute values and change from baseline in the values of selected biomarkers (Gene signature panel in blood; Human Myxovirus A (MXA), interferon gamma-induced protein 10 (IP-10), IFNbeta, highly sensitive C-reactive protein, (hsCRP), muscle and bone blood biomarkers, autoantibody concentrations in serum/plasma) at all scheduled time points.*
- *Absolute values and change from baseline in the values of the PhGA, VAS at all scheduled time points.*
- *Absolute values and change from baseline in the values of PROs including the Patient Global Assessment (PtGA), the Health Assessment Questionnaire - Disability Index (HAQ-DI), the 5-D Pruritus Scale,* CCI [redacted] *the Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-F), the Short Form (SF-36 v2 Acute), and the European Quality of Life Five Dimension, Five Level Scale (EQ-5D-5L) & EQ*

Visual Analogue Scale (EQ-VAS) at all scheduled time points. The brief description of the scales is presented in Section 9.4.

- CCI [redacted]
- [redacted]
- [redacted]

3.5. Other Endpoints

Medical history, pregnancy testing, physical examination and recording of concomitant medications data will be analyzed according to Pfizer standard practices.

3.6. Baseline Variables

The day 1 measurements will serve as the baseline for clinical reported outcomes (CDASI Scores, Physician Global Assesment Score), patient-reported outcomes (SF36, DLQI, 5D Itch Scale), medical assessments (skin biopsies, ECG, weight, vital signs), laboratory safety measurements CCI [redacted] Demographics and medical history data will be collected at the baseline.

4. ANALYSIS SETS

4.1. Full Analysis Set

Multiple full analysis sets are defined. They are listed briefly here and described in more detail in Sections 4.1.1 through 4.1.2.

- The Full Analysis Set (FAS) includes all subjects who received at least one dose of randomized treatment in any study stage.
- The Full Analysis Set in Stage 1 (FAS1) includes all subjects who received at least one dose of randomized treatment in Stage 1.
- The Full Analysis Set in Stage 2 (FAS2) includes all subjects who received at least one dose of randomized treatment in Stage 2.
- The Full Analysis Set in Amended Stage 2 (FASA2) includes all subjects who received at least one dose of randomized treatment in Amended Stage 2.

- The Full Analysis Set in Stage 1 and Stage 2 (FAS12) includes all subjects who received at least one dose of randomized treatment in Stage 1 or Stage 2.

- The Pooled Full Analysis Set, skin cohort (PFASS) includes all subjects who received at least one dose of randomized treatment in Stage 1, Stage 2, or Amended Stage 2.

- The Full Analysis Set in Stage 3 (FAS3) includes all subjects who received at least one dose of randomized treatment in Stage 3.

- The Pooled Full Analysis Set, fixed sequence stages (PFASF) includes all subjects who received at least one dose of randomized treatment in Amended Stage 2 or Stage 3.

4.1.1. Full Analysis Sets in Stage 1, Stage 2, and Amended Stage 2

The FAS1 is defined as all subjects who have received at least one dose of randomized treatment in Stage 1. This is the primary analysis population for the safety and treatment compliance in Stage 1. In the analyses it is assumed that subjects are assigned to the randomized treatment regardless of what treatment was received. Note that Section 9.2 of the protocol references a modified intent to treat (mITT) set. In Stage 1, the mITT set in Stage 1 is synonymous with the FAS1 mentioned in this document.

The FAS2 is defined as all subjects who have received at least one dose of randomized treatment in Stage 2. This is the primary analysis population for the safety and treatment compliance in Stage 2.

The FAS12 is defined as all subjects who have received at least one dose of randomized treatment in Stage 1 or Stage 2. The FAS12 may be used for analyses assessing the durability of response for subjects in Stage 1 or Stage 2.

The FASA2 is defined as all subjects who have received at least one dose of randomized treatment in Amended Stage 2. This analysis population will be used only for endpoints unique to Amended Stage 2 that were not collected in other stages.

The PFASS is defined as all subjects who have received at least one dose of randomized treatment in Stage 1, Stage 2, or Amended Stage 2. This is the primary analysis population for the safety and treatment compliance across Stage 1, Stage 2, and Amended Stage 2. In the analyses it is assumed that subjects are assigned to the randomized treatment regardless of what treatment was received. Note that Section 9.2 of the protocol references a mITT set. The mITT set for Stage 1, Stage 2, and Amended Stage 2 is synonymous with the PFASS mentioned in this document.

4.1.2. Full Analysis Sets in Stage 3

The FAS3 is defined as all subjects who have received at least one dose of randomized treatment in Stage 3. This is the primary analysis population for the safety and treatment compliance in Stage 3. In the analyses it is assumed that subjects are assigned to the randomized treatment regardless of what treatment was received. Note that Section 9.2 of

the protocol references a mITT set. In Stage 3, the mITT set in Stage 3 is synonymous with the FAS in Stage 3 mentioned in this document.

The PFASF is defined as all subjects who have received at least one dose of randomized treatment in Amended Stage 2 or Stage 3. The PFASF may be used for analyses assessing the durability of response for subjects with baseline CDASI activity score of 14 or greater in Amended Stage 2 or Stage 3.

4.2. Per Protocol Analysis Set

4.2.1. Per Protocol Analysis Set in Stage 1

The Per Protocol Analysis Set in Stage 1 (PPAS1) is a subset of the FAS1 and includes subjects who did not have major protocol deviations. The list of major protocol deviations will be finalized by the project team prior to un-blinding of the study. The PPAS1 may be used in sensitivity analyses. For evaluation of sensitivity of the primary analysis in Stage 1 we will not use data collected after Week 12.

4.2.2. Per Protocol Analysis Set in Stage 2 and Amended Stage 2

The per protocol analysis set for pooled Stage 1, Stage 2, and Amended Stage 2, skin cohort (PPASS) is a subset of the PFASS and includes subjects who did not have major protocol deviations. The list of major protocol deviations will be finalized by the project team prior to un-blinding of the study. The PPASS may be used in sensitivity analyses. For evaluation of sensitivity of the primary analysis we will not use data collected after Week 12.

4.2.3. Per Protocol Analysis Set in Stage 3

The per protocol analysis set in Stage 3 (PPAS3) is a subset of the FAS3 and includes subjects who did not have major protocol deviations. The list of major protocol deviations will be finalized by the project team prior to un-blinding of the study. The PPAS3 may be used in sensitivity analyses. For evaluation of sensitivity of the secondary analysis in Stage 3 we will not use data collected after the participant receives 12 weeks of active treatment.

4.3. Safety Analysis Set

Multiple safety analysis sets are defined.

- The safety analysis set (SAS) includes all subjects who received at least one dose of randomized treatment in any stage.
- The safety analysis set in Stage 1 (SAS1) includes all subjects who received at least one dose of randomized treatment in Stage 1.
- The safety analysis set in Stage 2 (SAS2) includes all subjects who received at least one dose of randomized treatment in Stage 1 or Stage 2.
- The safety analysis set in Amended Stage 2 (SASA2) includes all subjects who received at least one dose of randomized treatment in Amended Stage 2.

- The safety analysis set in Stage 3 (SAS3) includes all subjects who received at least one dose of randomized treatment in Stage 3.

Subjects who are not treated will be excluded from the safety analysis.

CCI

- In Stage 1, the biomarker analysis population is defined as all enrolled subjects in Stage 1 who received at least one dose of PF-06823859 with at least one biomarker assessment.

- For Stage 1, Stage 2, and Amended Stage 2, the biomarker analysis population is defined as all enrolled participants in Stage 1, Stage 2 and Amended Stage 2 who received at least one dose of PF-06823859 with at least one biomarker assessment.

- In Stage 3, the biomarker analysis population is defined as all enrolled participants in Stage 3 who received at least one dose of PF-06823859 with at least one biomarker assessment.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and Decision Rules in Stage 1

The primary hypothesis to be tested is that there is no difference between the effects of treatment and placebo exposures on the CDASI activity score in Stage 1.

The primary endpoint in Stage 1 is the mean change from baseline in CDASI activity score at Week 12. The placebo-adjusted treatment effect is defined as the difference (mean change from baseline at Week 12 in the active treatment group minus the mean change from baseline at Week 12 in the placebo group) in the mean change of CDASI activity score from baseline at Week 12. A successful treatment should *decrease* CDASI activity score and therefore the

desired treatment effect is negative. The value -5 is viewed as the minimal clinically relevant value of the treatment effect.

The treatment is declared efficacious if each of the following conditions is satisfied:

- C1: Estimated value of the treatment effect is lower than -5.
- C2: An upper limit of 2-sided 90% confidence interval for the treatment effect is below zero.

5.1.2. Hypotheses and Decision Rules in Stage 2 and Amended Stage 2

The primary endpoint for Stage 2 and Amended Stage 2 is the change from baseline in CDASI activity score at Week 12 and uses data from Stage 1, Stage 2, and Amended Stage 2. No p-values will be presented for the comparisons listed in the objectives and endpoints of this pooled analysis. The placebo-adjusted treatment effect is defined as the difference (mean change from baseline at Week 12 in the active treatment group minus the mean change from baseline at Week 12 in the placebo group) in the mean change of CDASI activity score from baseline at Week 12. No hypotheses will be tested in the pooled analysis.

5.1.3. Hypotheses and Decision Rules in Stage 3

The primary endpoint in Stage 3 is safety, including incidence of adverse events (AEs) laboratory abnormalities, changes in vital signs, and electrocardiogram (ECG) findings. Major efficacy endpoints, including the TIS, the components of the TIS, and the CDASI activity and damage scores, are assessed as secondary endpoints. No hypotheses will be tested in Stage 3.

5.2. General Methods

Pfizer's standards will be used to describe distributions of outcomes for each of the visits and treatment groups in each stage. In addition to that the following statistical methods will be used. To explore the data further and to assess the goodness-of-fit of all statistical models, separate SAS datasets and common separated value files will be provided by the clinical programmer to the statistician.

5.2.1. Analyses for Continuous Longitudinal Data

5.2.1.1. Longitudinal Analysis of Covariance (LANCOVA) Model

A discussion of this model may be found in Lu (2010). We will use the following variations of the method: LANCOVA-P and LANCOVA-S where letters P and S denote the versions that are used for primary (P) and sensitivity (S) analyses.

The LANCOVA-P model uses the change from baseline as an outcome and the baseline value as a covariate. The model assumes visit-independent adjustment of the mean outcome for baseline, saturated treatment by time model for the mean and an un-structured within subject error covariance matrix. Treatment and time are considered as factor variables and treatment by time is considered as unstructured interaction effect.

The LANCOVA-S model also uses the change from baseline as an outcome and the baseline value as a covariate. The model assumes a saturated baseline by time and treatment by time fixed effects and an un-structured within subject error covariance. Treatment and time are considered as factor variables and baseline by time and treatment by time are considered as unstructured interaction effects. This allows the baseline and treatment effects to vary over time.

Variations of LANCOVA-P and LANCOVA-S models with the covariance matrices that have less general structure (eg, compound symmetry) may be considered if there are convergence issues with the primary versions of the model.

The analysis will provide an estimate and a two-sided 90% confidence interval for the difference (for each of the active doses against placebo) in the expected values of the outcome at multiple visits.

The sample SAS code for the LANCOVA calculations is presented in Section 9.6.

5.2.1.2. Conventional ANCOVA Model

The conventional ANCOVA method will use the difference between post-baseline and baseline values as an outcome and the baseline value as a covariate and will provide an estimate and a two sided 90% confidence interval for the difference (each of the active doses against placebo) in the expected values of the outcome at each of the individual visits.

The ANCOVA-based estimate of the treatment effect at the particular visit incorporates only a visit-specific post-baseline observations of outcome while the LANCOVA-based estimate simultaneously estimates treatment effect at all treatment visits using all of the observed data. The ANCOVA-based estimate will be used for sensitivity analysis. A qualitative similarity of the LANCOVA and ANCOVA-based estimates of the treatment effect will support the robustness of the primary (LANCOVA-based) estimate.

The sample SAS code for the ANCOVA calculations is presented in Section 9.6.

5.3. Methods to Manage Missing Data

- In the longitudinal modeling (LANCOVA) of repeated measures of continuous outcomes (both efficacy measures and covariates) it is assumed that missing data are generated by the “missing at random” mechanism and, therefore imputation of the missing data is not necessary.

CCI

[redacted]

- For other endpoints, statistical analyses will be based on available data. Missing data will not be imputed.

6. ANALYSES AND SUMMARIES

6.1. Analyses and Summaries During Stage 1

6.1.1. Primary Endpoint in Stage 1

The primary endpoint is the change from baseline in CDASI activity score at Week 12.

6.1.1.1. Primary Analysis in Stage 1

- Analysis time point**:** Week 12.
- Analysis population: subset of FAS1 that *includes baseline observations and observations collected at weeks 1-12.* Note that subjects with some missing data are included.
- Analysis methodology**:** change from baseline will be analyzed using the LANCOVA-P model (as described in Section 5.2.1.1).
- Supporting objective: Primary Objective.
- Reporting results (for CDASI activity score change from baseline at Week 12):
  - The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm.
  - The LS means, 90% confidence interval for the LS means, difference between the LS means of treatment groups (active-placebo) and the corresponding 90% confidence interval for the estimate. These summary statistics will also be reported for the percent change from baseline in CDASI activity score at Week 12.

6.1.1.2. Sensitivity Analyses 1-3 in Stage 1

- The sensitivity analyses will explore sensitivity of the result of the primary analysis to the following variatons of the data analysis methods (as described in Section 5.2.1) and data sets (as described in Sections 4.1.1, 4.2.1).

Table 2. Details of Approaches to Sensitivity Analyses of Primary Analysis

Sensitivity Analysis	Data Set (population)	Included Observations (Time Window)	Method
1	FAS1	Baseline, weeks 1-12	LANCOVA-S
2	FAS1	Baseline, week 12	ANCOVA
3	PPAS1	Baseline, weeks 1-12	LANCOVA-S

- Here FAS1 and PPAS1 denote full and per-protocol data sets defined in Sections 4.1.1 and 4.2.1. The reporting results will include the LS means, 90% confidence interval for the LS means, difference between the LS means of treatment groups (active-placebo) and the corresponding 90% confidence interval.

- Note that in the sensitivity analyses (similarly to the primary analysis) we *do not use data collected after Week 12 visit*.

- In addition, the analysis for the CDASI-A in Section 6.1.1.1 may be repeated adjusting for subjects meeting protocol discontinuation criteria for use of prohibited medications as an intercurrent event (if any). Data from visits occurring after meeting discontinuation criteria for use of prohibited medications will be censored from this analysis.

6.1.2. Secondary Endpoint(s) in Stage 1

- Absolute values and change from baseline of CDASI activity and CDASI damage scores at all scheduled time points.

In the analyses (both the secondary analysis and corresponding sensitivity analyses) *the observations collected after Week 12 are included into the modeling*. The main goal of the analyses is to produce the estimate and two-sided confidence interval (with 90% coverage) of the treatment effect at all scheduled time points (ie, 1, 4, 8, 12, 16, 20, 24, 28 weeks). In addition to that the standard descriptive statistics for each of the two outcomes (CDASI activity and CDASI damage score), each of the treatment groups and each visit will be presented for absolute values of the scores. The descriptive statistics will include number of observations, mean, standard deviation, median, minimum and maximum values.

6.1.2.1. Analysis of Secondary Endpoints in Stage 1

- Analysis time points**:** 1, 4, 8, 12, 16, 20, 24, 28 weeks.

- Analysis population: FAS1.

- Analysis methodology**:** change from baseline will be analyzed using the LANCOVA-P model (as described in Section 5.2.1.1).

- Supporting objective: Secondary Objective.

- Reporting results (for CDASI scores changes at each visit):

- The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm.

- The LS means, 90% confidence interval for the LS means, difference between the LS means of treatment groups (active-placebo) and the corresponding 90% confidence interval for the estimate. Plots of this data will also be presented.

6.1.2.2. Sensitivity Analyses 1-3 in Stage 1

- The analyses will explore sensitivity of the result of the primary analysis to the following variatons of the data analysis methods (as described in Section 5.2.1) and data sets (as described in Sections 4.1.1, 4.2.1).

Table 3. Details of Approaches to Sensitivity Analyses of Secondary Analysis

Sensitivity Analysis	Data Set (population)	Included Observations	Method
1	FAS1	All	LANCOVA-S
2	FAS1	All	ANCOVA
3	PPAS1	All	LANCOVA-S

- Here FAS1 and PPAS1 denote full and per-protocol data sets defined in Sections 4.1.1 and 4.2.1. The reporting results will include the LS means, 90% confidence interval for the LS means, difference between the LS means of treatment groups (active-placebo) and the corresponding 90% confidence interval.

CCI [redacted]

[redacted]

6.1.3.3. Additional Biomarkers (Gene Signature Panel in Blood and Tissue, Muscle Bone and Blood Biomarkers, Autoantibody Concentrations in Serum/Plasma) in Stage 1

These analyses will be reported in the additional exploratory analysis report.

6.1.3.4. Patient Reported Outcomes (SF36, DLQI, 5D Itch Scale) and Physician's Evaluation of Disease (PhGA), CDASI Activity and CDASI Damage Subscores in Stage 1

- Analysis population: FAS1.
- Analysis time points: all visits with samples taken.
- Analysis methodology: summary statistics. For SF36(PCS and MCS scores), DLQI, 5D itch scale and PhGA scores the change from baseline will be also investigated by applying the LANCOVA-P model (as described in Section 5.2.1.1).
- Supporting objective: Exploratory Objective.
- Reporting results:
  - Change from baseline: The sample size, mean, standard deviation, median, minimum and maximum by visit.
  - Absolute values: The sample size, mean, standard deviation, median, minimum and maximum by visit.
- For the SF36 (short form Health Survey) observations two summary component scores (PCS and MCS) and 8 scale subscores (PF, RP, BP, GH, VT, SF, RE, and MH) will be summarized.

- For the DLQI (Dermatology and Life Quality Index), 5D Itch Scale and physician's evaluation of disease (PhGA) scores a single score summarizes each of the corresponding patient-reported outcomes.

- For the CDASI activity score we will analyze the following subscores: total extent score (ES), total erythema score (ER), total scaling (SC) and total erosion/ulceration scores (EU), Gottorn hands score (GHS), peringual score (PS), and alopecia score (AS).

- For the damage score we will analyze the following subscores: total poilkiloderma score (POLS), total calcinosis score (CALS) and Gotorn's hands damage score (GHDS).

We emphasize that the LANCOVA-P models will only be created for SF36 (PCS and MCS scores), DLQI, 5D itch scale and PhGA scores. Ad-hoc analyses for PROs with minimum clinically important differences may be performed.

6.2. Analyses and Summaries During Stage 2 and Amended Stage 2

6.2.1. Primary Endpoint in Stage 2 and Amended Stage 2

The primary endpoint is the change from baseline in CDASI activity score at Week 12.

6.2.1.1. Primary Analysis in Stage 2 and Amended Stage 2

The treatment effect is defined as the difference (between each pooled active treatment group and the pooled placebo group across Stage 1, Stage 2, and Amended Stage 2 in the mean change of CDASI activity score from baseline at Week 12. The pooled 600 mg treatment arm will include data from baseline to Week 12 from participants assigned to the 600 mg in Stage 1, the 600 mg arm in Stage 2, and the 600 mg -> placebo arm in Amended Stage 2. The pooled 150 mg treatment arm will include data from baseline to Week 12 from participants assigned to the 150 mg arm in Stage 2 or the 150 mg -> placebo arm in Amended Stage 2. The pooled placebo arm will include data from baseline to Week 12 from participants assigned to the placebo arm in Stage 1, the placebo arm in Stage 2, the placebo ->150 mg arm in Amended Stage 2, and the placebo ->600 mg arm in Amended Stage 2.

- Analysis time point**:** Week 12.

- Analysis population: subset of FAS2 that *includes baseline observations and observations collected at weeks 1-12 for subjects in Stage 1, Stage 2, and Amended Stage 2.* Note that subjects with missing data are included.

- Analysis methodology**:** change from baseline will be analyzed using the LANCOVA-P model (as described in Section 5.2.1.1).

- Supporting objective: Primary Objective.

- Reporting results [for CDASI activity score change (between each active treatment group and placebo group across Stage 1, Stage 2, and Amended Stage 2) from baseline at Week 12]:

    - The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm. Results will also be presented by study stage for these summary statistics.

    - The LS means, 90% confidence interval for the LS means, difference between the LS means of treatment groups (each active treatment arm - placebo across the two stages) and the corresponding 90% confidence interval for the estimate. No p-values will be computed for the analysis in this stage.

6.2.1.2. Secondary Analysis in Stage 2 and Amended Stage 2

Data permitting, the secondary analysis for the primary endpoint in Stage 2 will be the PF-06823859 exposure-response analysis of the combined data from Stage 1, Stage 2, and Amended Stage 2. The Population Modeling Analysis Plan will describe the evaluation of the relationship between PF-06823859 exposures and response at Week 12.

6.2.1.3. Sensitivity Analyses in Stage 2 and Amended Stage 2

The sensitivity analyses from Stage 1 (Section 6.1.1.2) may be repeated using the pooled data from Stage 1, Stage 2, and Amended Stage 2 using the PFASS. For these sensitivity analyses, estimates and CIs may be produced. No p-values will be produced.

The analysis for the CDASI-A in Section 6.2.1.1 using PFASS may be repeated adjusting for subjects meeting protocol discontinuation criteria for use of prohibited medications as an intercurrent event (if any). Data from visits occurring after meeting discontinuation criteria for use of prohibited medications will be censored from this analysis.

6.2.1.4. Supportive Analyses in Stage 2 and Amended Stage 2

Supportive analyses may be reported in Stage 2. For onesupportive analysis, summary statistics will be reported by treatment sequence and visit for the participants in the FASA2 from baseline to Week 28. The sample size, mean, standard deviation, median, minimum and maximum will be presented.

6.2.2. Secondary Endpoint(s) in Stage 2 and Amended Stage 2

The analyses and summaries of the secondary endpoints in Stage 2 will be the same as those in Stage 1 (Section 6.1.2) except that they will be conducted for the data across Stage 1, Stage 2, and Amended Stage 2 and use the appropriate Stage 2 analysis set. No p-values will be presented for these analyses.

6.2.3. Exploratory Efficacy Endpoints in Stage 2, and Amended Stage 2

The analyses and summaries of the exploratory efficacy endpoints in Stage 2 will be the same as those in Stage 1 (Section 6.1.3) except that they will be conducted for the data across Stage 1, Stage 2, and Amended Stage 2 and use the appropriate Stage 2 analysis set.

Statistical summaries will stratified by treatment group. No p-values will be presented for these analyses.

Percent change from baseline in the CDASI activity score will also be analyzed and reported in the same manner as the change from baseline CDASI activity score as described in Section 6.2.1.1.

For PROs newly added to Amended Stage 2 that were not collected in Stage 2, including the PtGA, HAQ-DI, CCI [redacted] and EQ-5D-5L & EQ-VAS, descriptive statistics will be provided at each visit using the FASA2 and reported by treatment group. The reporting results for these PROs will be the same as those in Section 6.1.3.4, with some additional details provided below:

- For the PtGA and EQ-VAS scores, a single score summarizes each of the corresponding patient-reported outcomes (see Sections 9.4.5 and 9.4.8).
- For the CCI [redacted] each of the three questions will be summarized by a single score (see Section 9.4.7).
- For the EQ-5D-5L, the summed number of checked boxes under each of the five sections will be reported (see Section 9.4.8).
- For the HAQ-DI, a single averaged score summarizes each of the corresponding patient-reported outcomes under the eight sections (see Section 9.4.6). In sections of the HAQ-DI relating to aids or devices or help from another person, each checkable box will be treated as binary data and summarized with proportions and frequency counts. A single score summarizes the severity of pain VAS.

6.3. Analyses and Summaries During Stage 3

6.3.1. Primary Endpoint in Stage 3

The primary endpoint in Stage 3 includes the incidence of adverse events (AEs) laboratory abnormalities, changes in vital signs, and electrocardiogram (ECG) findings. All safety summaries and analyses will be performed as specified in Section 6.7.

6.3.2. Secondary Endpoint(s) in Stage 3

6.3.2.1. Total Improvement Score in Stage 3

- Analysis population: FAS3.
- Analysis time points: Week 12 and intermediate scheduled time points.
- Analysis methodology: The TIS at post baseline visits are scored relative to baseline and the TIS at baseline is 0 by definition; thus all analyses are performed directly on the TIS. The estimates of the treatment effect from baseline to Week 12 will be obtained by MMRM where the TIS will be used as a dependent variable. This model uses fixed effects for treatment, time (visit), treatment by time, and a random effect for subject, Due to the unknown nature of the longitudinal data, different covariance

structures among repeated measures will be examined based on model diagnostics starting with the unstructured variance-covariance model.

- Supporting objective: Secondary objective.
- Reporting results: The absolute value of the TIS at each post baseline visit will be summarized by the following descriptive statistics:
  - The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm (active and placebo).
  - The LS means, 90% confidence interval for the LS means, difference between the LS means of treatment groups (active treatment arm - placebo) and the corresponding 90% confidence interval for the estimate. Plots of this data will also be presented. No p-values will be computed for the analysis in this stage.
  - The proportion of subjects achieving minimal, moderate, and major responses in their TIS score (defined as scores of ≥20, ≥40, and ≥60 in the TIS respectively) will also be reported by treatment.

6.3.2.1.1. Supportive and Sensitivity Analyses for the TIS in Stage 3

Additional analyses may be performed for the TIS in Stage 3.

- Summary statistics will be reported by treatment sequence and time for the participants in the FAS3 from baseline to Week 24. The sample size, mean, standard deviation, median, minimum and maximum will be presented.
- The analysis for the TIS in Section 6.3.2.1 may be repeated adjusting for subjects meeting protocol discontinuation criteria for use of prohibited medications as an intercurrent event (if any). Data from visits occurring after meeting discontinuation criteria for use of prohibited medications will be censored from this analysis.
- The analysis for the TIS in Section 6.3.2.1 may be repeated using treatment sequence as a covariate from baseline to Week 24.
- The analysis for the TIS in Section 6.3.2.1 may be repeated using last observation carried forward (LOCF) for missing TIS values.

6.3.2.2. Components of the TIS in Stage 3

The absolute values and change from baseline in each of the components of the TIS, including the PhGA (from the MDAAT), PtGA, MMT, HAQ-DI, muscle enzymes, and extramuscular global assessment, will be summarized by descriptive statistics at each visit, including number of observations, mean, standard deviation, median, minimum and maximum values.

Estimates of treatment effect for selected components of the TIS (PhGA, PtGA, extramuscular global assessment, HAQ-DI, and MMT-8) will be obtained by LANCOVA in the same manner as described in Section 6.1.2 where the change from baseline of a specified component will be used as a dependent variable. The absolute values of these five components will also be assessed using an MMRM model. Estimates and the confidence intervals for treatment effect will be presented for each of these methods. Accompanying plots may also be provided.

6.3.2.3. CDASI Activity and CDASI Damage Scores in Stage 3

The absolute values and change from baseline of CDASI activity and damage scores at all scheduled time points will be summarized by descriptive statistics, including number of observations, mean, standard deviation, median, minimum and maximum values.

The estimates for treatment effect may be obtained by LANCOVA in the same manner as described in Section 6.1.2 where the change from baseline of a corresponding score (ie, activity or damage score) will be used as a dependent variable. Estimates and the confidence intervals for treatment effect will be presented for each of these methods. Accompanying plots may also be provided.

A supportive analysis that includes subjects with skin disease (baseline CDASI activity score of 14 or greater) in Stage 3 (FAS3) may also be performed for the CDASI activity and damage score using the models specified in Section 6.1.2.

6.3.3. Exploratory Efficacy Endpoints in Stage 3

Exploratory efficacy endpoints will be summarized in Stage 3 in the same manner as the corresponding endpoints in Stage 1, Stage 2, or Amended Stage 2 using the appropriate Stage 3 analysis set (Section 6.1.3; Section 6.2.3). The only newly added PRO to Stage 3 is the FACIT-F. The average ranking across all questions of the FACIT-F for each patient will be summarized in the same manner as the PhGA (see Sections 6.1.3.4 and 9.4.9).

CCI [redacted]

6.5. Subset Analyses

Subset analysis is not planned.

6.6. Baseline and Other Summaries and Analyses

6.6.1. Baseline Summaries

In (a) Stage 1, (b) Stage 1, Stage 2, and Amended Stage 2, and (c) Stage 3, a breakdown of all subjects will be provided by demographic characteristics (age, race, weight, body mass index, and height), extent of disease and prior use of conventional therapy in accordance with Pfizer data standards. Mean steroid use and duration of disease will also be presented by stage. Baseline CDASI (activity and damage) will be summaried descriptively.

Subject evaluation groups will show end of study subject disposition and will show which subjects were analyzed for PK, as well as for safety (adverse events and laboratory data). Frequency counts will be supplied for subject discontinuation(s) by treatment.

6.6.2. Study Conduct and Subject Disposition

In (a) Stage 1, (b) Stage 1, Stage 2, and Amended Stage 2, and (c) Stage 3, subject evaluation groups will show end of study subject disposition and will show which subjects were analyzed in the full analysis set, and as well as for safety, PK and PD. Frequency counts will be supplied for subject discontinuations by treatment. Disposition will also be summarized separately for subjects who discontinue.

Data will be reported in accordance with reporting standards.

6.6.3. Study Treatment Exposure

In (a) Stage 1, (b) Stage 1, Stage 2, and Amended Stage 2, and (c) Stage 3, number of the doses of active and control treatments will be recorded and summarized by the treatment group.

6.6.4. Concomitant Medications and Non-Drug Treatments

All concomitant medication(s) as well as non-drug treatment(s) will be provided in the listings. These will be listed for both stages and stage will be indicated. A separate listing may be provided for DM-related prior and concomitant medications. Additional plots may be provided that illustrate concomitant medication use prior to and during the study period.

6.7. Safety Summaries and Analyses

The safety will be monitored over the study period that includes treatment and follow-up periods in all stages.

6.7.1. Adverse Events

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. Separate AE tables will show the data from (a) Stage 1, (b) Stage 1, Stage 2, and Amended Stage 2, and (c) Stage 3.

6.7.2. Laboratory Data

Laboratory data will be listed and summarized by treatment in accordance with the sponsor reporting standards. The laboratory data listing will be listed for all stages and stage will be indicated.

6.7.3. Vital Signs

Systolic blood pressure, diastolic blood pressure and pulse rate will be listed and tabulated by dose group and week with descriptive statistics (N, mean, standard deviation, median, minimum and maximum). Change from baseline (defined as mean pre-dose value collected at Week 0) will also be summarized using the same descriptive statistics by dose group and week in (a) Stage 1, (b) Stage 1, Stage 2, and Amended Stage 2, and (c) Stage 3.

The numbers of subjects with vital sign outcomes within the specified categories (see Section 9.7) will be tabulated by treatment and study week.

6.7.4. Electrocardiogram

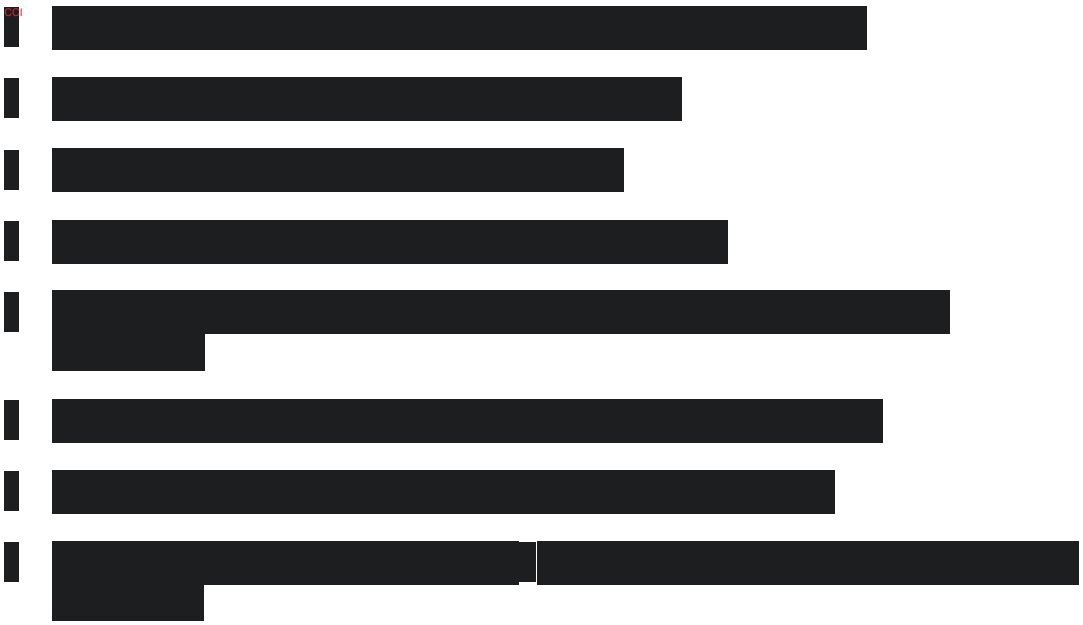
The following ECG data will be listed: QT, QTc (Fridericia's), heart rate, QRS duration, PR interval. QTc based on Fridericia's correction will be derived.

Baseline and change from baseline for QT, QTcF, heart rate, QRS, PR will be summarized using descriptive statistics by treatment and study week in (a) Stage 1, (b) Stage 1, Stage 2, and Amended Stage 2, and (c) Stage 3. For QTcF a classification of absolute values and increase from baseline will be used.

The numbers of subjects with ECG outcomes within the specified categories (see Section 9.7) and number of subjects with uncorrected QT values ≥500 ms will be summarized by treatment and study week.

CCI [redacted]

CCI [redacted]



6.7.7. Physical Examination

All physical exam data will be provided in the listings.

6.7.8. Other Safety Data

Prior medication(s), non-drug treatment(s), medical history, lab data and physical examination will be listed in accordance with the sponsor reporting standards. Any other data that is captured on the study database, will be listed.

7. INTERIM ANALYSES

7.1. Introduction

In Stage 1 of the study, an interim analysis was performed when 32 participants were randomized and completed Visit 6 (Week 12). Additional interim analyses during later stages of the study *may b*e conducted depending on the enrollment rate. If an additional interim analysis is done in the future, an internal review committee (IRC) independent from the blinded study team will review efficacy and safety data for business decision planning, if applicable. The results of the interim analysis will not be shared with the study team, participants, or investigators.

It is expected that in addition to the study team, an unblinded team may review the CCI (ie, any data not explicitly stated as primary or secondary endpoint) and exposure data on an ongoing basis. The analyses of the group will not be included into the report described in the current SAP.

8. REFERENCES

1. Lu. K. (2010) On efficiency of constrained longitudinal data analysis versus longitudinal analysis of covariance. Biometrics; 66: 891–896.

2. Aggarwal R, Rider LG, Ruperto N, et al. 2016 American college of rheumatology/European league against rheumatism criteria for minimal, moderate, and major clinical response in adult dermatomyositis and polymyositis: an international myositis assessment and clinical studies group/paediatric rheumatology international trials organisation collaborative initiative. Arthritis & Rheumatology. 2017;69(5):898-910.

9. APPENDICES

9.1. Definition and Use of Visit Windows in Reporting

The currently assumed assignment of the observations to visits is based on the subject's arrival day. It assumes that the arrival times for each visit occur within the (-5,7) days time window centered around the target day as specified in the schedule of activities. We will check the validity of the assumption by the analysis of the visit specific, distributions of actually observed arrivival days at the end of the study before data unblinding and may update the assignment rules slightly.

Table 4. Definition and Use of Visit Windows in Reporting

Visit Label	Targeted Day	Analysis window
Screening	35 days prior to Day 1	1-35 days prior to Day 1
Baseline	Day 1 (reference day)	Day 1 (reference day)
Week 1	Day 8	Days 3-16
Week 4	Day 29	Days 24-38
Week 8	Day 57	Days 49-64
Week 12	Day 85	Days 71-99
Week 16	Day 113	Days 108-122
Week 20	Day 141	Days 136-155
Week 24	Day 169	Days 163-188
Week 28	Day 197	Days 191-218
Week 32	Day 226	Days 222-236
Week 36	Day 252	Days 245-259
Week 40	Day 280	Days 273-287

If more than one observation from the same subject falls into the same window then the association of the observation with the visit will be done after the consultation with the study clinician and lead statistician.

9.2. CDASI

Table 5. CDASI Scale

(b) Cutaneous dermatomyositis disease area and severity index (CDASI) *v2*

Select the score in each anatomical location that describes the most severely affected dermatomyositis-associated skin lesion

Extent	Activity				Damage		
Anatomical location	Erythema	Scale	Erosion/ Ulceration		Poikiloderma (Dyspigmentation or Telangiectasia)	Calcinosis	Anatomical location
	0-absent 1-pink; faint erythema 2-red 3-dark red	0-absent 1-scale 2-crust; lichenification	0-absent 1-present		0-absent 1-present	0-absent 1-present	
Scalp							Scalp
Malar area							Malar area
Periorbital							Periorbital
Rest of the face							Rest of the face
V-area neck (frontal)							V-area neck (frontal)
Posterior neck							Posterior neck
Upper back & shoulder							Upper back & shoulder
Rest of back & buttocks							Rest of back & buttocks
Abdomen							Abdomen
Lateral upper thigh							Lateral upper thigh
Rest of leg & feet							Rest of leg & feet
Arm							Arm
Mechanic's hand							Mechanic's hand
Dorsum of hands (not over joints)							Dorsum of hands (not over joints)
Gottron's – not on hands							Gottron's – not on hands

Gottron's – Hands

Examine patient's hands and double score if papules are present		Ulceration		Examine patient's hands and score if damage is present	
0-absent 1-pink; faint erythema 2-red erythema 3-dark red				0-absent 1-dyspigmentation 2-scarring	

Periungual

Periungual changes (examine)			
0-absent 1-pink; red erythema/microscopic telangiectasias 2-visible telangiectasias			

Alopecia

Recent Hair loss (within last 30 days as reported by patient)			
0-absent 1-present			

Total activity score (For the activity score, please add up the scores of the left side, i.e. Erythema, Scale, Erosion/ Ulceration, Gottron's, Periungual, Alopecia)		Total damage score (For the damage score, add up the scores of the right side, i.e. Poikiloderma, Calcinosis, Gottron's)	

Figure 2 presents the CDASI scale. The main outcomes of interest are the activity score and damage score.

9.2.1. The Activity Score (AS)

The Activity Score (AS) is calculated as a sum of the contribution from the extent score (ES), Gottorn hands score (GHS), peringual score (PS) and allopecia score (AS).

$$AS = ES+GHS +PS+AS$$

The total extent score (ES) is obtained by summing up scores for the total erythema (ER, quantifies redness of the skin or mucous membranes), total scaling (SC, quantifies peeling of the skin) and total erosion/ulceration (EU, quantifies presence of the deeper wound).

$$ES = ER+SC+EU$$

Total erythema (ER), scaling (SC) and erosion/ulceration (EU) scores are calculated as a sum of the contributions from 15 individual areas of the body. The range of the total erytherma score is 0-45, the range of the total scaling score is 0-30 and the range of total erosion/ulceration is 0-15. The resulting range of the extent score is 0-90.

Gottorn hands score (GHS) characterizes papules (swellings) on hand and is a sum of the papule's characterization score (range 0-6) and ulceration score (range = 0-1).

Peringual score (PS) characterizes abnormalities around nails and its range is 0-2. The alopecia score (AS) characterizes hair loss and has the range 0-1.

The resulting range of the activity score is 0-100.

9.2.2. Damage Score

The Damage Score (DS) is calculated as a sum of the total poilkiloderma score (POLS), total calcinosis score (CALS) and Gotorn's hands damage score (GHDS).

$$DS = POLS +CALS+GDHS$$

The polikiolderma score characterizes specific dispigmentation in the particulal area and calcinosis score characterizes calcification of the skin in the particular area. The total polikiolderma score (POLS) and the total calcinosis score (CALS) are summed up over 15 individual areas in the body and each of them has range 0-15. The Gotorn's hands damage score (GHDS) has the range 0-2 so that the damage score (DS) has the range 0-32.

In our analysis we will look at the activity score (AS), damage score (DS).

For the activity score we will analyze the following subscores: total extent score (ES), total erythema score (ER), total scaling (SC) and total erosion/ulceration scores (EU), Gottorn hands score (GHS), peringual score (PS), and alopecia score (AS).

For the damage score we will analyze the following subscores: total poilkiloderma score (POLS), total calcinosis score (CALS) and Gotorn's hands damage score (GHDS).

9.3. Additional Information for Safety Endpoints

9.3.1. Adverse Events

An adverse event (AE) is considered treatment emergent relative to a given treatment if:

- the event occurs for the first time during the effective duration of treatment and was not seen prior to the start of treatment (for example, during the baseline or run-in period), or
- the event was seen prior to the start of treatment but increased in severity during treatment.

9.3.2. Laboratory Data

The description of the laboratory data is given in Section 7.8 of the protocol.

9.3.3. Vital Signs

Vital signs (blood pressure, pulse, respiratory rates and temperature) will be measured after 5 minutes of rest as indicated in the Schedule of Activities. Baseline will be the last measurement prior to receiving study treatment.

9.3.4. Electrocardiogram

Single 12-lead ECGs should be collected at times specified in the Schedule of Activities.

The screening ECG values will serve as each subject's baseline values. To ensure safety of the subjects, a qualified individual (investigator or sub-investigator) at the investigator site will make comparisons to baseline measurements. A paper or digital copy of the ECG should be filed in the subject's chart and must be available to the sponsor upon request. Any clinically significant changes will be recorded as AEs and evaluated further, as clinically warranted. In some cases, it may be appropriate to repeat abnormal ECGs to rule out improper lead placement as contributing to the ECG abnormality.

9.4. Patient-Reported Outcomes and physician's Evaluation of Disease

9.4.1. SF36: The Short Form Health Survey

The SF-36 v.2 (Acute version) is a 36-item generic health status measure. It measures 8 general health concepts or domains: physical functioning (PF), role physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role emotional (RE) and mental health (MH). These domains are also summarized as physical and mental component scores (PCS and MCS). All domains and summary components are scored such that a higher score indicates a higher functioning or health level.

9.4.2. DLQI: Dermatology Life Quality Index

The DLQI is calculated by summing the responses to each of the 10 questions (see Appendix 7 of the protocol) resulting in maximum of 30 and minimum of 0. The higher value of the score denotes higher impairment/lower quality of life.

9.4.3. 5D Itch Scale

The scores of each of the five domains in the questionary (see Appendix 8 of the Protocol) are calculated separately and then summed together to obtain a total 5-D score. 5-D scores can potentially range between 5 (no pruritus) and 25 (most severe pruritus).

Single-item domain scores (duration, degree and direction) are equal to the value indicated below the response choice (range 1–5). The disability domain includes four items that assess the impact of itching on daily activities: sleep, leisure/social activities, housework/errands and work/school. The score for the disability domain is calculated by taking the highest score on any of the four items.

For the distribution domain, the number of affected body parts is tallied (potential sum 0–16) and the sum is sorted into five scoring bins: sum of 0–2 = score of 1, sum of 3–5 = score of 2, sum of 6–10 = score of 3, sum of 11–13 = score of 4, and sum of 14–16 = score of 5.

9.4.4. PhGA: Physician Global Assessment

The score is a subjective assessment of the severity of disease by the physician. The physician uses the visual analog scales and puts a mark on 0-10 cm scale where higher score indicates worse status.

9.4.5. PtGA: Patient Global Assessment

The score is a subjective assessment of the severity of disease by the patient. The patient uses a visual analog scales and puts a mark on the scale where higher score indicates worse status (see Appendix 13 of the protocol).

9.4.6. HAQ-DI: Health Assessment Questionnaire and Disease Index

The HAQ-DI contains eight sections (including dressing & grooming, arising, eating, walking, hygiene, grip, reach, and activities). Each section has multiple questions that the participant can use to rank their functionality (see Appendix 12 of the protocol). For each participant, the average ranking will be reported for each of the eight sections.

The HAQ-DI also contains sections where the participant can indicate use of aids or devices or help from another person from a list of choices.

The HAQ-DI also includes a VAS that the participant can use to indicate the severity of their pain, where a higher score indicates more pain.

CCI [redacted]

9.4.8. EQ-5D-5L & EQ-VAS

The EQ-5D-5L contains five sections, including mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (see Appendix 10 of the protocol). Each of the five sections will be summarized by the number of checked boxes under their respective headers.

The EQ-VAS ranges from 0 to 100, where 100 is the best (ie, most healthful) score.

9.4.9. FACIT-F

The FACIT-F contains 13 questions with ranked scores (see Appendix 11 of the protocol). Each patient's average ranked score will be used in reporting.

CCI

CCI

CCI

C-CASA Event Code	C-CASA Event	C-SSRS Response
1	Completed suicide	As captured in the safety database
2	Suicide attempt	"Yes" on "Actual Attempt"
3	Preparatory acts towards imminent suicidal behavior	"Yes" on any of the following: • "Aborted attempt", or • "Interrupted attempt", or • "Preparatory Acts or Behavior"
4	Suicidal ideation	"Yes" on any of the following: • "Wish to be dead", or • "Non-Specific Active Suicidal Thoughts", or • "Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act", or • "Active Suicidal Ideation with Some Intent to Act, without Specific Plan", or • "Active Suicidal Ideation with Specific Plan and Intent"
7	Self-injurious behavior, no suicidal intent	"Yes" on "Has subject engaged in Non-suicidal Self-Injurious Behavior?"

9.6. Examples of SAS Code for the LANCOVA and ANCOVA Analyses

Create data set ds where for each subject and each visit the following information is stored in the row of the data set.

Table 7. Input Parameters for LANCOVA and ANCOVA Analyses

Id	Subject's id
Cfb	Change from baseline
Time	time of interest (treated as a class variable). Only the post-baseline visits are included
Trt	Treatment group (stored as a class variable, trt="Active" denotes active treatment and "Control" denotes control treatment)
Base	Baseline observation of the outcome

The analysis will be implemented as follows. In all of the LANCOVA and ANCOVA analyses we use the reference baseline value XXX (see below) which is the mean of the baseline values for all participants.

```
%let basemean=XXX;
```

LANCOVA-P

```
proc mixed data=ds ;
        class         id trt (ref=”Control”)  time (ref=”2”);
        model         cfb  =  base time trt  time*trt  / ddfm=kr;
        repeated      time /subject=  id type= un;
        lsmeans       visit*trt / pdiff at base = &basemean alpha=0.10 cl;
        ods output  diffs= dlancovaP  lsmeans= dlancovaMP;
run;

data  dlancovaP;
  set dlancovaP;
  if( (trt="Active" & _trt="Control") & (time=_time));
run;
```

The resulting estimates of the visit-dependent difference in means of cfb (active-control treatments) and means of cfb for active and control groups are stored in the dlancovaP and dlancovaMP data sets.

LANCOVA-S

```
proc mixed data=ds ;
        class         id trt (ref=”Control”)  time (ref=”2”);
        model         cfb = base time trt  time*trt  time*base / ddfm=kr;
        repeated      time /subject= id type= un;
        lsmeans       visit*trt / pdiff at base= &basemean alpha=0.10 cl;
        ods output    diffs= dlancovaS  lsmeans= dlancovaMS;
run;

data  dlancovaS;
  set dlancovaS;
  if( (trt="Active" & _trt="Control") & (time=_time));
run;
```

The only difference between the LANCOVA-S and LANCOVA-P analyses is the addition of the time*base term to the model.

The resulting estimates of the visit-dependent difference in means of cfb (active-control treatments) and means of cfb for active and control groups are stored in the dlancovaS and dlancovaMS data sets.

ANCOVA

Create data set ds as described in the Table 7 and sort it by visit. The analysis will be implemented as follows

```
proc mixed data=ds ;
        by        visit;
        class     id trt (ref="Control")  visit;
        model     cbaseline = trt baseline /ddfm=kr;
        lsmeans trt  / at baseline = &basemean  alpha=0.10 cl pdiff;
        ods output   diffs= dancovaS  lsmeans= dancovaMS;
run;
```

The resulting estimates of the visit-dependent difference in means of cfb (active-control treatments) and means of cfb for active and control groups are stored in the dancovaS and dancovaMS data sets.

9.7. Categories for ECG and Vital Sign Outcomes

The following Pfizer's standard will be used. For the sitting measurements of pulse rate the categories for the supine pulse rate (min <40 bpm and max >120 bpm) will be used.

Categories for QTcF

QTcF (msec)	450≤ max. <480	480≤ max.<500	max. ≥500
QTcF (msec) increase from baseline	30≤ max. <60	max. ≥60	

Categories for PR and QRS

PR (msec)	max. ≥300	
PR (msec) increase from baseline	Baseline >200 and max. ≥25% increase	Baseline ≤200 and max. ≥50% increase
QRS (msec)	max. ≥200	
QRS (msec) increase from baseline	Baseline >100 and max. ≥25% increase	Baseline ≤100 and max. ≥50% increase

Categories for Vital Signs

Systolic BP (mm Hg)	min. <90	
Systolic BP (mm Hg) change from baseline	max. decrease ≥30	max. increase ≥30
Diastolic BP (mm Hg)	min. <50	
Diastolic BP (mm Hg) change from baseline	max. decrease ≥20	max. increase ≥20
Supine pulse rate (bpm)	min. <40	max. >120
Standing pulse rate (bpm)	min. <40	max. >140

Measurements that fulfill these criteria are to be listed in the study report.

9.8. TIS Calculation in DM

The total improvement score is the sum of all 6 improvement scores associated with the change in each core set measure. A total improvement score of ≥20 represents minimal improvement, a score of ≥40 represents moderate improvement, and a score of ≥60 represents major improvement.[2]

ACR/EULAR CRITERIA FOR CLINICAL RESPONSE IN ADULT DERMATOMYOSITIS AND POLYMYOSITIS

Table 3. Final myositis response criteria for minimal, moderate, and major improvement in adult dermatomyositis/polymyositis (DM/PM) and combined adult DM/PM and juvenile DM clinical trials and studies*

Core set measure, level of improvement based on absolute percent change	Improvement score
Physician global activity	
Worsening to 5% improvement	0
>5% to 15% improvement	7.5
>15% to 25% improvement	15
>25% to 40% improvement	17.5
>40% improvement	20
Patient global activity	
Worsening to 5% improvement	0
>5% to 15% improvement	2.5
>15% to 25% improvement	5
>25% to 40% improvement	7.5
>40% improvement	10
Manual muscle testing	
Worsening to 2% improvement	0
>2% to 10% improvement	10
>10% to 20% improvement	20
>20% to 30% improvement	27.5
>30% improvement	32.5
Health Assessment Questionnaire	
Worsening to 5% improvement	0
>5% to 15% improvement	5
>15% to 25% improvement	7.5
>25% to 40% improvement	7.5
>40% improvement	10
Enzyme (most abnormal)	
Worsening to 5% improvement	0
>5% to 15% improvement	2.5
>15% to 25% improvement	5
>25% to 40% improvement	7.5
>40% improvement	7.5
Extramuscular activity	
Worsening to 5% improvement	0
>5% to 15% improvement	7.5
>15% to 25% improvement	12.5
>25% to 40% improvement	15
>40% improvement	20

The **total improvement score** is the sum of all 6 improvement scores associated with the change in each core set measure. A total improvement score of **≥20** represents **minimal improvement**, a score of **≥40** represents **moderate improvement**, and a score of **≥60** represents **major improvement**.

* Note that these response criteria are also proposed for use in combined adult DM/PM and juvenile DM trials (20). For comparison, the thresholds of improvement in the total improvement score for juvenile DM are ≥30 for minimal improvement, ≥45 for moderate improvement, and ≥70 for major improvement. Also note that the criteria for major improvement for adult DM/PM are preliminary.

How to calculate the improvement score: The absolute percent change ([final value − baseline value]/range × 100) is calculated for each core set measure. For muscle enzymes, the most abnormal serum muscle enzyme level at baseline (creatine kinase, aldolase, alanine transaminase, aspartate aminotransferase, lactate dehydrogenase) is used. The enzyme range was calculated based on a 90% range of enzymes from natural history data (34,46), which for creatine kinase is 15 times the upper limit of normal (ULN), for aldolase is 6 times the ULN, and for lactate dehydrogenase, aspartate aminotransferase, and alanine transaminase is 3 times the ULN. The ULN is determined according to the individual laboratories in the participating centers. The ranges for physician global activity, patient global activity, manual muscle testing, Health Assessment Questionnaire, and extramuscular global activity are based on the instrument scale used (3,26). An improvement score is assigned for each core set measure based on the absolute percent change in the core set measure according to the definition. These individual core set measure improvement scores are then totaled among the 6 core set measures to give the total improvement score. The thresholds for minimal, moderate, and major improvement are provided. The total improvement score itself may also be compared among treatment arms in a trial. A total improvement score between 0 and 100 corresponds to the degree of improvement, with higher scores corresponding to a greater degree of improvement.