STATISTICAL ANALYSIS PLAN (SAP)

A Randomized, Double-blind, Placebo-controlled, Repeat-dose, Multicenter Trial to Evaluate the Efficacy, Safety, Tolerability and Pharmacokinetics of HZN-825 in Patients with Diffuse Cutaneous Systemic Sclerosis

Study Number:	**HZNP-HZN-825-301**
Study Product:	**Fipaxalparant (HZN-825)**
Development Phase:	Phase 2b
Sponsor:	Horizon Therapeutics DAC (a wholly owned subsidiary of Amgen Inc.) 70 St. Stephen's Green Dublin 2 D02 E2X4 Ireland
Version:	1.2
Version Date:	28OCT2024
Compliance:	This study will be conducted in accordance with standards of Good Clinical Practice (as defined by the International Council for Harmonisation) and all applicable national and local regulations.

COMMERCIAL IN CONFIDENCE

Table of Contents

[redacted]

List of Figures

List of Tables

1 Modification History

Version	Effective Date	Author of Modification	Summary of Change
1.0	08Dec2021	[redacted]	N/A – First Version
1.1	12OCT2023	[redacted]	1. Changes per FDA comments dated 01March2023. 2. Updated eCRF and Clinical Study Protocol versions and dates 3. Analysis Visit Windows (Table 1) modifications for appropriate treatment exposure measurement. 4. Changes from "Actual stratification value from the clinical database will be used for analysis if there is a stratification error in the randomization process." and updated as "The stratification information from IRT will be used in efficacy analysis." per ITT definition. 5. Add related SAE process. 6. EAIR (event adjusted incidence rate) added for TEAE event rate summary. 7. The SMQ Search Criteria for AESI of [redacted] will be performed as needed. Table 5 moved to the study TFL shells and specifications. 8. Abnormal ECG values added based on FDA guidance.

1.2	30AUG2024	[redacted]	The SAP were updated per Protocol Amendment 4 as follows. 1. Amending the order and content of the secondary objectives to increase the ability to demonstrate efficacy for lung, skin, and overall disease parameters. The statistical significance for the primary and secondary end points will be evaluated in the hierarchical order. Additionally, to adapt to new updates in the definition of Revised Composite Response Index in Systemic Sclerosis (CRISS 25). 2. Clarifying when all assessments must be completed to ensure study completion before subsequent treatment if joining another study. 3. Updating Physician Global Assessment to Clinician Global Assessment throughout the document to align with industry standards. 4. Introducing the revised CRISS (CRISS 25) into the endpoints and associated definitions to align throughout the protocol. 5. Amending statistical language to align with changes in unblinding, simplify descriptions of analyses to be performed, and clarify that the independent data monitoring committee (IDMC) will review futility data with pre-specified IDMC recommendation criteria.

Version	Effective Date	Author of Modification	Summary of Change
			6. Removed dose selection which is no longer applicable. 7. Removing Covid 19 related analysis per FDA feedback for HZNP-HZN-825-303 (IPF). 8. [redacted] 9. All PK related analyses/results will be provided by Amgen PK group separately. 10. [redacted] 11. Using covariates for randomization stratification variables in logistic and Cox models to adjust for confounding effects within a unified modeling framework. 12. The overall statistically significant level will be 0.05 (2-sided).

2 List of Abbreviations

Abbreviation	Definition
ACR-CRISS	American College of Rheumatology-Composite Response Index in Systemic Sclerosis
AE	adverse events
AESI	adverse event of special interest
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BID	twice daily
CGA	Clinician Global Assessment
MDGA	Physician Global Assessment
CMH	Cochran -Mantel-Haenszel
COVID-19	Coronavirus disease of 2019
CS	clinically significant
CSH	heterogeneous compound symmetry
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CV%	coefficient of variation
DBL	database lock
[redacted]	[redacted]
DRM	data review meeting
ECG	electrocardiogram
eCRF	electronic case report form
EAIR	Event Adjusted Incidence Rate
EDC	electronic data capture
ESR	erythrocyte sedimentation rate
[redacted]	[redacted]
FCS	fully conditional specification
FVC %	forced vial capacity percent
GIT	gastrointestinal tract
HAQ-DI	health assessment questionnaire - disability index
[redacted]	[redacted]
hsCRP	high sensitivity C-reactive Protein
ICE	intercurrent event
ICH	International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use

Abbreviation	Definition
ID	identification
IDMC	independent data monitoring committee
ILD	interstitial lung disease
ITT	intention-to-treat
LS	least squares
MAR	missing at random
MCS	mental component score
MDGA	Physician Global Assessment
MedDRA	Medical Dictionary for Regulatory Activities
mmHg	millimeters of mercury
MMRM	mixed model repeated measures
mRSS	modified Rodnan skin score
NCS	not clinically significant
PAH	pulmonary hypertension
PCS	physical component score
PD	protocol deviation
PK	pharmacokinetics
PMM	pattern mixture model
PP	per-protocol
PT	preferred term
PTGA	Patient Global Assessment
PYE	person years of exposure
Q1	first quartile
Q3	third quartile
QD	once daily
RCTC	Rheumatology Common Toxicity Criteria
SAP	statistical analysis plan
SAS	statistical analysis software
SD	standard deviation
[redacted]	[redacted]
[redacted]	[redacted]
SI	International System of Units
SMQ	standard MedDRA query
SOC	system organ class

Abbreviation	Definition
SPP	statistical programming plan
SSc	systemic sclerosis
SSC GIT	scleroderma gastrointestinal tract
[redacted]	[redacted]
SSPRO-18	scleroderma skin patient-reported outcome
TEAE	treatment-emergent adverse event
TOEP	Toeplitz
TOEPH	heterogeneous Toeplitz
[redacted]	[redacted]
VAS	visual assessment scale

3 Purpose

This SAP provides a detailed and complete description of the planned statistical analyses of the study HZNP-HNZ-825-301 to support the Clinical Study Report (CSR).

This SAP complies with the International Council for Harmonization (ICH) E9 'Statistical Principles for Clinical Trials' and E9(R1) 'Statistical Principles for Clinical Trials: Addendum on Estimands and Sensitivity Analysis in Clinical Trials', and is based upon the following study documents:

- Clinical Study Protocol Version 5.0 (dated 22AUG2024)
- electronic Case Report Form (eCRF), 23Mar2023)

All decisions regarding the final analysis of the study results, as defined in this SAP, have been made before database lock (DBL) of the study data.

Deviations from the analyses in this SAP will be detailed in the CSR.

4 Study Design

This trial will be conducted at approximately 135 sites globally.

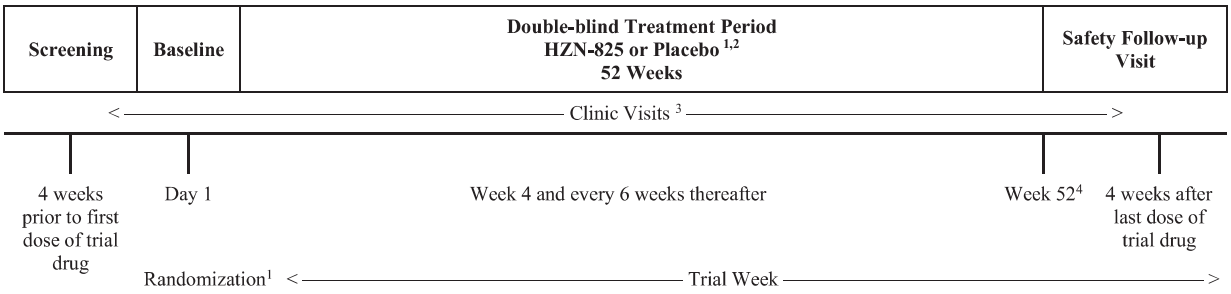
This is a randomized, double-blind, placebo-controlled, repeat-dose, multicenter trial. Subjects will be screened within 4 weeks prior to the Baseline (Day 1) Visit. Approximately 300 subjects who meet the trial eligibility criteria will be randomized on Day 1 in a 1:1:1 ratio to receive fipaxalparant (HZN-825) 300 mg once daily (QD), HZN-825 300 mg twice daily (BID) or placebo for 52 weeks. Randomization will be stratified according to Screening use of mycophenolate mofetil (yes/no) and presence of interstitial lung disease (ILD) (yes/no) based on a Screening [redacted] scan.

The trial will include up to a 4-week Screening Period and a 52-week Double-blind Treatment Period. Subjects will take their first dose of trial drug at the clinic and will return to the clinic for trial visits at Week 4 and every 6 weeks thereafter until Week 52. All subjects who complete the Double-blind Treatment Period (Week 52) will be eligible to enter a 52-week extension trial (HZNP-HZN-825-302). All Week 52 assessments should be performed before the first dose of the extension trial drug is administered. Subjects not entering the extension trial will return to the clinic for a Safety Follow-up Visit 4 weeks after the last dose of trial drug.

A futility analysis will be conducted after approximately 50% of the subjects reach Week 52 or discontinue the trial before Week 52 and these unblinded efficacy and safety data will be provided to an independent data monitoring committee (IDMC). The IDMC will also review accumulating safety data to ensure subject safety at the pre-specified time points during the trial conduct.

An overview of the trial design is presented in Figure 1.

Figure 1 Schematic of Trial Design



BID=twice daily; QD=once daily

1. Subjects will be randomized in a 1:1:1 ratio to receive HZN-825 300 mg QD, HZN-825 300 mg BID or placebo.
2. If a subject prematurely discontinues trial drug, he/she will be asked to remain in the trial, participating in the scheduled trial visits through Week 52. If a subject prematurely discontinues trial drug and does not wish to continue in the trial, he/she will return for a clinic visit and undergo the Week 52 assessments. Subjects not entering the 52-week extension trial (HZNP-HZN-825-302) will return to the clinic 4 weeks after the last dose of trial drug for a Safety Follow-up Visit.
3. Visit windows are ±3 days for Week 4, ±5 days for Week 10 to Week 52, inclusive, and ±14 days for the Safety Follow-up Visit.
4. Subjects who complete the Double-blind Treatment Period (Week 52) will be eligible to enter a 52-week extension trial (HZNP-HZN-825-302).

4.1 Objectives and Endpoints

The overall objective is to investigate the efficacy, safety and tolerability of 2 dose regimens of HZN-825, a selective antagonist of $LPAR_1$, administered QD or BID for 52 weeks in the treatment of subjects with diffuse cutaneous systemic sclerosis (SSc).

4.2 Primary Objective

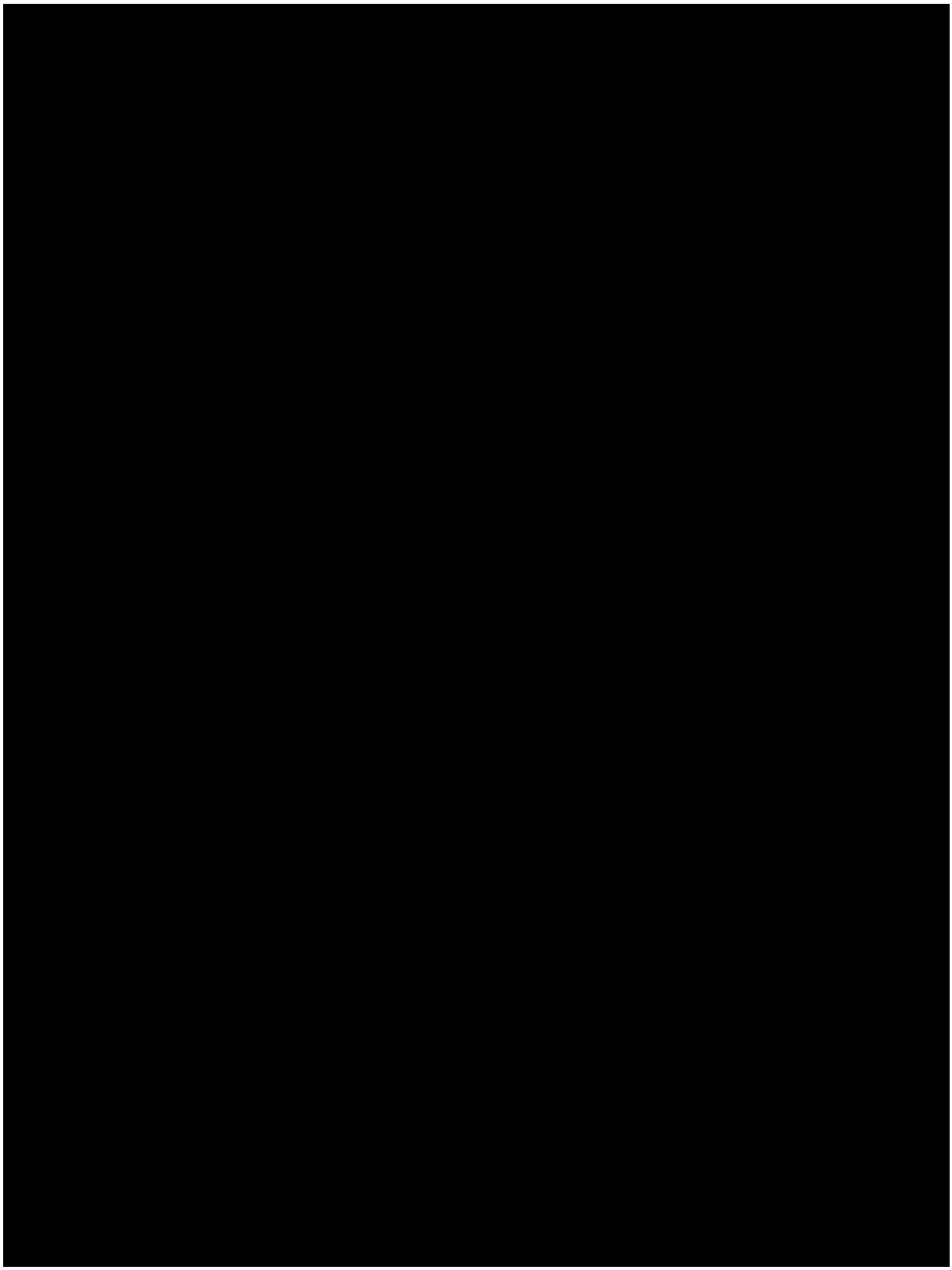
The primary objective is to demonstrate the efficacy of 1 or 2 dose regimens of HZN-825 versus placebo in subjects with diffuse cutaneous SSc, as determined by a comparison of change in FVC % predicted after 52 weeks of treatment.

4.3 Secondary Objectives

1. Evaluate the effect of 2 dose regimens of fipaxalparant (HZN-825) versus placebo on the modified Rodnan skin score (mRSS), after 52 weeks of treatment.

2. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on the Revised Composite Response Index in Systemic Sclerosis (Revised CRISS [CRISS 25]) after 52 weeks of treatment.
3. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on the Health Assessment Questionnaire – Disability Index (HAQ-DI) after 52 weeks of treatment.
4. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on Clinician Global Assessment (CGA) after 52 weeks of treatment.
5. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on Patient Global Assessment (PTGA) after 52 weeks of treatment.
6. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on the Physical Effects subscale of the scleroderma skin patient-reported outcome (SSPRO-18) after 52 weeks of treatment.
7. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on the Physical Limitations subscale of the SSPRO-18 after 52 weeks of treatment.
8. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on proportion of subjects with clinically important change in the mRSS, after 52 weeks of treatment.
9. Evaluate the effect of 2 dose regimens of HZN-825 versus placebo on American College of Rheumatology-Composite Response Index in Systemic Sclerosis (ACR-CRISS), defined as improvement from Baseline in mRSS, HAQ-DI, PTGA, CGA and FVC % predicted, after 52 weeks of treatment.
10. Assess safety and tolerability of HZN-825 based on adverse events (AEs), the adverse event of special interest (AESI) ([redacted]), concomitant medication use, vital signs, 12-lead electrocardiogram (ECG) and clinical safety laboratory evaluations (hematology, chemistry, lipids, coagulation panel and urinalysis).
11. Evaluate the PK of HZN-825 and metabolite(s).

[redacted]



4.5 Study Treatments

On Day 1 of the Double-blind Treatment Period, subjects will be randomized in a 1:1:1 ratio to receive:

1. HZN-825 300 mg QD, or
2. HZN-825 300 mg BID, or
3. Placebo.

Subjects will take 2 tablets (150 mg each or matching placebo) of trial drug orally in the morning and evening with a meal.

4.6 Randomization Procedures and Blinding

4.6.1 Method of Assigning Patients to Treatment Groups

A randomization schedule will be generated by an unblinded statistician not otherwise associated with the trial prior to shipment of any trial drug to the clinical sites. On Day 1 of the Double-blind Treatment Period, once all Baseline procedures other than administration of trial drug have been completed, authorized site personnel will use the Interactive Response Technology (IRT) system to randomize the subject (this is different from what is specified in the protocol, as electronic data capture [EDC] was originally planned to be used, but the decision was made to use IRT instead).

Subjects will be randomized in a 1:1:1 ratio to receive HZN-825 300 mg QD, HZN-825 300 mg BID, or placebo. The randomization will be stratified by Screening use of mycophenolate mofetil (yes/no) and presence of interstitial lung disease (ILD) (yes/no) based on a Screening [redacted] scan. The Investigator or designee will then use the EDC system to obtain dosing information and dispense the appropriate trial drug.

4.6.2 Blinding and Unblinding

Placebo tablets match the appearance of active tablets. HZN-825 150 mg and placebo tablets will be packaged in blinded blister packs. The subject, Investigator, and all other trial site personnel, including Sponsor or designee monitors, will be blinded to the trial drug being administered. Subjects assigned to 300 mg QD will receive two 150 mg HZN-825 tablets in the morning and two placebo tablets in the evening. Subjects assigned to 300 mg BID will receive two 150 mg HZN-825 tablets each in the morning and in the evening. Subjects assigned to placebo will receive two placebo tablets in the morning and two placebo tablets in the evening.

The trial blind should be broken only if the safety of a subject is at risk and the treatment plan depends on which trial drug he or she received.

4.7 Determination of the Sample Size

Based on prior trials of tocilizumab [Khanna et al., 2016; Khanna and Lin et al., 2020] and abatacept [Khanna and Spino et al., 2020] in a similar subject population, change in FVC % predicted is expected to have a standard deviation of 8 to 8.5 percentage points after 52 weeks of treatment. Assuming a clinically important difference between HZN-825 and placebo of 4 percentage points and a standard deviation of 8.5, a sample size of 100 subjects per treatment group will provide 85% power to demonstrate an improvement between a dose regimen of HZN-825 and placebo using α=0.025, 2-sided.

4.8 Planned Futility Analyses and Reviews

When the first half of the planned total number of subjects (150 of 300) have been followed for the planned duration of 52 weeks or have permanently discontinued the study, a futility analysis will take place. This will require final disposition on the first 150 subjects randomized to be complete and all data on these subjects to be entered and clean before the analysis occurs. Data on enrollment, disposition, exposure, efficacy and safety will be summarized. Specific data and formats will be agreed between the sponsor and IDMC members before the futility analysis.

The conditional power for concluding superiority of each posology of HZN-825 over placebo will be calculated, as defined in Section 4.8.1. The calculations will be provided for the conditional power based on current trend [Lan et al., 1988], assuming the treatment effect observed thus far in the study, as well as for the hypothesized trend. The calculation most emphasized for each dose will be the conditional power which assumes that the current trend in efficacy continues. The IDMC may also consider conditional power assuming the hypothesized efficacy, if it is markedly different and there is a reason to believe that the conditional power assuming the current trend is biased.

All available data on the first 150 subjects randomized will be used for the conditional power calculation.

The IDMC will recommend one of the following options:

(1) **Stop the study for futility** if the conditional power for both doses is less than 10%.

(2) **Continue the study** if both BID and QD have conditional power>=10% and placebo adjusted FVCpp change from baseline at Week 52 > 5.0% with at least one of the doses has mRSS placebo adjusted reduction from baseline at Week 52 > 4.1 (Assuming the true reduction is 5 with SD=8.9.)
(3) **Deliberate the results** if there is uncertainty about whether to stop or continue the study.

If the IDMC recommends stopping the study for futility or deliberating the results, an unblinded internal team will be established in accordance with an internal data access plan (DAP) (SOP-433026) to review the unblinded data. This team will confirm the IDMC's recommendation and conduct ad hoc analyses to evaluate the findings, ultimately making the final decision on whether to continue or terminate the study. The IDMC may also recommend stopping only one posology only if there is a clear safety signal that precludes continuation of one posology. Unless this occurs, the IDMC will recommend continuing or discontinuing both posologies in the study based on the above rules. The interim analysis is not designed for the purpose of stopping the trial earlier based on efficacy. There will be no possibility of stopping the study early for efficacy.

4.8.1 Calculation of Conditional Power

At the interim, the conditional power for the current trend and the hypothesized trend will be calculated [Lan et al (1988)]. The conditional power will be presented for the comparison of each dose regimen to the placebo. The alpha is selected based on a conservative Bonferroni 0.025 even split.

Then Conditional Power at time *t* is given by:

Conditional Power = $\Phi[Z_{CP}]$,

where Φ is the standard normal cumulative distribution function and

Current Trend: $Z_{CP}(t)=\dfrac{\frac{Z_t}{\sqrt{t}}-Z_{\alpha/2}}{\sqrt{1-t}}$ at t=0.5 and α = 0.025

Hypothesized (H1) trend: $Z_{CP}(t)=\dfrac{\left(Z_t*\sqrt{t}\right)+(1-t)*\left(Z_{\alpha/2}+Z_\beta\right)-Z_{\alpha/2}}{\sqrt{1-t}}$

at t=0.5 and α = 0.025, β = 0.150

The statistic Z_t will be calculated as Z_t = (LS Mean Difference)/Standard Error, with these values obtained from the SAS statistical output.

4.8.2 Independent Data Monitoring Committee (IDMC)

An external IDMC will be convened to review data for safety and efficacy, with the possibility of the IDMC recommendation on trial design modification per a pre-defined IDMC charter.

The IDMC will include at least 1 statistician and at least 2 clinicians experienced in clinical trials and the therapeutic area. The IDMC charter has included decision rules for stopping the trial based on the comparative, unblinded summary of safety and efficacy.

5 Changes from the Protocol Planned Analyses

The multiplicity handling approach for primary and secondary efficacy endpoints has been updated from the graphical testing procedure to the Hochberg testing procedure. Since the Hochberg method is more powerful when the intention is to show significance in one endpoint at a time, this change to the multiplicity approach allows the study result to prioritize a positive outcome for the primary endpoint more directly. In addition, there is not a strong endpoint correlation between the primary and key secondary endpoints, so the graphical approach is less advised in this situation.

[redacted]

6 Study Analysis Sets

6.1 ITT Analysis Set

The intent-to-treat (ITT) analysis set will include all subjects who are randomized to treatment regardless of whether they received study drug or not. All subjects who are randomized will be included in the primary efficacy analyses (ITT analysis set).

6.2 Safety Analysis Set

The Safety analysis set consists of all subjects who received any amount of study drug after being randomized into the study. This analysis set will be analyzed according to the treatment actually received. The safety analysis set is the primary analysis set for safety evaluation.

The subjects' actual treatment (dose regimen) will be derived from the Exposure data set.

- A subject who received only placebo will be classified as a subject in the placebo group.
- A subject who received any active treatment will be classified as a subject in the active treatment group.

A subject who received different active doses at different dosing visits during the study will be included in the group of the highest active dose that he/she received. A supplemental analysis will be added for key safety analysis parameters based on randomized dose groups if there are more than 5% patients with dose error in the trial. The risk differences for the comparisons between treatment and placebo groups with their 95% confidence intervals will be calculated as needed.

6.3 Per Protocol Analysis Set

The Per-Protocol (PP) analysis set consists of all ITT subjects with no major protocol deviations that compromise the evaluation of efficacy and safety. The protocol deviation plan pre-specifies which major deviations qualify a subject for exclusion from the PP analysis set. Protocol deviations resulting in exclusion from the PP analysis set will be documented in the data review meeting (DRM) minutes before DBL unblinding.

The PP analysis set will be used for supportive analysis of the primary endpoint and the first secondary endpoint.

6.4 Pharmacokinetic Analysis Set

The PK analysis set will include all subjects who receive at least 1 dose of HZN-825 and have at least 1 PK sample post HZN-825 treatment. This analysis set will be analyzed using the treatment as the subject actually received.

7 General Considerations

SAS version 9.4 or higher will be used to perform all data analyses.

Summaries of continuous variables will be in terms of the number of observations, mean, standard deviation, median, first quartile (Q1), third quartile (Q3), minimum and maximum. Other descriptive statistics (e.g. standard error, coefficient of variation) may be reported when appropriate. Categorical variables will be summarized using frequency counts and percentages. Analyses that use other descriptive statistics will have the specific descriptive statistics identified with the analysis in the applicable SAP section. All confidence intervals,

statistical tests, and P-values will be reported as 2-sided and will be assessed at 2-sided 5% significance level unless otherwise specified.

8 Data Handling

8.1 Study Day

Study day will be calculated as (date of interest - treatment start date) + 1 if the date of interest occurs on or after the treatment start date. If the date of interest occurs before the treatment start date, then the study day will be calculated as (date of interest – treatment start date). There will be no study day zero.

8.2 Durations and Time to Event Data

Durations are calculated in days as:

- event end date – event start date + 1, if end time or start time not available.
- event end date / time – event start date / time, if both end time and start time available.

Thus, there will be no duration of 0 if end time or start time are not available. If an event has missing or partially missing start or end date, no duration will be calculated.

For elapsed time (e.g. the time to event), use:

- event date / time – reference date /time, (if time available).

Thus, an event which happens on the same date as the reference date will have an elapsed time of 0, if event time or reference time are not available.

8.3 Baseline Definition

Unless otherwise specified, baseline is defined as the last non-missing observation prior to the first dose of study drug.

For observations occurring on the same date as the first dose of study drug where time is not collected but the protocol specifies the evaluation must occur prior to administration of study drug, the result occurring on the day of first study drug administration will be considered baseline.

8.4 Visit Window

For endpoints that present visit-based data, the variables will be summarized based on the scheduled visits with derived analysis visit windows. No visit windows will be derived for

the screening period. For the visits on and after the 1st dose date, the actual visit date will be mapped to the derived analysis visit windows based on the study day (see Section 8.1).

Visit windows have been constructed so that every observation (unscheduled visits included) collected can be allocated to a specific visit. The actual assessment day will be mapped to the windows defined for each scheduled study visit with following rules:

- If more than 1 assessment falls within a visit window, the closest non-missing valid assessment to the scheduled day will be used in the analysis.
- If 2 non-missing assessment actual dates are equidistant from the target day, the later visit will be used in the analysis.
- For retest values of laboratory and spirometry data, the retest value (the last valid observation assessed corresponding to the same visit) will be chosen.

The Analysis Visit Window for the primary endpoint FVC % predicted and key secondary endpoint mRSS are as indicated in Table 1. Analysis visit windows for other endpoints that are to be collected at different time points will be also defined and provided in the table, listing, and figure (TLF) shell document.

Table 1 Analysis Visit Windows

Scheduled Visit	Target Day of the Visit	Analysis Visit Window
Analysis visit windows for primary analysis		
Baseline	Day 1	Day 1
Week 16	Week 16 (Day 113)	Day 2 to Day 155
Week 28	Week 28 (Day 197)	Day 156 to Day 239
Week 40	Week 40 (Day 281)	Day 240 to Day 323
Week 52	Week 52 (Day 365)	Day 324 to Day 407 or last value after Day 323

Data from all assessments (scheduled and unscheduled), including multiple assessments, will be included in listings.

8.5 Stratification

Randomization to dose regimen will be stratified according to Screening use of mycophenolate mofetil (yes/no) and presence of interstitial lung disease (yes/no) based on a Screening [redacted] scan. The stratification information from IRT will be used in efficacy analysis.

9 Study Population

9.1 Subject Disposition

The following summaries will be provided by dose regimen:

Subject disposition presenting the number and percentages of:

- subjects screened (count only; no percentage is calculated)
- subjects ineligible/failed screening
- subjects completed Double-Blind Treatment Period
- subjects discontinued study drug during Treatment Period (before Week 52) and reasons for discontinuation
- subjects discontinued study during Treatment Period (before Week 52) and reasons for discontinuation
- subjects completed Safety Follow-up Visit (Subjects not entering extension study)
- subjects who entered the Extension Trial
- subjects available by visit (Treatment Period), a subject will be considered to have attended a visit if a visit date is recorded

The number of subjects screened will be the denominator for calculation of the number of subjects ineligible/failed screening. For all other disposition percentages, including the analysis sets below, the number of subjects randomized into the treatment group will be used as the denominator for the calculation.

Number and percentage of subjects in the analysis sets:

- ITT Analysis Set
- Safety Analysis Set
- PP Analysis Set
- PK Analysis Set

Subject study duration in weeks will be summarized for subjects in ITT by treatment group and overall. Subject study duration will be calculated from the Day 1 visit to the end of study date, on which the end of study assessment (i.e., Week 52) is performed. The number (percent) of subjects with results for each scheduled visit will also be presented. Visit windowing will not be applied for the summary and a subject will be considered to have completed a visit if they had any results recorded in the eCRF for the visit.

The following by-subject listings will be provided by randomized treatment group and by subject identification (ID) number in ascending order to support the above summary tables:

- Reasons for screen failure (will be provided by screening ID number in ascending order)
- A disposition listing will be provided including date of informed consent, date of randomization, date of first treatment (Treatment Period), study treatment completion, reason if not completed, study completion, reason if not completed, Follow-up Visit date, and date of enrollment to Extension Trial
- Assignment to analysis sets with reasons for exclusion.

9.2 Protocol Deviations

Protocol deviations (PDs) occurring after subjects entered the trial are documented during routine monitoring. The PDs are reviewed and categorized as major and minor prior to the database lock. The number and percentage of subjects with major protocol deviations by deviation reason (e.g., nonadherence to study drug, violation of select inclusion/exclusion criteria) will be summarized by treatment group per ITT. A by-subject listing will be provided for all protocol deviations.

A by-subject listing will be provided for those subjects who did not meet at least 1 eligibility (inclusion or exclusion) criterion. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that subjects did not meet and related comments, if collected. The following listings will be provided:

- All inclusion and exclusion criteria protocol deviations
- All other protocol deviations

9.3 Demographic and Baseline Characteristics

Descriptive summaries of demographic and baseline characteristics will be presented overall and by dose regimen for the ITT and Safety Analysis Set. These characteristics include age, sex, race, ethnicity, height, weight, body mass index, region, and child-bearing potential (yes, no, not applicable). The use of mycophenolate mofetil (yes/no) and the presence of interstitial lung disease (ILD) (yes/no) based on a Screening [redacted] scan will also be summarized, based on the values collected in the clinical database. If the values that come from the IVRS system are different from what's in the clinical database, the IVRS values will be summarized separately.

Systemic sclerosis history variables will include: time since SSc diagnosis (months), whether the patient has symptoms of SICCA syndrome , time with [redacted] (months), time since

first [redacted] (months), [redacted] (% predicted), incidence of subjects with gastrointestinal medical history (based on the SOC "gastrointestinal disorders"), incidence of the following symptoms: [redacted] as well as other related information collected. Systemic sclerosis history will be summarized by treatment group for the ITT Analysis Set.

Demographic data and Baseline characteristics will be provided in subject listings.

9.4 Medical History

Medical history information will be coded using the most recent version of Medical Dictionary for Regulatory Activities (MedDRA), summarized and presented overall and by treatment group based on the Safety Analysis Set. Summaries will be ordered alphabetically by system organ class (SOC) and then, within a SOC, alphabetically by preferred term (PT). Medical history will be provided in subject listings.

9.5 Medications

Prior and concomitant medication verbatim terms in the eCRFs will be mapped to Anatomical Therapeutic Chemical (ATC) level 4 and preferred terms using the WHO Drug Global B3 Mar2020. Prior and concomitant medications will be summarized by presenting the counts and percentage of subjects using medications overall and by each treatment group for the Safety Analysis Set. Summaries will be provided by ATC Level 4 term and PT. Medication summaries will be sorted alphabetically by ATC Level 4 and by PT within ATC Level 4. Subjects will be counted only once for each medication class and each preferred drug name.

Prior and concomitant medications will be listed together with a designation to identify the medications as prior and/or concomitant and sorted by start date.

9.5.1 Prior Medications

Prior medications will be presented separately from concomitant medications in a summary table. Any medication with a stop date prior to the date of first dose will be considered a prior medication.

9.5.2 Concomitant Medications

Any medication that is ongoing, has a start date on or after the first dose of study drug, or a stop date on or after the first dose date will be considered a concomitant medication.

Prior and concomitant medications will be summarized for each treatment by ATC level 4 term and preferred name. These summaries will present the number and percentage of subjects using each medication. Subjects may have more than 1 medication per ATC class and preferred name but will only be counted once in the summary.

Prior and concomitant medications and concomitant procedures will be listed for all subjects.

10 Efficacy Analyses

All efficacy analyses will be analyzed using the ITT Analysis Set. The PP analysis set will be used for supportive analysis of the primary endpoint and the first secondary endpoint.

10.1 Primary Estimand

The primary objective is to demonstrate the efficacy of 1 or 2 dose regimens of HZN-825 versus placebo in subjects with diffuse cutaneous SSc, as determined by a comparison of the change in FVC% predicted after 52 weeks of treatment.

The estimand framework is summarized for the primary endpoint in Table 2.

Table 2 Estimand Framework for Primary Endpoint

Endpoint	Treatment	Population	Intercurrent Events (ICE)	Population-Level Summary
Change from Baseline in FVC % predicted at Week 52	HZN-825 300 mg BID, HZN-825 300 mg QD and Placebo	ITT Analysis Set (all randomized subjects)	The intercurrent events (ICEs) which may impact analysis results include rescue medicine use, early treatment discontinuation, adverse events leading to missing data at the visit, death, early study discontinuation or subjects being lost to follow-up. The treatment policy strategy will be used for these intercurrent events. All observed data will be used. A mixed model for repeated measures (MMRM) will be fit to the data for the observed changes in FVC % predicted values from all planned	Point estimate (The least squares mean (LS mean) difference for the change from Baseline to Week 52 based on MMRM mixed model with its 95% confidence interval and P-value.

Endpoint	Treatment	Population	Intercurrent Events (ICE)	Population-Level Summary
			post-baseline visits (Weeks 16, 28, 40 and 52). The least squares mean (LS mean) difference for the change from Baseline at Week 52 will be estimated from this model.	

10.2 Primary Efficacy Analysis

The estimand for the primary efficacy analyses is constructed to compare the primary endpoint between each dose regimen of HZN-825 and placebo, using the treatment policy strategy approach to intercurrent events. All subjects who are randomized will be included in the primary efficacy analyses (ITT).

The primary efficacy endpoint will be the change from Baseline in FVC % predicted at Week 52. FVC% predicted is collected at Screening, Day 1, Week 16, Week 28, Week 40, and Week 52.

The primary analysis will be based on a mixed model for repeated measures (MMRM) analysis of covariance (ANCOVA) model using observed change in FVC % predicted values from all planned post-baseline assessments (Weeks 16, 28, 40 and 52) and including effects for baseline FVC% predicted, treatment (300 mg HZN-825 QD, 300 mg HZN-825 BID, placebo), the Baseline factors used for stratifying randomization as covariates (use of mycophenolate mofetil [yes/no] and presence of ILD [yes/no]), visit week, and treatment by visit week interaction. An unstructured covariance matrix will be used for the primary analysis; if the model does not converge, the following three variance-covariance matrices or by visit week will be attempted in order until one converges: heterogeneous Toeplitz (TOEPH), heterogeneous compound symmetry (CSH), and Toeplitz (TOEP).

For the comparison between each dose regimen and placebo, the point estimate (the least squares mean (LS mean) difference for the change from Baseline at Week 52, with its 95% confidence interval and P-value will be provided. Similar analyses will be presented for the other post-baseline time points.

A line plot of the least squares mean value for FVC% predicted change from Baseline will be plotted by treatment group for each visit week over time, including the estimated 95% CI around each LS mean.

10.3 Sensitivity Analyses of the Primary Endpoint

Sensitivity analyses will be used to support the results from the primary analysis for the primary endpoint.

The intercurrent events (ICEs) which may impact analysis results include: rescue medicine use, early treatment discontinuation, adverse events leading to missing data at the visit, death, early study discontinuation, or subjects being lost to follow-up.

The primary analysis is based on the treatment policy strategy for intercurrent events, which reflects the ITT principle and assumes the treatment in real world conditions. Under the treatment policy, the data will be assessed as data collected regardless of whether an intercurrent event has occurred or not. Sensitivity analyses will be performed to evaluate the robustness of the primary analysis results based on different assumptions for the ICEs. The key ICEs are the ones leading to missing data during the study.

Multiple imputation with the fully conditional specification (FCS) method will be used to handle missing FVC % predicted data at Week 52. The imputation model will be similar to the statistical model of the primary analysis. Two sensitivity analyses will be performed for subjects with missing FVC % predicted data at Week 52. Additional sensitivity analysis will be performed for only subjects who took rescue therapy.

First, the hypothetical strategy will be used to support those subjects with missing data at Week 52 due to intercurrent events. For these subjects, we would assume that their values would be like the ones without missing data at Week 52 for each treatment group respectively, and that the treatment effect is preserved.

Multiple imputation will be used as the column of Imputation Cohorts [A] in Table 3. Similar to the primary analysis, the sensitivity analysis will include observed data with the imputed FVC % predicted data at Week 52 under the hypothetical strategy. There are 12 patterns and imputation rules listed in Table 3.

Table 3 Imputation Rules

Cohort	Treatment	On treatment Before WK 52	%FVC predicted During the study	Imputation Cohorts [A]	Imputation Cohorts [B]
1	HZN BID	Yes	Non-Missing		
2	HZN QD	Yes	Non-Missing		

Cohort	Treatment	On treatment Before WK 52	%FVC predicted During the study	Imputation Cohorts [A]	Imputation Cohorts [B]
3	Placebo	Yes	Non-Missing		
4	HZN BID	No	Non-Missing		
5	HZN QD	No	Non-Missing		
6	Placebo	No	Non-Missing		
7	HZN BID	Yes	Missing	1	3, 6
8	HZN QD	Yes	Missing	2	3, 6
9	Placebo	Yes	Missing	3, 6	3, 6
10	HZN BID	No	Missing	4	3, 6
11	HZN QD	No	Missing	5	3, 6
12	Placebo	No	Missing	3, 6	3, 6

A pattern mixture model (PMM) with the FCS imputation method will be used to impute missing FVC % predicted data at Week 52. Missing data will be imputed based on Baseline, previous FVC % predicted data, treatment, and stratification factors used in randomization. One thousand imputed datasets will be created with results from the analysis of each imputed dataset combined using Rubin's method via PROC MIANALYZE.

SAS code example is as below for missing data imputation in cohorts [A] (1,2,3,6,7,8,9).

```
proc mi data=FVCpp nimpute=1000 seed=123456 out=FVCppOut;
    where cohort in (1,2,3,6,7,8,9);
    class Stratification TRT;
        var WK16 WK28 WK40 WK52 Baseline Stratification TRT;
        fcs reg (WK52=Baseline WK16 WK28 WK40 Stratification TRT);
    run;
```

Another sensitivity analysis will be performed for subjects with missing FVC % predicted data at Week 52 under the hypothetical strategy as the column of Imputation Cohorts [B] showed in Table 3. The assumption for this analysis is that subjects with intercurrent events lose any treatment effect after the last non-missing data. As such, subjects with missing data at Week 52 would be assumed to be like those subjects in the placebo group without missing data in the study after last-non-missing data. Similar to the primary analysis, the sensitivity

analysis will include observed data with the imputed FVC % predicted data at Week 52 based on placebo subjects with non-missing data.

Tipping point analyses will be performed to evaluate the robustness of the primary analysis to the missing at random (MAR) based on primary analysis method. In the tipping point analysis, data from HZN-825 and placebo groups will be imputed under Missing Not at Random (MNAR) assumption. First, the missing values in each treatment group will be imputed separately based on observed values in each group, respectively. Then adjustments for each treatment group (placebo and HZN-825 dose selected for Phase 3) will be added to the imputed data and vary to find conditions with non-significant treatment effect.

Two additional sensitivity analyses will be done for missing data due to death before Week 52 (See Table 4) because it is an important assessment of the treatment efficacy.

Table 4 Sensitivity Analysis for Missing Data at Week 52 Due to Death

Analysis	Missing Week 52 data in subjects but still alive at Week 52		Missing Week 52 data due to death before/at Week 52	
	Handling of missing data at Week 52	Assumption for treatment effect after last non-missing data	Handling of missing data at Week 52	Assumption for treatment effect after last non-missing data
Primary	No imputation	Treatment policy strategy - Assuming missing at random	No imputation	Treatment policy strategy - Assuming missing at random
Sensitivity 1	A pattern mixture model (PMM) with the fully conditional specification (FCS) imputation method will be used for missing Week 52 data based on non-missing data from the treatment and placebo subjects respectively	Hypothetical strategy - Assuming subjects with missing data would be like those ones without missing data in the study for the treatment and placebo subjects respectively	Use the worst observed value of the FVC % predicted in the study to impute missing Week 52 data, which will be more severe outcome than that in subjects who survived	Hypothetical strategy – Assuming subjects with death have the worst outcomes in the study
Sensitivity 2	A pattern mixture model (PMM) with the fully conditional specification (FCS) imputation method will be used for missing data at Week 52 based on non-missing data from the placebo subjects with non-missing data only	Hypothetical strategy - Assuming subjects with missing data after last-non-missing data would be like those ones in placebo without missing data in the study		

There will not be specific data handling and data imputation rules for rescue medicine use in the primary analysis assuming patients with rescue medicine will have the same treatment effect during the study. If the treatment is efficacious, the analysis results with this assumption may underestimate the treatment effect in the primary analysis because placebo patients may have more rescue medication use and get help to prevent the further decline in FVC % predicted change from Baseline at Week 52. However, if the treatment is not efficacious, the rescue medicine use will be observed similarly among treatment and placebo subjects and there will be less likely to have biased trend in the results.

To assess the potential impact of rescue medicine use in the primary analysis, A sensitivity analysis will be performed for rescue medicine used for improving FVC per the hypothetical strategy. The observed data will be used and analyzed like the primary analysis but the observed data after start of rescue medicine use will not be included in the analysis assuming

this intercurrent event hypothetically does not exist and the time of rescue medicine use indicates the end of data collection. Supplementally, the proportion of subjects who qualify for and who receive rescue medication at each potential time point will be summarized for each treatment group to determine whether there are different use rates between treatment groups.

As a supplemental analysis to assess the robustness of the primary analysis results, the effect of each dose regimen of HZN-825 will be compared to placebo using the following measurements and methods:

- Change in FVC (in mL) from Baseline to Week 52 using a similar MMRM analysis method as described for the primary endpoint
- Cumulative response curve comparison using Cochran-Mantel-Haenszel (CMH) test on:
    - The range of improvement (≥0% to ≥15% with a 5% increment) and worsening (< -15% to <0% with a 5% increment) based on the change in FVC % predicted value from Baseline at Week 52
    - The range cut-off values from less than -300 mL to greater 100 mL with a 100-mL increment based on the change in FVC (in mL) from Baseline at Week 52
- The time to the first FVC % predicted decline >=10% from Baseline or death will be explored using a Cox proportional hazard model including Baseline FVC % predicted, treatment, and the randomization stratifications (use of mycophenolate mofetil [yes/no] and presence of ILD [yes/no]).

Additional sensitivity analyses for missing data evaluation may be performed as needed.

10.4 Secondary Efficacy Endpoint Analyses

As described in Section 10.7, analysis of the secondary endpoints will be conducted in the order they are listed below, with statistical significance for an endpoint within a dose declared only if the preceding endpoint was declared significant within that dose.

The estimands for the secondary efficacy analyses are constructed to compare the treatment effect between each dose regimen of HZN-825 and placebo, using the treatment policy strategy approach to intercurrent events. All subjects who are will be included in the secondary efficacy analyses (ITT).

10.4.1 Change from Baseline in mRSS at Week 52

The key secondary endpoint for the trial will be change in mRSS from Baseline to Week 52.

Analysis will follow that of the primary analysis of the primary efficacy endpoint, except that the analysis will be based only on observed data. Statistical significance on change in mRSS at a given dose will be concluded based on Hochberg testing procedure as described in Section 10.7.

10.4.2 CRISS-25 Responder Rate at Week 52

The proportion of subjects responding to treatment based on CRISS 25 at Week 52 will be evaluated using logistic model. The Revised CRISS (CRISS 25) is defined as improvement in at least 2 components: ≥5% increase for FVCpp and/or ≥25% decrease for mRSS, HAQ-DI, PTGA, CGA (MDGA) and worsening in no more than one component: ≥5% decrease percent predicted FVC and/or ≥25% increase for mRSS, HAQ-DI, PTGA, CGA (MDGA), at 52 weeks. Revised CRISS (CRISS-25) is also a 2-step process. If the subject meets ACR-Criss Step 1 (as defined in Section 10.4.9), they are considered not improved, given a percentage change of 0% for each core set item. In Step 2, the five core set measures are individually collected and scored.

10.4.3 Change from Baseline in HAQ-DI at Week 52

The HAQ-DI from Baseline to Week 52 will be a secondary endpoint. The HAQ-DI assesses the subject's level of functional ability and includes questions of fine movements of the upper extremity, locomotor activities of the lower extremity, and activities that involved both upper and lower extremities. There are 20 questions in 8 categories of functioning including dressing, rising, eating, walking, hygiene, reach, grip and usual activities [Cole et al., 2006]. The subject's ability to accomplish each activity in the past week is indicated as: without any difficulty, with some difficulty, with much difficulty, and unable to do. Any devices that are usually used to complete activities and any categories for which help form another person is needed are also assessed.

In addition, the subject will be asked the following anchor (additional) questions at the time points specified in Section 2.1 of the protocol:

How did your systemic sclerosis limit your daily activities in the last week?

- 0: did not limit activity
- 1: mildly limited my activity
- 2: moderately limited my activity
- 3: severely limited my activity
- 4: very severely limited my activity

How has limitation of your daily activities changed since the start of the trial?

- +3: very much less limited
- +2: much less limited
- +1: a little less limited
- 0: no change
- -1: a little more limited
- -2: much more limited
- -3: very much more limited

The HAQ-DI value itself is derived from a combination of the scores of the 8 categories of function and use of devices to result in one disability index, which ranges from 0 to 3 [Bruce B, Fries JF et al, 2003]. The HAQ-DI is calculated by scoring the answer to each question in the HAQ from 0 to 3, with 0 representing the ability to do without any difficulty, and 3 representing inability to do. Any activity that requires assistance from another individual or requires the use of an assistive device raises a 0 or 1 score to a 2. The highest score for each of the 8 domains is summed (range from 0 to 24) and divided by 8 to yield, on a scale with 25 possible values, a Functional Disability Index with a range from 0 to 3.

If a domain has all responses missing (including aids/devices) then the domain is considered missing. If some of the questions have a result and/or the aids/devices is checked, then use available values to calculate a domain score. The disability index is based on the number of domains answered and is computed only if the subject completes answers to at least 6 domains. That is, if 6 domains are non-missing, then the average of the 6 available scores will determine the Functional Disability Index. If < 6 domains are non-missing then the HAQ-DI score will be missing.

If "Other" option is checked in either of the "AIDS AND DEVICES" sections of the questionnaire, the corresponding "Other, Specify" field will be reviewed and categorized by Horizon Therapeutics into an appropriate domain of function, so it can be incorporated into the score.

Analysis will follow that of the primary analysis of the primary efficacy endpoint, except that the analysis will be based only on observed data. Statistical significance on change in HAQ-DI at a given dose will only be concluded if statistical significance was achieved for the primary endpoint at that dose, as described in Section 10.7.

10.4.4 Change from Baseline in CGA at Week 52

The CGA (Clinician Global Assessment or Physician Global Assessment (MDGA)) is an 11-point scale ranging from 0 to 10 (0=excellent to 10=extremely poor) on which the physician rates the subject's overall health over the past week.

The CGA will be tested in the same manner as that described for the primary analysis of the primary endpoint in Section 10.2, but based only on observed data.

10.4.5 Change from Baseline in PTGA at Week 52

The Patient Global Assessment (PTGA) is an 11-point scale ranging from 0 to 10 (0=excellent to 10=extremely poor) on which the patient rates their overall health over the past week.

The PTGA will be tested in the same manner as that described for the primary analysis of the primary endpoint in Section 10.2, but based only on observed data.

10.4.6 Change from Baseline in the Physical Effects subscale of the SSPRO-18 at Week 52.

The SSPRO-18 is an 18-item, patient-reported outcome instrument that specifically assesses skin-related quality of life in patients with SSc and was developed with extensive patient input and according to the FDA patient-reported outcomes guidance [Man et al., 2017]. The SSPRO-18 comprises 4 major conceptual constructs—physical effects, emotional effects, physical limitations and social effects—and has reproducibility and high internal consistency. This instrument reflects how subjects feel and function from several different health perspectives. Recall is the past 4 weeks. Response options are on a 0-6 point scale where 0 is 'not at all' and 6 is 'very much.' The physical effects or symptoms domain includes 5 items, physical limitations have 4 items, emotional effects include 6 items and social effects has 3 items. A total score and domain specific scores (physical effects, physical limitations, emotional effects, and social effects) can be calculated and transformed to a 0 – 100 scale, with higher scores indicating more severe impact of skin problems on the patient's quality of life. The SSPRO-18 Physical Effects subscale will be tested in the same manner as that described for the primary analysis of the primary endpoint in Section 10.2, but based only on observed data.

Missing items will be imputed according to the instructions for the questionnaire. If the scoring manual for the SSPRO-18 does not indicate how to impute missing data, then the subscale with missing items will be considered missing.

10.4.7 Change from Baseline in the Physical Limitations subscale of the SSPRO-18 at Week 52.

The SSPRO-18 Physical Limitations subscale will be tested in the same manner as that described for the primary analysis of the primary endpoint in Section 10.2, but based only on observed data.

Missing items will be handled as indicated in Section 10.4.6.

10.4.8 Proportion of subjects with an mRSS decrease of ≥5 points and 25% from Baseline at Week 52.

The mRSS is a validated method for estimating skin thickening [Khanna et al, 2017]. Seventeen different body areas are scored as normal (0), mild thickening (1), moderate thickening (2) and severe thickening (3). The scores selected for each body area are summed for a total score, which has a maximum score of 51.

Missing body area scores will be imputed according to the scoring instructions for this tool. If the scoring manual for the mRSS does not indicate how to impute missing data, then the body area with missing items and corresponding total score will be considered missing.

The proportion of subjects with an mRSS decrease in ≥ 5 points and 25% from baseline at Week 52 will be analyzed with observed data using a logistic regression model. The model will include mRSS Baseline, treatment, use of mycophenolate mofetil [yes/no] and presence of ILD [yes/no]) at randomization. The risk differences between each dose and placebo with its 95% confidence intervals will be presented. We will use NRI-MI (Non-Responder Imputation in conjunction with Multiple Imputation) method to impute missing mRSS score at Week 52. The missing values in each treatment group at Week 52 will be imputed separately based on observed values at Week 52 in each group, respectively. Subjects who died before Week 52 will be categorized as non-responders. Inferences will be made if allowed under the algorithm described in Section 10.7.

10.4.9 Responder rate (defined as ACR-CRISS [predicted probability] of at least 0.6) at Week 52.

Subjects will be evaluated using the ACR-CRISS, an outcome measure for diffuse cutaneous SSc. The ACR-CRISS includes core items that assess change in 2 prominent manifestations of early diffuse cutaneous SSc (skin and ILD), functional disability (HAQ-DI) and patient and physician global assessments. In addition, the score captures a clinically meaningful worsening of internal organ involvement requiring treatment.

The ACR-CRISS is a 2-step process that assigns a probability of improvement for a subject that ranges from 0.0 (no improvement) to 1.0 (marked improvement). Step 1 will be evaluated as part of the AE assessment, at which time the Investigator will assess if a subject has developed new or worsening cardiopulmonary and/or renal involvement due to SSc, as outlined below. Step 1 events will be adjudicated.

- New scleroderma renal crisis, defined as follows (adapted from Steen et al., 2003):
  Hypertensive scleroderma renal crisis:
  1. New onset hypertension, defined as any of the following:
     a. systolic blood pressure ≥140 mmHg
     b. diastolic blood pressure ≥90 mmHg
     c. rise in systolic blood pressure ≥30 mmHg
     d. rise in diastolic blood pressure ≥20 mmHg

  AND

  2. One of the following 5 features:
     a. increase in serum creatinine by ≥50% over Baseline OR serum creatinine ≥120% of ULN for local laboratory
     b. proteinuria ≥2+ by dipstick
     c. hematuria ≥2+ by dipstick or ≥10 red blood cells/high-powered field
     d. thrombocytopenia: <100,000 platelets/mm^3
     e. hemolysis, defined as anemia not due to other causes and either of the following:
        1. schistocytes or other red blood cell fragments seen on blood smear
        2. increased reticulocyte count

  Normotensive scleroderma renal crisis:

  1. Increase in serum creatinine >50% over Baseline OR serum creatinine ≥120% of ULN for local laboratory:

  AND

  2. One of the following 5 features:
     a. proteinuria ≥2+ by dipstick
     b. hematuria ≥2+ by dipstick or ≥10 red blood cells/high-powered field
     c. thrombocytopenia: <100,000 platelets/mm^3
     d. hemolysis, defined as anemia not due to other causes and either of the following:
        1. schistocytes or other red blood cell fragments seen on blood smear

2. increased reticulocyte count

e. Renal biopsy findings consistent with scleroderma renal crisis (microangiopathy)

- Decline in FVC % predicted ≥15% (relative), confirmed by another FVC % within a month, high resolution computed tomography to confirm ILD (if previous scan did not show ILD) and FVC % predicted <80% predicted;
- New onset of left ventricular failure (defined as ejection fraction ≤45%) requiring treatment;
- New onset of pulmonary hypertension (PAH) on right heart catheterization requiring treatment.
- Gastrointestinal dysmotility requiring enteral (tube feeding) or parenteral nutrition
- Digital ischemia with gangrene, amputation, or hospitalization requiring treatment

If a subject meets any of these criteria, the subject is assigned a probability of 0. Otherwise, in Step 2, the probability of improvement is calculated based on the 5 core measures incorporated into the ACR-CRISS, including changes in mRSS, FVC % predicted, HAQ-DI, PTGA and CGA [Khanna et al., 2016].

The predicted probability of improvement for each subject will be computed using the following equation in Step 2 (equation to derive predicted probabilities from a logistic regression model):

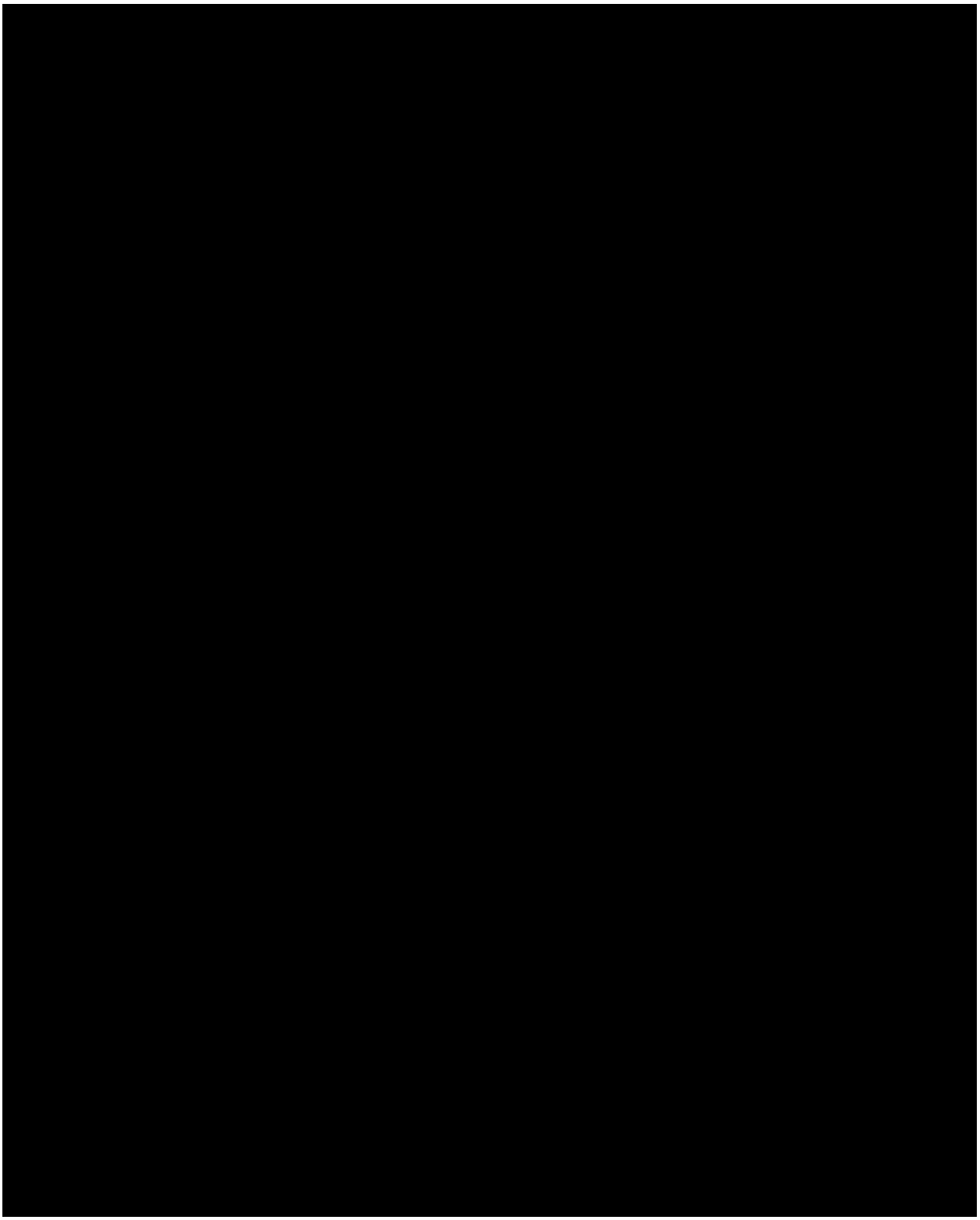
$$\frac{\exp(-5.54-0.81*\Delta MRSS+0.21*\Delta FVC\%-0.40*\Delta Pt-glob-0.44*\Delta MD-glob-3.41*\Delta HAQ-DI)}{1+\exp(-5.54-0.81*\Delta MRSS+0.21*\Delta FVC\%-0.40*\Delta Pt-glob-0.44*\Delta MD-glob-3.41*\Delta HAQ-DI)}$$

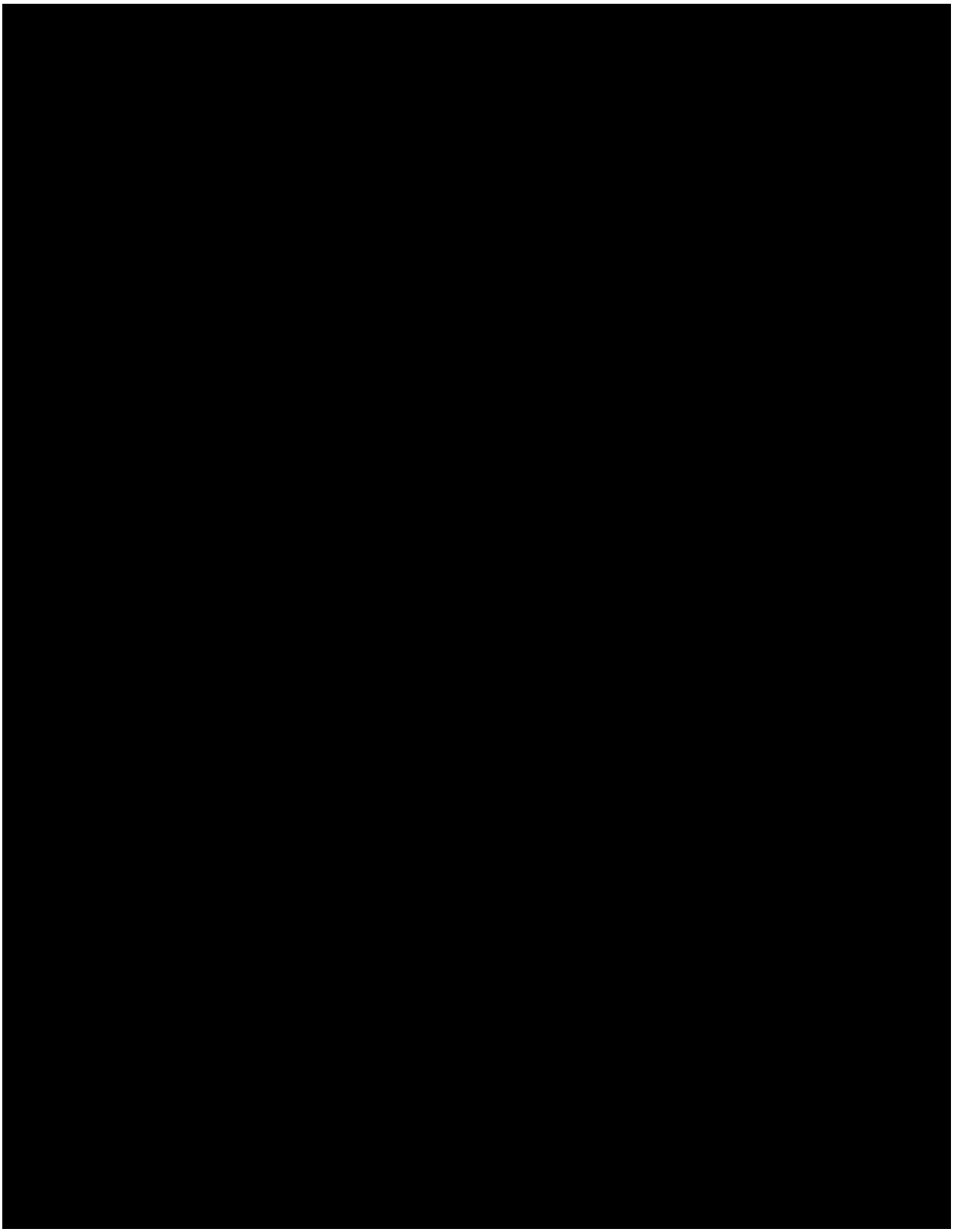
where Δ_{MRSS} indicates the change in mRSS from Baseline, $\Delta_{FVC\%}$ denotes the change in FVC% predicted from Baseline, $\Delta_{Pt\text{-}glob}$ indicates the change in PTGA, $\Delta_{MD\text{-}glob}$ denotes the change in CGA, and $\Delta_{HAQ\text{-}DI}$ is the change in HAQ-DI. All changes are absolute change ($Time_2 - Time_{baseline}$). The NRI-MI (Non-Responder Imputation in conjunction with Multiple Imputation) method will be used for subjects missing one or more components of the equation similar as Section 10.4.8. The CGA and PTGA are the 11 point scales rating the subject's overall health during the preceding week by physician and patient, respectively.

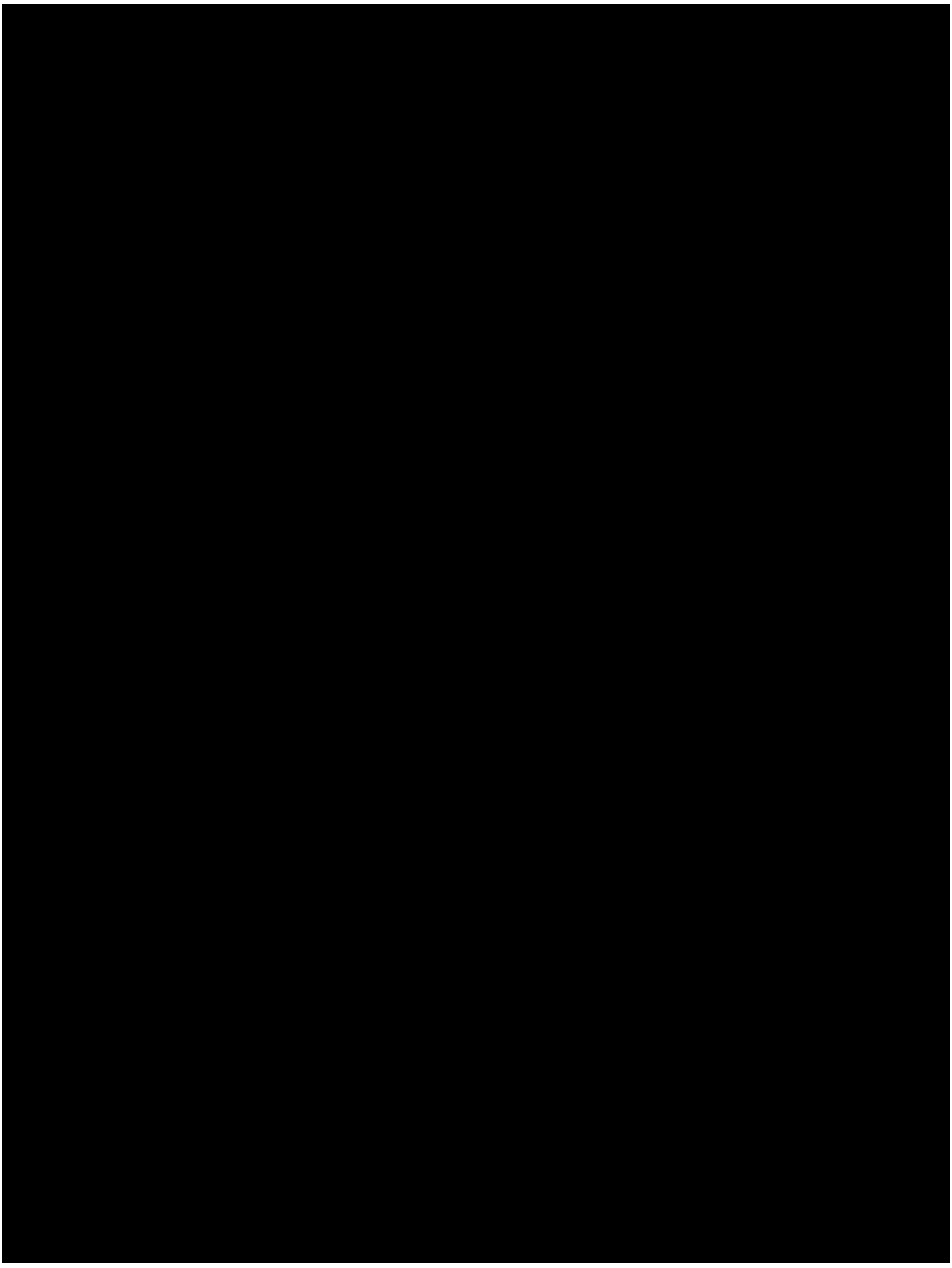
If a subject does not have an event in Step 1 of the calculation and the equation in Step 2 shows a probability of at least 0.6 for an individual subject at Week 52, that subject will be classified as a responder.

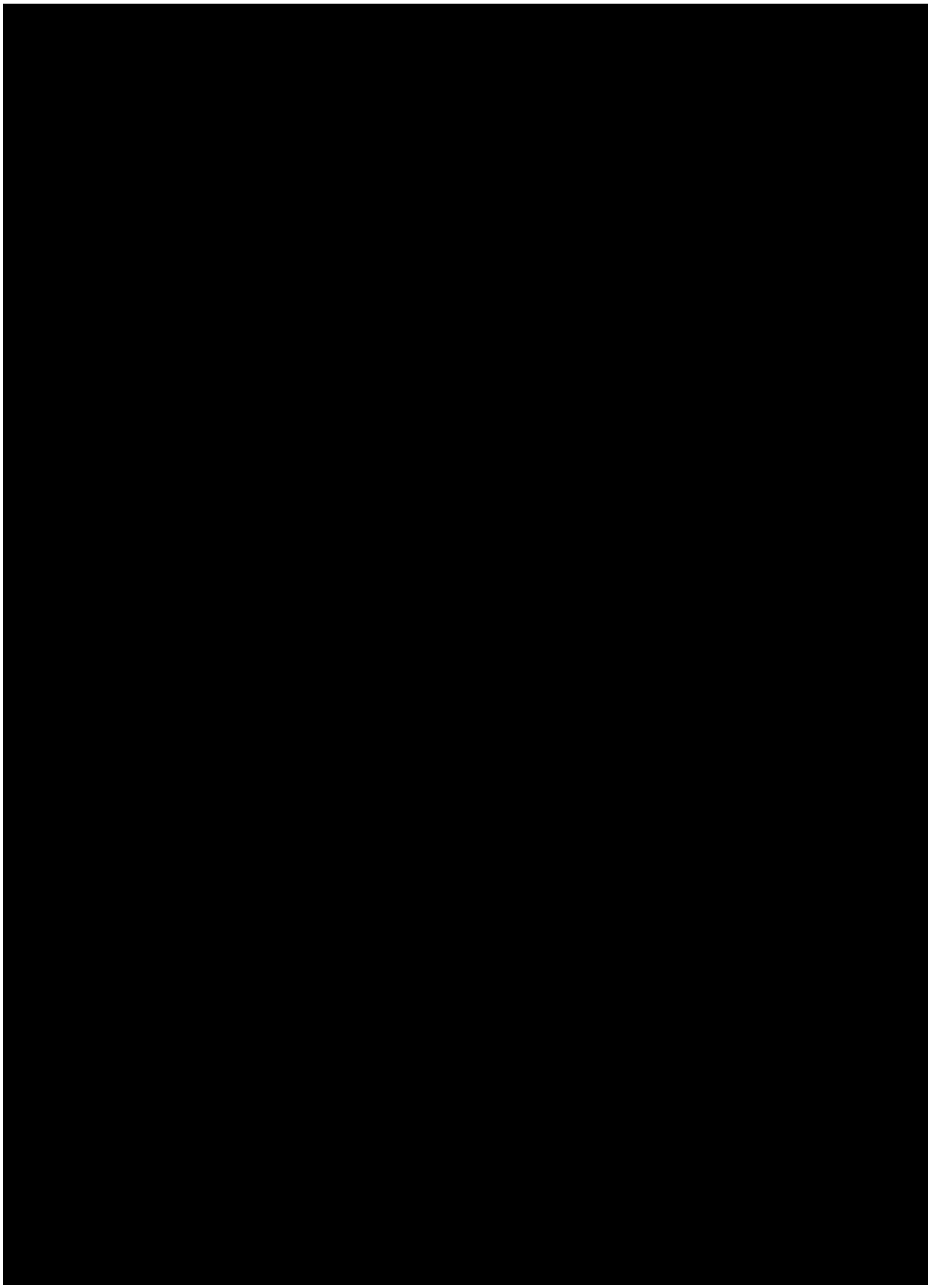
The responder rate by ACR-CRISS at Week 52 will be analyzed with a logistic regression model. The model will include treatment, use of mycophenolate mofetil [yes/no] and presence of ILD [yes/no]) at randomization. The risk differences between each dose and

placebo with its 95% confidence intervals will be presented. Inferences will be made if allowed under the algorithm described in Section 10.7.









[redacted]

10.6 Examination of Subject Subgroups

Subgroupings based on the baseline characteristics listed below may be explored for analysis of the primary endpoint and key secondary efficacy endpoint. Selected safety analyses may also be performed.

The presumed prognostic baseline characteristics include the following:

- Age (< 65 years and ≥ 65 years)
- Sex (male and female)
- Race (American Indian or Alaska Native, Asian, Black/African American, Native Hawaiian or Other Pacific Islander, White, or Other)

- Baseline FVC% predicted (≤70%, >70%)

If the actual size of the individual subgroup category is too small (≤10% of treated subjects), then the individual groups will be pooled in a meaningful way. If pooling cannot be done (e.g. if there are only two categories in the subgroup) then the subgroup analysis will not be performed.

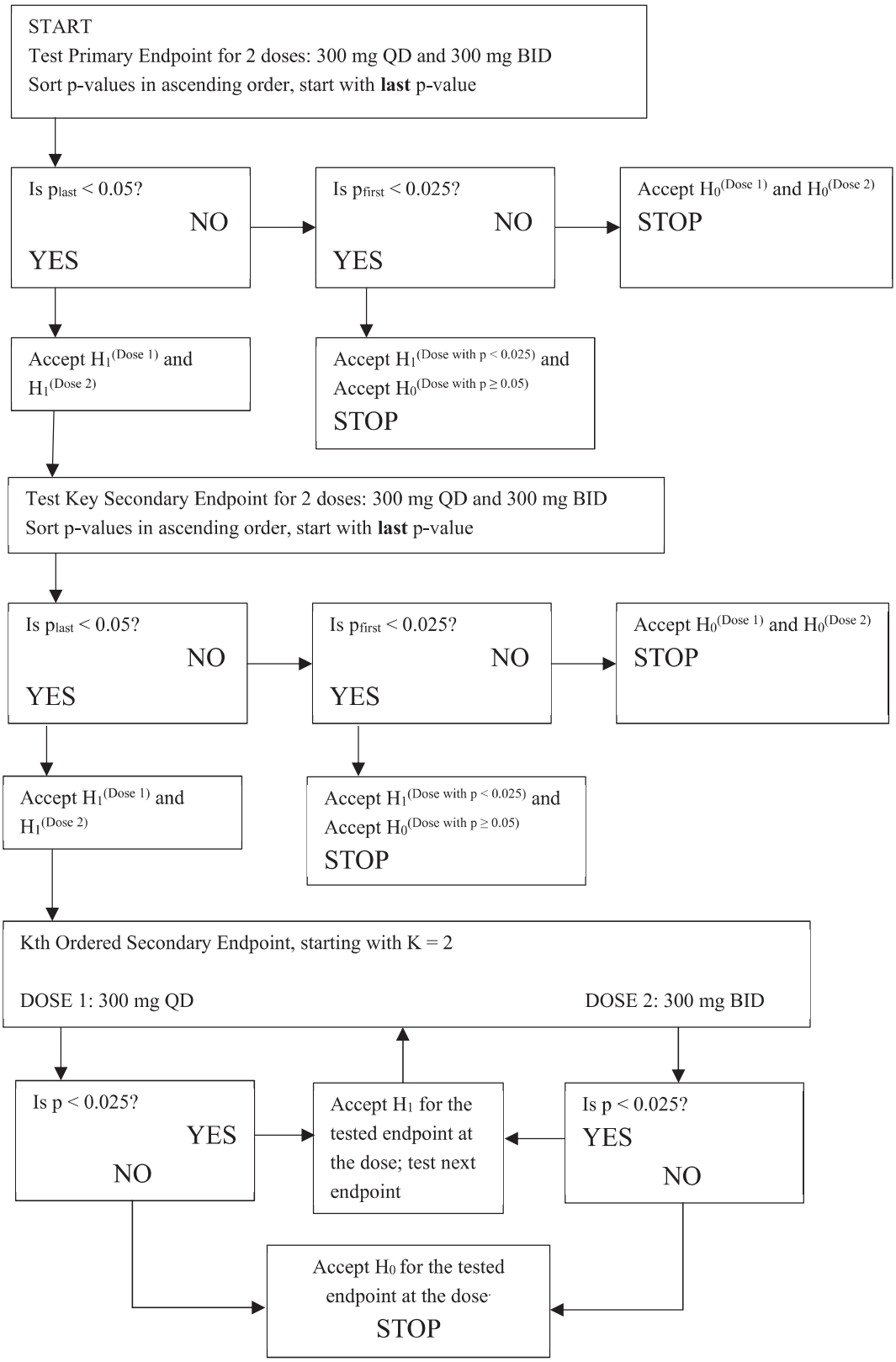
10.7 Adjustments for Multiplicity

The overall statistical level is α=0.05 (2-sided), except where otherwise specified.

Since 2 dose regimens of HZN-825 will be compared to placebo in the final analysis, adjustment for multiplicity will be used to preserve the family-wise error rate for multiple comparisons in the primary analysis. For the primary endpoint change from Baseline in FVC % predicted at Week 52, a Hochberg testing procedure will be used for the comparisons of HZN-825 BID vs. placebo and HZN-825 QD vs placebo. The resulting p-values will be ranked, and the larger p-value will be evaluated at the α = 0.05 (2-sided) threshold. If statistically significant, then both comparisons will be considered significant. If the larger of the two P-values is not statistically significant at α = 0.05 (2-sided), then the smaller P-value will be compared to an α = 0.025 (2-sided).

If one or both dose regimens of the primary endpoint are not rejected the null hypothesis (i.e., not difference between HZN-825 and placebo) after completing testing, no further testing will occur due to exhaustion of all alpha. If both HZN-825 doses are statistically significantly better than the placebo for the primary efficacy endpoint at α = 0.05 (2-sided), the same Hochberg testing procedure will subsequently be used to evaluate the key secondary endpoint. If both dose regimens of HZN-825 for the key secondary endpoint are considered significant, subsequent secondary endpoints will be tested sequentially within each dose in the order shown in Section 10.4, using α = 0.025 (2-sided) in a sequential testing procedure. If all preceding sequentially tested hypotheses are rejected at α = 0.025 (2-sided) the next hypothesis will be tested. Once a null hypothesis is not rejected, all subsequent endpoints within the same dose will not be tested due to alpha exhaustion. A graphic representation of the proposed multiplicity testing approach is presented in Figure 2 Schematic of Hochberg Testing Procedure

Figure 2 Schematic of Hochberg Testing Procedure



Although P-values will be provided for [redacted], they will not be used for inferential purposes.

11 Safety Analyses

Safety and tolerability of assessments of HZN-825 will be based on AEs, the AESIs ([redacted]), concomitant medication use, vital signs, 12-lead ECG and clinical safety laboratory evaluations (hematology, chemistry, lipids and coagulation panel, and urinalysis).

All safety analyses will be based on the Safety Analysis Set.

11.1 Extent of Exposure

Exposure to HZN-825 will be descriptively summarized by treatment group: drug exposure, measured as duration of treatment (days treated), is the number of days on treatment based on the first and last days of treatment with the study medication (last day of study medication – first day of study medication + 1). This will be based on the study drug dispensation CRF page.

Summary statistics will be provided for the duration of exposure to study medication (days) by treatment group. Additionally, a categorical breakdown of duration of treatment will be provided, including the following categories: 0 – 3 months, > 3 to 6 months, > 6 to 9 months, and > 9 to 12 months.

Summary statistics will also be presented for the total number of HZN-825 doses and total dose administered by treatment group.

11.2 Treatment Compliance

The calculation of overall compliance is based on all doses of HZN-825.

The formula for compliance (%) is calculated as: (cumulative actual dose / prescribed dose for time spent on study treatment)*100.

The prescribed dose for time spent on study treatment is calculated as: (date of last administration – date of first administration + 1) * daily prescribed dose.

The following summaries will be provided by treatment group:

- Summary statistics for percentage overall compliance.
- Percentage overall compliance categorized by frequencies < 70%, 70%-80%, 80%-90%, and > 90%.

The following listings will be provided:

- Randomized and actual treatments.
- Overall compliance.

11.3 Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as 1 or any of the following:

- Any AEs with an onset date on or after the study drug start date and time (if known) and no later than 28 days after permanent discontinuation of study drug or entry into the extension study (Week 52 visit).
- Any AEs with an onset date on or after the study drug start date and time (if known) and being indicated as 'Related to Study Drug' by the principal investigator which are not attributed to the extension study.

Refer to the table, listing, and figure shell document for determination of TEAE status in adverse events that are missing critical AE date start and stop information.

Imputed dates are only used for classification of TEAEs.

Prior AEs are defined as any AE with a start date prior to the date of first dose of study treatment. Prior AEs will not be summarized but will be included in listings.

Follow-up AEs are defined as any AE occurring > 28 days after permanent discontinuation of study drug for subjects who prematurely discontinue the study medication prior to 52 weeks (i.e. 392 days after first trial drug intake) or who do not enter the extension study. Follow-up AEs will be identified in the listings as such. Related SAEs will be followed up until resolved or until the event stabilizes and the overall clinical outcome has been ascertained.

The Verbatim terms in the eCRFs will be mapped to preferred terms (PTs) and system organ classes (SOCs) using the current version of MedDRA. Adverse events will be graded for severity using the Rheumatology Common Toxicity Criteria (RCTC) v2.0 [Woodworth et al., 2007].

For summaries by SOC, PT, and maximum severity, a subject will only be counted once for each SOC based on the maximum severity level reported for that SOC and once for each unique PT within that SOC level at the maximum intensity level reported for that PT.

For summaries by SOC and PT only, a subject will be counted at most once at the SOC level and at most once at each unique PT within the SOC level. Summaries presenting the frequency of TEAEs by SOC and PT will be ordered alphabetically by SOC and then, within a SOC, alphabetically by PT.

Event Adjusted Incidence Rate (EAIR, per 100 subject-years) is defined as number of subjects with events / total subject-years *100. Subject-years for each subject will be calculated as (date of first event occurred – first dose date +1)/365.25. For subjects without an event, subject-years for each subject = (last day in study – first dose date +1)/365.25. EAIR will be summarized by SOC and PT as needed for TEAE.

Events for which the investigator did not record relationship to study drug will be considered related to study drug for summary purposes. However, by-subject data listings will show the relationship as missing. No imputation will be done in case of missing study treatment relationship for non-treatment emergent AEs.

The severity grade of events for which the investigator did not record severity will be categorized as “missing” for tabular summaries and data listings. The missing category will be listed last in summary presentation.

Treatment-emergent AEs will be summarized based on the Safety Analysis Set.

Summaries of the following types will be presented:

The number and percentage of subjects who experienced at least 1 TEAE will be provided and summarized by SOC, PT, and treatment group.

The following AEs summaries will be provided by SOC, PT, and treatment group. AE summaries followed by an asterisk(*) will include the number of events and the corresponding incidence of events per PYE.:

- All TEAEs by maximum severity
- All treatment-emergent SAEs*
- All treatment-related TEAEs*
- All treatment-related TEAEs by maximum severity
- All treatment-related treatment-emergent SAEs
- All TEAEs meeting CRISS criteria
- All TEAEs leading to permanent withdrawal of any trial drug
- All TEAEs leading to death (i.e., outcome of death)
- All follow-up adverse events
- All serious follow-up adverse events
- Subject incidence of AESIs: [redacted]

A brief, high-level summary of AEs described above will be provided by treatment group for the main study treatment period and by the number and percentage of subjects who experienced the above AEs. All deaths observed in the trial will also be included in this summary. For the overall summary, the category of "TEAEs by maximum severity" described above will be replaced by AEs with severity > Grade 3.

A similar high-level overall summary of AEs may be created for follow-up AEs, and may include the following categories: All AEs, SAEs, AEs with severity > Grade 3, AEs meeting CRISS criteria, AEs leading to death, and AESIs.

The frequency of TEAEs occurring in > 5% of subjects in any treatment group will be summarized by treatment, primary SOC, and preferred term. A similar summary will be created for serious TEAEs occurring in > 3% of subjects in any treatment group by primary SOC and preferred term.

An adverse event plot will be created with frequency and risk differences (with corresponding exact 95% confidence interval [SAS® Documentation (2018]) plotted together for the adverse events occurring in >5% of subjects in any treatment group. A similar plot will be created for SAEs occurring in > 3% of subjects in any treatment group.

In addition, data listings will be provided for the following:

- All AEs, indicating whether the event is treatment-emergent
- SAEs
- Deaths
- AEs leading to permanent withdrawal of trial drug
- Adverse events of special interest: [redacted]

11.3.1 Adverse Events of Special Interest

The following AESI is identified for this trial:

[redacted]

11.4 Clinical Laboratory Evaluations

11.4.1 Summaries of Numeric Laboratory Results

Laboratory data collected during the trial will be analyzed and summarized using both quantitative and qualitative methods. Summaries of laboratory data will be provided for the Safety Analysis Set. Laboratory parameters to be summarized are listed in Table 5.

Table 5 Lab Parameters

Chemistry	Hematology	Lipids and Coagulation	Urinalysis
Total protein, albumin, sodium, phosphate, potassium, calcium, chloride, bicarbonate, blood urea nitrogen, creatinine, creatine kinase, uric acid, glucose, lactate dehydrogenase; liver function tests (alanine aminotransferase, aspartate aminotransferase, gamma glutamyl transferase, alkaline phosphatase, total bile acid, total bilirubin, conjugated and unconjugated bilirubin, if applicable;	Hemoglobin, hematocrit, red blood cells, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, red blood cell distribution width, reticulocyte count White blood cells count and differential (neutrophils, eosinophils, basophils, monocytes, lymphocytes), platelets, high sensitivity C-reactive Protein (hsCRP) and erythrocyte sedimentation rate (ESR)	Lipids: total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol and triglycerides Coagulation (if applicable): prothrombin time, partial thromboplastin time, international normalized ratio and fibrinogen	Ketone, Specific Gravity, Urine Glucose, Urine Protein, Urobilinogen, and pH Urinalysis also reports blood

Laboratory results collected in conventional units will be converted to International System of Units (SI) for all summaries and listings. Clinical laboratory test results (Hematology, Chemistry, Lipids and Coagulation, and Urinalysis) and their changes from baseline will be summarized by visit and treatment group using descriptive statistics.

For hematology, chemistry, and lipids and coagulation, results will be categorized as low, normal, or high based on their normal ranges. If toxicity grading is applicable to the laboratory result, toxicity grades will also be evaluated and assigned to the result. If available, shift tables using categories of toxicity grade comparing laboratory test results from baseline to each visit will be presented with percentages based on subjects with a non-missing value at

baseline and post-baseline visit. If toxicity grade is not available, then categories of low, normal, and high will be used for the analogous presentation.

For urinalysis tests, results will be classified as normal or abnormal. Results out of range will be identified as such on subject listings. Shift tables for urinalysis results using categories of normal and abnormal, comparing laboratory test results from Baseline to each visit will be presented with percentages based on subjects with a non-missing value at Baseline and post-baseline visit.

A listing of urine pregnancy test results will be produced.

In addition, a by-subject listing for laboratory test results will be provided by subject ID number, and time point in chronological order for hematology, serum chemistry, lipids and coagulation, and urinalysis separately. Values falling outside of the relevant reference range and/or having a severity grade of 1 or higher based on the toxicity severity grade will be flagged in the data listings, as appropriate.

No formal statistical testing is planned.

Shift plots may be produced for selected laboratory values.

11.4.2 Summaries of Liver Function Tests

Liver-related abnormalities after initial study drug dosing will be examined and summarized by treatment group using the number and percentage of subjects who were reported to have the following laboratory test values for postbaseline measurements:

- Aspartate aminotransferase (AST): (a) > 3 times of the upper limit of reference range (ULN); (b) > 5 x ULN
- Alanine aminotransferase (ALT): (a) > 3 x ULN; (b) > 5 x ULN
- AST or ALT: (a) > 3 x ULN; (b) > 5 x ULN
- Total bilirubin: > 2 x ULN
- Alkaline phosphatase (ALP) > 1.5 x ULN
- Potential Hy's Law: AST or ALT > 3 x ULN and total bilirubin > 2 x ULN

The summary will include data from all postbaseline visits up to 28 days after the last dose of study drug.

For individual laboratory tests, subjects will be counted once based on the most severe postbaseline values. For both the composite endpoint of AST or ALT and total bilirubin,

subjects will be counted once when the criteria are met at the same postbaseline visit date. The denominator is the number of subjects in the Safety Analysis Set who have non-missing postbaseline values of all relevant tests at the same postbaseline visit date.

A listing of subjects who met at least 1 of the above criteria will be provided.

Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plots will be produced.

11.5 Vital Signs and Weight

Vital signs (blood pressure, heart rate, respiratory rate, temperature) will be measured at all clinic visits.

Weight will be measured at Screening, Day 1, Week 28, and Week 52. Height will be measured at Screening.

Descriptive summaries of observed and change from baseline values will be presented for each vital sign parameter and weight by treatment group and visit. A shift table for vital signs by CTCAE grade and visit will be summarized.

In the case of multiple values in an analysis window, data will be selected for analysis as described in the TLF shells document. No formal statistical testing is planned.

A by-subject listing of vital signs and weight will be provided by subject ID number and time point in chronological order. High or low values will be flagged.

11.6 ECGs

12-lead ECGs will be performed at Screening (Baseline) and Weeks 1, 4, 16, 28, and 52. The results will be recorded as normal or abnormal on the eCRF and all abnormal results will be evaluated as clinically significant (CS) or not clinically significant (NCS) by the investigator.

Descriptive summaries of observed and change from baseline values will be presented for each ECG parameter by treatment group and visit, including HR, PR, QRS, QT, and QTcF. ECG shift tables will be presented providing the count of subjects with each type of finding (normal, abnormal – NCS, or abnormal – CS) at baseline compared to each post-baseline visit by treatment group with percentages based on subjects in the safety analysis set with a non-missing value at the baseline and post-baseline visit.

Further, a summary will be provided of the count and percent of subjects with any post-Baseline assessment in the following categories:

The abnormal post-dose QTcF interval values obtained during the trial will be summarized within the following categories:

- • > 450 msec
- • > 480 msec
- • > 500 msec
- • QTcF increase from Baseline > 30 msec
- • QTcF increase from Baseline > 60 msec

Percentages will be based on the number of subjects in the safety analysis set with at least one post-baseline value. Similar summaries will be provided by visit with the denominator based on the number of subjects with data at the given visit.

11.7 Pharmacokinetics

Blood samples will be collected from all subjects to evaluate the PK of HZN-825 and metabolite(s) at the following visits: Day 1 (at 2 to 4 hours after the first dose of trial drug), Week 4 (pre-dose), Week 10 (anytime at the visit), Weeks 16 and 28 (pre-dose and 2 to 4 hours post-dose) and Weeks 40 and 52 (pre-dose). Note that all pre-dose samples will be collected prior to any trial drug administration for the day. For subjects not entering the 52-week extension trial, a sample will be collected anytime during the Week 52 Visit. PK sample collection time and the most recent dosing time prior to PK sample collection will be recorded for all PK samples.

The following presentations of subject plasma concentration data covered in this SAP will be provided for HZN-825 for the PK Analysis Set:

- o A listing including subject, week/time point (planned), treatment and plasma concentrations.
- o A table summary of plasma concentrations at each time point (n; arithmetic mean, geometric mean, SD, coefficient of variation (CV)% calculated as 100% × SD/mean, minimum, median and maximum)

Pre-dose PK samples will only be considered ‘pre-dose’ if they are collected within the 10-14 hours window post the most recent dose prior to the PK sample collection.

All PK related analysis results will be provided by Department of Clinical Pharmacology Modeling and Simulation (CPMS) at Amge. Population PK analysis and exposure-response analysis may be performed, with details reported separately.

[redacted]

12 References

Bruce B, Fries JF. The Stanford Health Assessment Questionnaire : dimensions and practical applications. Health Qual Life Outcomes. 2003 ;1 :20.

Cole JC, Khanna D, Clements PJ, et al. Single-factor scoring validation for the Health Assessment Questionnaire-Disability Index (HAQ-DI) in patients with systemic sclerosis and comparison with early rheumatoid arthritis patients. Qual Life Res. 2006;15(8):1383-94.

[redacted]

Khanna D, Berrocal VJ, Giannini EH, et al. The American College of Rheumatology Provisional Composite Response Index for Clinical Trials in Early Diffuse Cutaneous Systemic Sclerosis. Arthritis Rheumatol. 2016;68(2):299-311.

Khanna D, Denton CP, Jahreis A, et al. Safety and efficacy of subcutaneous tocilizumab in adults with systemic sclerosis (faSScinate): a phase 2, randomised, controlled trial. Lancet. 2016;387:2630-40.

Khanna D, Furst DE, Clements PJ, et al. Standardization of the modified Rodnan skin score for use in clinical trials of systemic sclerosis. J Scleroderma Relat Disord. 2017;2(1):11-8.

Khanna D, Lin CJF, Furst DE, et al. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Respir Med. 2020;8:963-74.

Khanna D, Spino C, Johnson S, et al. Abatacept in early diffuse cutaneous systemic sclerosis: results of a phase 2 investigator-initiated, multicenter, double-blind, randomized, placebo-controlled trial. Arthritis Rheumatol. 2020;72:125-36.

Lan KKG, Wittes, J. The B-Value: a tool for monitoring data. Biometrics. 1988;2:579-585.

Man A, Correa JK, Ziemek J, Simms RW, Felson DT, Lafyatis R. Development and validation of a patient-reported outcome instrument for skin involvement in patients with systemic sclerosis. Ann Rheum Dis. 2017;76:1374-80.

[redacted]

SAS® Documentation (2018). Base SAS® 9.4 Procedures Guide: Statistical Procedures, Sixth Edition.

[redacted]

Steen VD, Mayes MD, Merkel PA. Assessment of kidney involvement. Clin Exp Rheumatol. 2003;21(3Suppl29):S29-31.

[redacted]

Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34:220-33.

[redacted]

Woodworth T, Furst DE, Alten R, Bingham CO 3rd, Yocum D, Sloan V, Tsuji W, Stevens R, Fries J, Witter J, Johnson K, Lassere M, Brooks P. Standardizing assessment and reporting of

adverse effects in rheumatology clinical trials II: the Rheumatology Common Toxicity Criteria v.2.0. J Rheumatol. 2007 Jun;34(6):1401-14.

13 Appendix 1 Schedule of Assessments (Clinical Study Protocol V4.0)

		Double-blind Treatment Period [2]										Safety Follow-up Visit
Trial Visit	**SCR [1]**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11 4 weeks after last dose of trial drug**
Trial Week (W)	**-28 days**	**Day 1 [3]**	**W4**	**W10**	**W16**	**W22**	**W28**	**W34**	**W40**	**W46**	**W52/PD [4]**	
Visit Window (±days)			**(±3)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±14)**
Informed consent	X											
Review eligibility criteria	X	X										
Demographics	X											
Medical history [5]	X	X										
Weight	X	X					X				X	
Height	X											
Randomization [6]		X										
Trial drug dispensing		X	X	X	X	X	X	X	X	X	X [7]	
Compliance			X	X	X	X	X	X	X	X	X	
mRSS [8]	X	X			X		X		X		X	
FVC % predicted/spirometry	X	X			X		X		X		X	
MDGA		X			X		X		X		X	
[redacted]												
Patient-reported outcome assessments												
[redacted]												
PTGA		X			X		X		X		X	
[redacted]												
SSPRO-18		X			X		X		X		X	

		Double-blind Treatment Period [2]										Safety Follow-up Visit
Trial Visit	**SCR [1]**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11 4 weeks after last dose of trial drug**
Trial Week (W)	**-28 days**	**Day 1 [3]**	**W4**	**W10**	**W16**	**W22**	**W28**	**W34**	**W40**	**W46**	**W52/PD [4]**	
Visit Window (±days)			**(±3)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±14)**
[redacted]												
Anchor questions												
ACR-CRISS (last week)		X			X		X		X		X	
ACR-CRISS (overall health and change since start of trial)							X				X	
FVC (last week)		X			X		X		X		X	
FVC (change since start of trial)							X				X	
HAQ-DI (last week)		X			X		X		X		X	
HAQ-DI (change since start of trial)							X				X	
SSPRO-18 (last week)		X			X		X		X		X	
SSPRO-18 (change since start of trial)							X				X	

		Double-blind Treatment Period [2]										Safety Follow-up Visit
Trial Visit	**SCR [1]**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11 4 weeks after last dose of trial drug**
Trial Week (W)	**-28 days**	**Day 1 [3]**	**W4**	**W10**	**W16**	**W22**	**W28**	**W34**	**W40**	**W46**	**W52/PD [4]**	
Visit Window (±days)			**(±3)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±14)**
Subject interviews [14]											X	
Pregnancy test [15]	X	X	X		X		X		X		X	X
Physical examination [16]	X	X					X				X	X
Vital signs [17]	X	X	X	X	X	X	X	X	X	X	X	X
[redacted]												
12-lead electrocardiogram [18]	X	X	X		X		X				X	
Echocardiogram [18]	X											
Clinical laboratory safety tests												
Chemistry [19]	X	X	X	X	X	X	X	X	X	X	X	X
Lipids [20]	X				X						X	
Hematology [21]	X	X	X	X	X	X	X	X	X	X	X	X
Coagulation [22]	X			X								
Urinalysis	X	X	X	X	X	X	X	X	X	X	X	X
HBV and HCV serology	X											
[redacted]												
PK samples [25]		X	X	X	X		X		X		X	
[redacted]												

		Double-blind Treatment Period [2]										Safety Follow-up Visit
Trial Visit	**SCR [1]**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11 4 weeks after last dose of trial drug**
Trial Week (W)	**-28 days**	**Day 1 [3]**	**W4**	**W10**	**W16**	**W22**	**W28**	**W34**	**W40**	**W46**	**W52/PD [4]**	
Visit Window (±days)			**(±3)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±5)**	**(±14)**
Adverse event assessment [28]	X	X	X	X	X	X	X	X	X	X	X	X
Prior/concomitant medications [29]	X	X	X	X	X	X	X	X	X	X	X	X

ACR-CRISS=American College of Rheumatology-Composite Response Index in Systemic Sclerosis; BID=twice daily; [redacted]; FVC=forced vital capacity; HAQ-DI=Health Assessment Questionnaire – Disability Index; HBV=hepatitis B virus; HCV=hepatitis C virus; [redacted]; hsCRP=high-sensitivity C-reactive protein; MDGA=Physician Global Assessment; mRSS=modified Rodnan skin score; PD=premature discontinuation; PK=pharmacokinetic; PTGA=Patient Global Assessment; QD=once daily; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; SCR=Screening [redacted]; SSc=systemic sclerosis [redacted] SSPRO=scleroderma skin patient-reported outcome; VAS=visual analog scale; W=Week; WOCBP=women of childbearing potential

1. Screening procedures can take place over more than 1 day/clinic visit provided consent is obtained first and all assessments are completed within the designated window. An abnormal test during Screening may be repeated once during the Screening Period.
2. Subjects will take HZN-825 300 mg QD, HZN-825 300 mg BID or placebo depending upon randomization.
3. On Day 1 (Baseline), subjects will be randomized and receive the first dose of trial drug in the clinic. All Day 1 assessments should be performed before the first dose of trial drug is administered in the clinic except for the PK sample collected 2-4 hours post dose.
4. If a subject prematurely discontinues trial drug, he/she will be asked to remain in the trial, participating in the scheduled trial visits through Week 52. If a subject prematurely discontinues trial drug and does not wish to continue in the trial, he/she will be asked to return for a clinic visit and undergo the Week 52 assessments. Subjects not entering the 52-week extension trial will return to the clinic 4 weeks after the last dose of trial drug for a Safety Follow-up Visit.
5. Medical history, including SSc history and treatment, as well as substance use history.
6. Subjects will be randomized in a 1:1:1 ratio to receive HZN-825 300 mg QD, HZN-825 300 mg BID or placebo.
7. For subjects who are entering the extension trial.
8. The assessment should be performed by the Investigator (or designee) who is trained in skin scoring. Except when strictly unavoidable, the same person should perform the assessment at each evaluation during the trial.

[redacted]

[redacted]

14. Will be conducted for 20 subjects in the United States only.
15. Perform for WOCBP. Serum pregnancy test at Screening and Week 52 (or as needed). Urine pregnancy tests should also be done every 4 weeks after randomization, which includes both in-clinic testing at scheduled visits prior to dosing (Weeks 1, 4, 16, 28 and 40) and at home (also a ±5-day window) by the subject and reported to the site (Weeks 8, 12, 20, 24, 32, 36, 44 and 48).
16. A complete physical examination, including but not limited to cardiac, pulmonary, neurologic and skin assessments, as well as directed rheumatology assessments.
17. Vital signs (blood pressure, heart rate, respiratory rate, temperature) will be measured at each clinic visit.
18. Additional electrocardiograms or echocardiograms will be conducted, if clinically indicated. An echocardiogram that has been performed within the 3 months prior to Screening can serve as the Baseline echocardiogram if the subject has been clinically stable.
19. Includes total protein, albumin, sodium, phosphate, potassium, calcium, chloride, bicarbonate, blood urea nitrogen, creatinine, creatine kinase, uric acid, lactate dehydrogenase, hsCRP and liver function tests (alanine aminotransferase, aspartate aminotransferase, gamma glutamyltransferase, alkaline phosphatase, total bile acid, total bilirubin and conjugated and unconjugated bilirubin, if applicable).
20. Includes fasting glucose, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol and triglycerides.
21. Includes hemoglobin, hematocrit, red blood cell count (with morphology if blood cell count is abnormal), white blood cell counts with differential (neutrophils, eosinophils, basophils, monocytes, lymphocytes), platelet count and ESR (must be processed within one hour of blood draw).
22. To determine risk of bleeding for [redacted]; includes prothrombin time, partial thromboplastin time, international normalized ratio and fibrinogen.

[redacted]

25. PK samples will be collected at each of the following visits: Day 1 (at 2 to 4 hours after the first dose of trial drug), Week 4 (pre-dose), Week 10 (anytime during the visit), Weeks 16 and 28 (pre-dose and 2 to 4 hours post-dose) and Weeks 40 and 52 (pre-dose). Note: all pre-dose samples will be collected prior to any trial drug administration for the day. For subjects not entering the 52-week extension, a sample will be collected anytime during the Week 52 Visit.

[redacted]

28. Adverse events that occur after signing the informed consent form and prior to dosing on Day 1 will be considered medical history. Adverse events occurring or worsening after the first dose of trial drug through the Safety Follow-up Visit will be considered treatment-emergent adverse events. All adverse events that occur from the signing of informed consent through the Safety Follow-up Visit will be recorded. The subject should be assessed for the development of new onset of scleroderma renal crisis, new onset or worsening of lung fibrosis, new onset of pulmonary arterial hypertension requiring treatment or right heart catheterization or new onset of left ventricular failure.
29. Includes recording of herb/supplement use. See Table 9.1 for restrictions regarding medications.

14 Appendix 2 Approvals

Confirmation by the study biostatistician (or designee), biostatistics management (or designee), and the study clinical colleague or therapeutic lead (or designee) that the review of this statistical analysis plan is complete, and there is agreement on the content.

[redacted]
Director , Biostatistics

Name,
Title

Signature/Date

Signed by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 11/7/2024 | 8:03:24 PM GMT
[redacted]

[redacted]
Medical Director

Name,
Title

Signature/Date

Signed by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 11/8/2024 | 5:50:02 PM GMT
[redacted]

DocuSign

Certificate Of Completion

[redacted] Status: Completed

Subject: Complete with Docusign: SAP HZN 825_301_Final_V1Nov2024.pdf

Ensure Legal Names obtained for all recipients: Legal Names were confirmed

Required fields in document: Signature: I read SOP-432501, and understand Signature (and/or Initial) field is required

Source Envelope:

Document Pages: 65 | Signatures: 2 | Envelope Originator:

Certificate Pages: 5 | Initials: 0 [redacted]

AutoNav: Enabled

EnvelopeId Stamping: Enabled

Time Zone: (UTC) Monrovia, Reykjavik

IP Address: [redacted]

Record Tracking

Status: Original
11/7/2024 8:00:17 PM

Holder: [redacted]

Location: DocuSign

Signer Events	Signature	Timestamp
[redacted] Security Level: Email, Account Authentication (Required)	[redacted] Signature Adoption: Pre-selected Style Signature ID: [redacted] Using IP Address: [redacted] With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 11/7/2024 8:02:07 PM Viewed: 11/8/2024 5:49:14 PM Signed: 11/8/2024 5:50:19 PM
Electronic Record and Signature Disclosure: Accepted: 11/8/2024 5:49:14 PM ID: e93f1c67-1fa8-4bfc-aeb9-be26b19201a4		
[redacted] Security Level: Email, Account Authentication (Required), Login with SSO	[redacted] Signature Adoption: Pre-selected Style Signature ID: [redacted] Using IP Address: [redacted] With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 11/7/2024 8:02:07 PM Viewed: 11/7/2024 8:02:42 PM Signed: 11/7/2024 8:03:31 PM
Electronic Record and Signature Disclosure: Accepted: 11/7/2024 8:02:42 PM ID: [redacted]		

In Person Signer Events	Signature	Timestamp

Editor Delivery Events	Status	Timestamp

Agent Delivery Events	Status	Timestamp

Intermediary Delivery Events	Status	Timestamp

Certified Delivery Events	Status	Timestamp

Carbon Copy Events	Status	Timestamp

Witness Events	Signature	Timestamp

Notary Events	Signature	Timestamp

Envelope Summary Events	Status	Timestamps
Envelope Sent	Hashed/Encrypted	11/7/2024 8:02:07 PM
Certified Delivered	Security Checked	11/7/2024 8:02:42 PM
Signing Complete	Security Checked	11/7/2024 8:03:31 PM
Completed	Security Checked	11/8/2024 5:50:19 PM

Payment Events	Status	Timestamps

Electronic Record and Signature Disclosure

CONSENT TO USE OF ELECTRONIC SIGNATURE

CONSUMER DISCLOSURE

From time to time, Amgen Inc. (we, us or Company) may be required by law to provide to you certain written notices or disclosures. Described below are the terms and conditions for providing to you such notices and disclosures electronically through the DocuSign system. Please read the information below carefully and thoroughly, and if you can access this information electronically to your satisfaction and agree to this Electronic Record and Signature Disclosure (ERSD), please confirm your agreement by selecting the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

Getting paper copies

At any time, you may request from us a paper copy of any record provided or made available electronically to you by us. You will have the ability to download and print documents we send to you through the DocuSign system during and immediately after signing session and, if you elect to create a DocuSign signer account, you may access them for a limited period of time (usually 30 days) after such documents are first sent to you. After such time, if you wish for us to send you paper copies of any such documents from our office to you, you will be charged a $0.00 per-page fee. You may request delivery of such paper copies from us by following the procedure described below.

Compliance

You agree that any electronic signature used to sign this document shall be treated the same as handwritten signature for the purposes of validity, enforceability and admissibility.

Authorized Representative

Any agreement for signature accessed through a secure link to DocuSign is to be signed by a representative of your organization authorized to bind your organization ("authorized representative"). If you receive an agreement for signature and are not the authorized representative, please reassign the request to the authorized representative by selecting 'Other Actions' in the upper right corner and clicking 'Assign to Someone Else'. You will receive a copy of the fully executed agreement via email upon completion.

Withdrawing your consent

If you decide to receive notices and disclosures from us electronically, you may at any time change your mind and tell us that thereafter you want to receive required notices and disclosures only in paper format. How you must inform us of your decision to receive future notices and disclosure in paper format and withdraw your consent to receive notices and disclosures electronically is described below.

Consequences of changing your mind

If you elect to receive required notices and disclosures only in paper format, it will slow the speed at which we can complete certain steps in transactions with you and delivering services to you because we will need first to send the required notices or disclosures to you in paper format, and then wait until we receive back from you your acknowledgment of your receipt of such paper notices or disclosures. Further, you will no longer be able to use the DocuSign system to receive required notices and consents electronically from us or to sign electronically documents from us.

All notices and disclosures will be sent to you electronically

Unless you tell us otherwise in accordance with the procedures described herein, we will provide electronically to you through the DocuSign system all required notices, disclosures, authorizations, acknowledgements, and other documents that are required to be provided or made available to you during the course of our relationship with you. To reduce the chance of you inadvertently not receiving any notice or disclosure, we prefer to provide all of the required notices and disclosures to you by the same method and to the same address that you have given us. Thus, you can receive all the disclosures and notices electronically or in paper format through the paper mail delivery system. If you do not agree with this process, please let us know as described below. Please also see the paragraph immediately above that describes the consequences of your electing not to receive delivery of the notices and disclosures electronically from us.

How to contact Amgen Inc.

You may contact us to let us know of your changes as to how we may contact you electronically, to request paper copies of certain information from us, and to withdraw your prior consent to receive notices and disclosures electronically as follows:

To contact us by email send messages to: docusign@amgen.com

To advise Amgen Inc. of your new email address

To let us know of a change in your email address where we should send notices and disclosures electronically to you, you must send an email message to us at docusign@amgen.com and in the body of such request you must state: your previous email address, your new email address. We do not require any other information from you to change your email address. If you created a DocuSign account, you may update it with your new email address through your account preferences.

To request paper copies from Amgen Inc.

To request delivery from us of paper copies of the notices and disclosures previously provided by us to you electronically, you must send us an email to docusign@amgen.com and in the body of such request you must state your email address, full name, mailing address, and telephone number. We will bill you for any fees at that time, if any.

To withdraw your consent with Amgen Inc.

To inform us that you no longer want to receive future notices and disclosures in electronic format you may:

i. decline to sign a document from within your DocuSign session, and on the subsequent page, select the check-box indicating you wish to withdraw your consent, or you may;

ii. send us an email to docusign@amgen.com and in the body of such request you must state your email, full name, mailing address, and telephone number. We do not need any other information from you to withdraw consent. The consequences of your withdrawing consent for online documents will be that transactions may take a longer time to process.

Required hardware and software

The minimum system requirements for using the DocuSign system may change over time. The current system requirements are found: here.

Acknowledging your access and consent to receive materials electronically

To confirm to us that you can access this information electronically, which will be similar to other electronic notices and disclosures that we will provide to you, please confirm that you have read this ERSD, and (i) that you are able to print on paper or electronically save this ERSD for your future reference and access; or (ii) that you are able to email this ERSD to an email address where you will be able to print on paper or save it for your future reference and access. Further, if you consent to receiving notices and disclosures exclusively in electronic format as described herein, then select the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

By selecting the check-box next to 'I agree to use electronic records and signatures', you confirm that:

- You can access and read this Electronic Record and Signature Disclosure; and and
- You can print on paper this Electronic Record and Signature Disclosure, or save or send this Electronic; and
- Record and Disclosure to a location where you can print it, for future reference and access; and until or unless you notify Amgen Inc. as described above, you consent to receive exclusively through electronic means all notices, disclosures, authorizations, acknowledgements, and other documents that are required to be provided or made available to you by Amgen Inc. during the course of your relationship with Amgen Inc. br>