STATISTICAL ANALYSIS PLAN

Efficacy and safety of SAR156597 in the treatment of diffuse cutaneous Systemic Sclerosis (dcSSc): A randomized, double blind, placebo controlled, 24 week, proof of concept study

SAR156597-ACT14604

STATISTICIAN: [redacted]

Statistical Project Leader: [redacted]

DATE OF ISSUE: 21-Feb-2019

NCT02921971

Total number of pages: 78

TABLE OF CONTENTS

STATISTICAL ANALYSIS PLAN **1**

TABLE OF CONTENTS **2**

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS **5**

1 OVERVIEW AND INVESTIGATIONAL PLAN **7**

1.1 STUDY DESIGN AND RANDOMIZATION 7

1.2 OBJECTIVES 7

1.2.1 Primary objectives 7

1.2.2 Secondary objectives 7

1.2.3 Exploratory objectives 8

1.3 DETERMINATION OF SAMPLE SIZE 8

1.4 STUDY PLAN 9

1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL 9

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN 9

2 STATISTICAL AND ANALYTICAL PROCEDURES **11**

2.1 ANALYSIS ENDPOINTS 11

2.1.1 Demographic and baseline characteristics 11

2.1.2 Prior or concomitant medications 12

2.1.3 Efficacy endpoints 12

2.1.3.1 Primary efficacy endpoint(s) 13

2.1.3.2 Secondary efficacy endpoints 13

2.1.4 Exploratory endpoints 14

2.1.5 Safety endpoints 17

2.1.5.1 Adverse events variables 17

2.1.5.2 Deaths 18

2.1.5.3 Laboratory safety variables 18

2.1.5.4 Vital signs variables 20

2.1.5.5 Electrocardiogram variables 20

2.1.6 Pharmacokinetic variables 20

2.1.7 Immunogenicity variables 20

2.1.8 Biomarker endpoints 22

2.1.9 Quality-of-life endpoints 22
2.1.10 Health economic endpoints 22
2.2 DISPOSITION OF PATIENTS 22
2.2.1 Randomization and drug dispensing irregularities 24
2.3 ANALYSIS POPULATIONS 25
2.3.1 Efficacy populations 25
2.3.1.1 Intent–to-treat population 25
2.3.2 Safety population 25
2.3.3 Anti-SAR156597 antibody (ADA) population 26
2.3.4 Pharmacokinetics population 26
2.3.5 Biomarker population 26
2.4 STATISTICAL METHODS 26
2.4.1 Demographics and baseline characteristics 26
2.4.2 Prior or concomitant medications 27
2.4.3 Extent of investigational medicinal product exposure and compliance 27
2.4.3.1 Extent of investigational medicinal product exposure 27
2.4.3.2 Compliance 28
2.4.4 Analyses of efficacy endpoints 29
2.4.4.1 Analysis of primary efficacy endpoint(s) 29
2.4.4.2 Analyses of secondary efficacy endpoints 33
2.4.4.3 Multiplicity issues 34
2.4.4.4 Analyses of exploratory efficacy endpoints 34
2.4.5 Analyses of safety data 37
2.4.5.1 Analyses of adverse events 38
2.4.5.2 Deaths 40
2.4.5.3 Analyses of laboratory variables 40
2.4.5.4 Analyses of vital sign variables 41
2.4.5.5 Analyses of electrocardiogram variables 42
2.4.6 Analyses of pharmacokinetic variables 42
2.4.7 Analyses of immunogenicity data 42
2.4.8 Analyses of Biomarker data 43
2.4.9 Analyses of quality of life/health economics variables 43
2.5 DATA HANDLING CONVENTIONS 44
2.5.1 General conventions 44
2.5.2 Data handling conventions for efficacy variables 44
2.5.3 Missing data 46
2.5.4 Windows for time points 47
2.5.5 Unscheduled visits 48

2.5.6 Pooling of centers for statistical analyses 48
2.5.7 Statistical technical issues 48

3 INTERIM ANALYSIS 49

4 DATABASE LOCK 50

5 SOFTWARE DOCUMENTATION 51

6 REFERENCES 52

7 LIST OF APPENDICES 53

APPENDIX A POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES CRITERIA 54

APPENDIX B SUMMARY OF STATISTICAL ANALYSES 59

APPENDIX C MRSS 64

APPENDIX D SHAQ 65

APPENDIX E UCLA SCTC GIT 2.0 QUESTIONNAIRE 68

APPENDIX F EQ-5D-5L QUESTIONNAIRE (UK VERSION) 72

APPENDIX G PATIENT AND PHYSICIAN GLOBAL ASSESSMENTS OF OVERALL HEALTH USING THE 0-10 LIKERT SCALE 74

APPENDIX H TJC28 75

APPENDIX I DIGITAL ULCER COUNT 76

APPENDIX J CODELISTS FOR BACKGROUND THERAPY DEFINITIONS, AESIS AND RELATED HEPATIC DISORDER 78

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

ADA:	anti-drug antibodies
AE:	adverse event
AESI:	adverse event of special interest
ALP:	alkaline phosphate
ALT:	alanine transaminase
ANCA:	anti-neutrophil cytoplasmic antibodies
AST:	aspartate transaminase
ATC:	anatomic or therapeutic categories
AUC:	area under the curve
BM:	biomarker
BMI:	body mass index
CCL2:	chemokine (C-C motif) ligand 2
CMQ:	customized MedDRA query
COMP:	cartilage oligomeric matrix Protein
cpk:	creatinine phosphokinase
CRISS:	composite response index for diffuse cutaneous systemic sclerosis
dcSSc:	diffuse cutenous systemic sclerosis
EBLUP:	empirical best linear unbiaised prediction
ECL:	electrochemiluminescence
e-CRF:	electronic case report form
ELISA:	enzyme linked immuno sorbent assay
ESR:	erythrocyte sedimentation rate
FVC:	forced vital capacity
HEVA:	health economics and value assessment
HLGT:	high-level group term
HLT:	high-level term
HRCT:	high resolution computer tomography
hsCRP:	high-sensitivity C-reactive protein
ILD:	interstitial lung disease
IMP:	investigational medicinal product
IRT:	interactive response technology
ITT:	intent-to-treat
LLT:	lower level term
MAR:	missing at random
MedDRA:	Medical Dictionary for Regulatory Activities
MMRM:	mixte model with repeated measures
mRSS:	modified Rodnan skin score
NMAR:	not missing at random
PAH:	pulmonary arterial hypertension
PK:	pharmacokinetics
PT:	preferred term

QOL: quality of life
QQ: quantile
SAE: serious adverse event
SD: standard deviation
SHAQ: scleroderma health assessment questionnaire
SMQ: standardized MedDRA query
SOC: sytemic organ class
SSc: systemic sclerosis
TBILI: total bilirubine
TJC28: tender joint count 28
UCLA SCTC GIT 2.0: UCLA scleroderma clinical trial consortium gastrointestinal tract 2.0
WHO-DD: World Health Organization Drug Dictionnary

1 OVERVIEW AND INVESTIGATIONAL PLAN

1.1 STUDY DESIGN AND RANDOMIZATION

This study will be a multinational, randomized, double blind, placebo controlled, 2 parallel groups, proof of concept Phase 2 study to assess the efficacy and safety of SAR156597 200 mg administered subcutaneously once a week over a 24 week period to patients with diffuse Systemic Sclerosis (SSc).

After a screening phase of up to 4 weeks, patients will be centrally randomized (using allocation from block randomization schedule with stratifying factors) via Interactive Response Technology (IRT) in a 1:1 ratio to 1 of the 2 treatment groups and treated double-blind for approximately 6 months.

Randomization will be stratified based upon the patients' medical history of SSc ILD (yes or no).

1.2 OBJECTIVES

1.2.1 Primary objectives

To evaluate, in comparison with placebo, the efficacy of SAR156597 administered subcutaneously for 24 weeks on skin fibrosis in patients with diffuse cutaneous SSc (dcSSc).

1.2.2 Secondary objectives

The secondary objectives are:

- To evaluate the efficacy of SAR156597 compared to placebo on physical/functional disability in patients with dcSSc.
- To evaluate the efficacy of SAR156597 compared to placebo on respiratory function of patients with dcSSc.
- To evaluate the safety profile of SAR156597 compared to placebo in patients with dcSSc.
- To evaluate the potential for immunogenicity (Anti-drug antibodies (ADA) response) of SAR156597 in patients with dcSSc.
- To evaluate the Pharmacokinetics (PK) (trough plasma concentrations) of SAR156597 administered subcutaneously for 24 weeks.

1.2.3 Exploratory objectives

The exploratory objectives are:

- To explore the efficacy of SAR156597 compared to placebo on other manifestations of SSc (gastrointestinal, joint pain, cardiovascular and renal manifestations) in patients with dcSSc.
- To explore the effect of SAR156597 on the Quality Of Life in patients with dcSSc.
- To measure the effect of SAR156597 on biomarkers of the disease activity and the IL 4/IL 13 pathway.
- To identify biomarkers predictive of response to treatment as defined by mRSS endpoints.

1.3 DETERMINATION OF SAMPLE SIZE

The targeted treatment benefit is a difference of 4 points between treatment groups in the mean change in mRSS from baseline to Week 24.

The impact of treatment discontinuations in the context of an ITT analysis where all modified Rodnan Skin Score (mRSS) data will be included in the analysis (regardless of adherence to treatment) was evaluated. Assuming that 10% of patients will discontinue the treatment, the estimated treatment effect at 24 weeks will be decreased from 4 (targeted treatment effect if all patients adhered to treatment) to 3.6 (targeted treatment effect in all randomized patients).

Ninety four (94) patients (47 patients each in SAR156597 and placebo groups) will yield 80% power to detect a difference between SAR156597 and placebo groups of 3.6 in the mean change from baseline in mRSS at 24 weeks, assuming a Standard Deviation (SD) of 7 and using a 1 sided alpha of 5% (Type I error).

Table 1 provides power calculations depending on several assumptions for treatment effect (difference between treatment groups in the mean change in mRSS from baseline to Week 24) and SD of the response (1) (assumed to be the same in both groups (2, 3).

Table 1 - Power calculations for mRSS, depending on treatment effect and standard deviation, using a 1-sided alpha of 0.05

Treatment effect	SD	Power
Δ=3	8	57%
	7	67%
	6	78%
Δ=3.6	8	70%
	7	80%
	6	90%
Δ=4	8	78%
	7	87%
	6	94%

Calculations were made using East 6.3 software.

1.4 STUDY PLAN

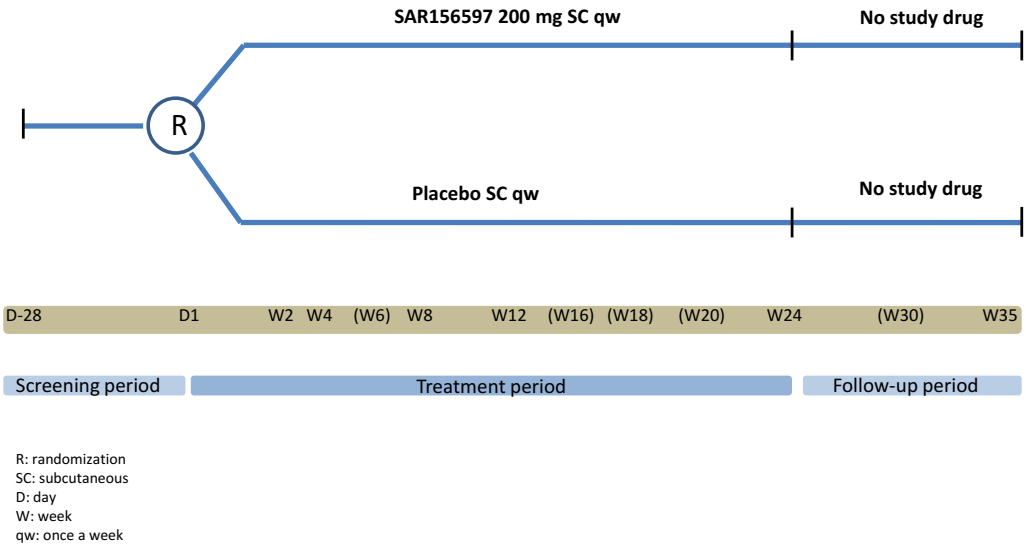
Approximately 94 patients will be randomized in a ratio 1:1 to the following two treatment groups:

- SAR156597 group (n=47): Patients will receive SAR156597 administered subcutaneously in 200 mg doses qw.
- Placebo group (n=47): Patients will receive placebo subcutaneously qw.

Randomization will be stratified based upon the patients' medical history of SSc-ILD (yes or no).

The following figure presents the graphical study design.

Figure 1 - Graphical study design (screening period, double-blind period and follow-up period)



1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

The statistical section of the protocol was never changed in an amendment.

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

Some minor modifications were done in this new version:

- A second definition of background therapy was added (Section 2.4.2).
- The description of all efficacy endpoints at each visit was added (Section 2.4.4).
- One sensitivity analysis on high mRSS was added (Section 2.4.4.1.6).

- Three subgroups were added to analyze the primary criteria: second definition of background therapy, geographical region and ANA staining pattern (Section 2.4.4.1.8).
- The Section 2.4.4.4.8 and Section 2.4.4.4.9 were updated according to the Appendix B from SAP.
- The observed FVC at Week 24 will be analyzed instead of the change in observed FVC from baseline to Week 24 in the linear mixed model (Section 2.4.4.4.2).

2 STATISTICAL AND ANALYTICAL PROCEDURES

2.1 ANALYSIS ENDPOINTS

2.1.1 Demographic and baseline characteristics

The baseline value is defined as the last available value before the first administration of the Investigation Medicinal Product (IMP).

All baseline safety and efficacy parameters are presented along with the on-treatment summary statistics in the safety and efficacy sections (Section 2.4.4 and Section 2.4.5).

Demographic characteristics

Demographic variables are gender (Male, Female), ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not reported, Unknown), race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Not reported, Unknown), age in years (quantitative and qualitative variable: <45, [45-65[, [65-75[and ≥75 years) and Body Mass Index (BMI) in kg/m² (quantitative variable and qualitative variable: <30 and ≥30).

Medical or surgical history

Medical (or surgical) history includes previous relevant medical and surgery history collected at baseline.

This information will be coded using the version of Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at Sanofi at the time of database lock.

Disease characteristics at baseline

Specific disease history includes disease duration from the time of first non-Raynaud's phenomenon manifestation (months), any systemic sclerosis history and mRSS at baseline. The auto-antibodies specific to systemic sclerosis (centromere antibodies, RNA polymerase III antibody, Scl-70 antibody and ANA staining pattern) and stratification factor (with/without medical history of SSc-ILD) as per IRT and as per electronic case report form (e-CRF) will be also presented.

Alcohol habits at baseline

Alcohol habits (frequency of alcoholic drinks in the last 12 months (Never/At least monthly/ At least weekly/At least daily/Occasionally), number of standard drinks (1 or 2/Greater than 2 where standard drink means 1 pint/bottle of beer, 1 glass of wine, 1 shot of hard liquor…) per day when drinking alcohol.

Other baseline characteristics

Other baseline characteristics include weight in kilograms (quantitative variable and qualitative variable: <50, [50-100[, ≥100).

Any technical details related to computation, dates, and imputations for missing dates are described in Section 2.5.

2.1.2 Prior or concomitant medications

All immunosuppressive and investigational drugs taken within 3 months before screening visit, all antibiotics taken within 1 month before screening visit, and all medications taken within 28 days prior to baseline visit and until the end of the study are to be reported in the case report form pages.

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) using the version currently in effect at Sanofi at the time of database lock.

- Prior medications are those the patient used prior (from 3 months before screening visit) to first investigational medicinal product (IMP) intake. Prior medications can be discontinued before first administration or can be ongoing during treatment phase.
- Concomitant medications are any treatments received by the patient concomitantly to the IMP, from randomization to the end of treatment + 84 days. A given medication can be classified both as a prior medication and as a concomitant medication. Concomitant medications do not include medications started during the post-treatment period (as defined in the observation period in Section 2.1.5).
- Post-treatment medications are those the patient took in the period running from the day after last administration of IMP + 84 days up to the end of the study.

Any technical details related to computation, dates, imputation for missing dates are described in Section 2.5.

2.1.3 Efficacy endpoints

All measurements, scheduled or unscheduled, will be assigned to analysis windows defined in Section 2.5.4 in order to provide an assessment of the efficacy parameter at each time point planned to be collected as per protocol. For all post-baseline time points, the value used for the analyses at a given time point (eg, at Week 24) is the valid value obtained within the corresponding analysis window. If multiple valid values of a variable exist within an analysis window, the nearest from the targeted study day will be selected. In case of ties, the best value will be considered. The baseline value is the last available and valid measurement obtained up to the date and time of the first double-blind IMP injection. For patients randomized and not treated, the baseline value is defined as the last available and valid value obtained up to the date and time of randomization. All efficacy endpoints will be analyzed using the ITT population.

2.1.3.1 Primary efficacy endpoint(s)

The primary efficacy endpoint will evaluate the efficacy of SAR156597 on skin thickening (ie, fibrosis) of patients with dcSSc by measuring:

- Change in mRSS from baseline to Week 24.

The mRSS assesses skin thickness in 17 body surface areas (face, chest, abdomen, right and left fingers, hands, forearms, upper arms, thighs, lower legs and feet). Each area is assessed for thickness on a 0 to 3 scale (0 = normal, 1 = mild but definite thickening, 2 = moderate skin thickening and 3 = severe skin thickening). The total mRSS score (the sum of score from all 17 body areas) ranges from 0 (no thickening) to 51 (severe thickening in all 17 areas). If one or more the scores from the 17 body areas is missing, then the mRSS is missing (see Appendix C). The lower the mRSS is, the better it is for the patient.

2.1.3.2 Secondary efficacy endpoints

Two secondary endpoints will evaluate the efficacy of SAR156597 on other aspects of dcSSc:

- Change in HAQ-DI, assessed with SHAQ, from baseline to Week 24.
- Change in respiratory function as measured by observed Forced Vital Capacity (FVC) and observed DLco (corrected for hemoglobin) from baseline to Week 24.

The HAQ-DI assesses the extent of the patient's functional ability; specifically their usual abilities using their usual equipment over the past week. It contains 8 domains of activity (dressing and grooming, arising, eating, walking, hygiene, reach, grip, and common daily activities) each of which has at least 2 questions, for a total of 20 items (see Appendix D). For each item, patients report the amount of difficulty experienced performing the activity. There are 4 possible responses for each item ranging from 0 (without any difficulty) to 3 (unable to do).

Each domain is scored separately. A domain score is determined from the highest score of the items in that domain. For example, if there are two items (as in the domain "arising"), and the patient responds with a 1 and 2, respectively, to the two items, the score for the "arising" domain will be a 2. If a question is left blank or is confusing, then the score for that category is determined by the remaining completed question(s).

For each of the eight domains there is an "aids or devices" companion variable(s) that is used to record the type of assistance, if any, a patient uses for his/her usual activities. The relationship between aids and devices and the domains (disability categories) is presented in Table 2. Use of aids or devices, or requirement of help from another person impacts on the scoring of the domains. When there are no aids or devices or help indicated for a domain, the category's score is not modified. When aids or devices or help are indicated by the patient, the score for the domain is raised from a 0 or a 1 to a 2, but if the patient's highest score for that domain is a 2 or 3, the score is kept as it is.

No adjustment is made if the patient indicated only "other" under aids or devices, without any other aid, device or help. The data entered at field "Other specify" will not be used for score adjustment.

Table 2 - Relationship between aids or devices and the domains (disability categories)

Disability Category	Aids or devices
Dressing & Grooming	Devices used for dressing (button hook, zipper pull, long handled shoe horn, etc.)
Arising	Built up or special chair
Eating	Built up or special utensils
Walking	Cane, walker, crutches, wheelchair
Hygiene	Raised toilet seat, bathtub seat, bathtub bar
	Long handled appliances in bathroom
Reach	Long handled appliances for reach
Grip	Jar opener (for jars previously opened)

The HAQ-DI composite score is calculated as the average of the scores of the domains. The patient must have a score for at least 6 of the 8 domains for the composite score to be calculated.

The HAQ-DI composite score is obtained by dividing the sum of the domains by the number of answered domains. This gives a score between 0 (no impairment in function) and 3 (maximal impairment of function).

If less than 6 domains have been scored then the HAQ-DI composite score is missing.

2.1.4 Exploratory endpoints

- Proportion of patients with improvement in mRSS of at least 20%, 40% and 60% from baseline to Week 24.
- Change in VAS for pain, breathing function, vascular function (Raynaud's phenomenon), gastrointestinal function, digital ulcers, and global assessment from SHAQ from baseline to Week 24 (see Appendix D).

Each VAS is scored between 0 and 100 measuring (using a ruler) how many mm from 0 the patient has placed their mark on the VAS (ie, where the mark crosses the VAS line). Each VAS is scored independently; no total score is calculated. Missing data is considered missing, and not imputed.

- Change in respiratory function as measured by % predicted FVC and % predicted DLco (corrected for hemoglobin) from baseline to Week 24.
- Change in UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract 2.0 (UCLA SCTC GIT 2.0) score from baseline to Week 24.

The UCLA SCTC GIT 2.0 Questionnaire includes 34 items and 7 multi-item scales (reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning) to assess quality of life (QOL) related to gastrointestinal function in patients with SSc. All scales are scored from 0.00 (better QOL) to 3.00 (worse QOL) except the diarrhea and

constipation (range from 0.00-2.00 and 0.00-2.50, respectively). The UCLA SCTC GIT 2.0 questionnaire provides a total score of GIT severity: it is calculated as the average of all scales (except constipation) and ranges from 0.00 to 2.83 (see Appendix E). If one item is missing, the UCLA SCTC GIT 2.0 score will be considered missing.

- Change in Tender Joint Count 28 (TJC28) from baseline to Week 24.

The TJC28 is an assessment of the overall joint pain based upon the examination of 28 key joints (see Appendix H). With the individual missing joint score (the "replaced or fused" joints are not taken into consideration with regards to tenderness) imputed as the mean of the scored joints, the tender joint count calculation after imputation is as follows:

- TJC28 = sum (scored tender joints) * (number of joints in the full joint set/number of scored tender joints).

The number of joints in the full joint set is defined as (28 - number of replaced or fused joints) and the scored joints refer to those with an answer (0 - no pain, 1 - pain).

- Change in digital ulcer count from baseline to Week 24.

The digital ulcer count corresponds to the sum of active or new plus indeterminate/healing digital ulcers secondary to SSc (see Appendix I).

- CRISS.

The Composite Response Index for Diffuse Cutaneous Systemic Sclerosis (CRISS) tool summarizes the changes in the clinical and patient-reported outcomes using a single composite score that reflects the probability that the patient with dcSSc has improved.

CRISS is obtained in a 2-step process as described below.

Step 1: Patients who develop new or worsening of cardiopulmonary and/or renal involvement due to SSc are considered as not improved (irrespective of improvement in other core items) and assigned a probability of improving equal to 0.0. Specifically if a subject develops any of the following:

- New scleroderma renal crisis.
- Decline in FVC% predicted ≥15% (relative), confirmed by another FVC% within a month, high resolution computer tomography (HRCT) to confirm ILD (if previous HRCT of chest did not show ILD) and FVC% predicted below 80% predicted*.
- New onset of left ventricular failure (defined as left ventricular ejection fraction ≤45%) requiring treatment*.
- New onset of Pulmonary Arterial Hypertension (PAH) on right heart catheterization requiring treatment*. PAH is defined as mean pulmonary artery pressure ≥25 mm Hg at rest and an end-expiratory pulmonary artery wedge pressure ≤15 mm Hg and a pulmonary vascular resistance >3 Wood units.

*Attributable to SSc

Step 2: For the remaining patients, Step 2 involves computing the predicted probability of improving for each subject using the following equation (equation to derive predicted probabilities from a logistic regression model):

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

ΔmRSS indicates the change in mRSS from baseline, ΔFVC denotes the change in FVC% predicted from baseline, ΔPt-glob indicates the change in patient global assessment (as per Likert scale), ΔMD-glob denotes the change in physician global assessment (as per Likert scale) (see Appendix G), and ΔHAQ-DI is the change in HAQ-DI. All changes are absolute change ($Time_2$ - $Time_{baseline}$).

- Change in EQ-5D-5L index from baseline to Week 24.

EQ-5D-5L (see Appendix F) is a generic measure of health status developed by the EuroQol Group in order to provide a simple measure of health for clinical and economic appraisal. It measures 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each of the dimensions is divided into 5 levels of perceived problems ranging from 1 (no problem) to 5 (extreme problem).

For each patient, a 5-digit score is obtained by combining each of the 5 dimensions. For example, score 11111 indicates no problems on any of the 5 dimensions, while score 12345 indicates no problems with mobility, slight problems with washing or dressing, moderate problems with doing usual activities, severe pain or discomfort and extreme anxiety or depression. For each patient, the EQ-5D-5L index, ranging from -0.594 to 1.0 is derived from this score using the United Kingdom population-based scoring (4, 5).

In addition, the EQ-5D has a single VAS to evaluate perceived current health. The VAS is scored between 0 and 100 by using the values indicated on the VAS where the patient has placed their mark, or by using the value written by the patient.

For index and VAS, missing data will not be imputed. Ambiguous values (eg, 2 boxes are ticked for a single dimension) on the health state/index score are treated as missing.

- Change in efficacy endpoints (mRSS, HAQ-DI, VAS from SHAQ, observed FVC, % predicted FVC, observed DLco [corrected for hemoglobin], % predicted DLco [corrected for hemoglobin], UCLA SCTC GIT 2.0 score, TJC28, digital ulcer count, CRISS, and EQ-5D-5L index and VAS) from baseline to Week 35 (up to end of follow-up period) and proportion of patients with improvement in mRSS of at least 20%, 40% and 60% from baseline to Week 35.

2.1.5 Safety endpoints

The safety analysis will be based on the reported adverse events and other safety information, such as clinical laboratory data, vital signs, ECG.

Observation period

The observation period will be divided into 4 epochs:

- The **pre-treatment** epoch is defined as the time from the signed informed consent date up to the first administration of IMP.
- The **treatment** epoch is defined as the time from the first administration of the IMP to the last administration of the IMP + 7 days.
- The **residual treatment** epoch is defined as the time from the last administration of the IMP + 8 days to the last administration of the IMP + 84 days.

The treatment-emergent adverse event (TEAE) period will include both **treatment** and **residual treatment** epochs.

- The **post-treatment** epoch is defined as the period of time starting the day after the end of the TEAE period up to the end of the study (defined as last protocol-planned visit or last study assessment).

The on-study observation period is defined as the time from the first administration of IMP until the end of the study (defined as last protocol planned visit or last study assessment).

2.1.5.1 Adverse events variables

Adverse event observation period

- Pre-treatment adverse events are adverse events that developed or worsened or became serious from the signed informed consent date up to first administration of IMP.
- Treatment-emergent adverse events are adverse events that developed or worsened or became serious during the treatment-emergent adverse event period.
- Post-treatment adverse events are adverse events that developed or worsened or became serious during the post-treatment period.

All adverse events (including serious adverse events and adverse events of special interest (AESIs)) will be coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the version of MedDRA currently in effect at Sanofi at the time of database lock.

Occurrence of adverse events (including Serious Adverse Events (SAEs) and Adverse Event of Special Interest (AESIs)) are recorded from the time of signed informed consent until the end of the study.

Adverse events of special interest (AESIs) include the following terms (their complete descriptions are provided in the protocol):

- Pregnancy of a female subject entered in a study as well as pregnancy occurring in a female partner of a male subject entered in a study with IMP, selected using the e-CRF "Pregnancy" tick box on the AE page.
- Symptomatic overdose (serious or non-serious) with IMP, selected using the e-CRF "symptomatic overdose of the IMP" tick box on the AE page.
- Alanine transferase (ALT) >3 ULN, selected using laboratory data.
- [redacted]
- Anaphylactic reactions or acute allergic reactions that require immediate treatment, selected using a CMQ coding list (see Appendix J).
- Severe injection site reactions, selected using HLT = "Injection site reaction" and an intensity equal to severe.
- Tuberculosis, selected using a CMQ coding list and/or Initiation of medications for suspected tuberculosis, selected using a WHO-DD CDG00737 "initiation of medications for suspected tuberculosis" (see Appendix J).
- Acute renal failure, selected using a CMQ coding list (see Appendix J) or using the e-CRF "Acute renal failure" tick box on the AE page.

2.1.5.2 Deaths

The deaths observation period are per the observation periods defined above.

- Death on-study: deaths occurring during the on-study observation period.
- Death on-treatment: deaths occurring during the treatment-emergent adverse event period.
- Death post-study: deaths occurring after the end of the study.

2.1.5.3 Laboratory safety variables

Clinical laboratory data consists of blood analysis, including hematology, clinical chemistry, and urinalysis. Clinical laboratory values after conversion will be analyzed into standard international units and international units will be used in all listings and tables.

Blood samples for clinical laboratories will be taken at Visit 1 (Screening), Visit 2 (Day 1), Visit 4 (Week 4), Visit 5 (Week 8), Visit 6 (Week 12), Visit 7/EOT (Week 24), early termination and Visit 8/EOS (Week 35) unless otherwise specified. The laboratory parameters will be classified as follows:

- Hematology
  - **Red blood cells and platelets and coagulation**: Hemoglobin, hematocrit, red blood cell count, erythrocyte sedimentation rate (ESR) and platelet count,
  - **White blood cells:** white blood cell count with differential blood count (neutrophils, lymphocytes, monocytes, basophils, eosinophils),
  - **Tuberculosis screen:** a Quantiferon®-TB gold evaluation.
- Biochemistry
  - **Metabolism:** Fasting glucose, total protein, albumin, high-sensitivity C-reactive protein (hsCRP),
  - **Electrolytes**: sodium, potassium, chloride, calcium, phosphorous, bicarbonate,
  - **Renal function**: serum creatinine, creatinine phosphokinase (CPK), blood urea nitrogen, clearance of creatinine according to Cockroft and Grault formula (mL per minute) (qualitative variable: <30, [30-50[, [50-80], >80),
  - **Pregnancy test**: (For women of childbearing potential) β-HCG blood test and urine pregnancy tests will be performed,
  - **Liver function**: alkaline phosphate (ALP), Alanine transaminase (ALT), Aspartate transaminase (AST), total bilirubine (TBILI),
  - [redacted]
- Serology
  - **Hepatitis screen:** hepatitis B surface antigen, hepatitis B surface antibody, HBV DNA, hepatitis C antibody, hepatitis C RNA, HIV-1/HIV-2 antibody,
  - [redacted]
- Urine sample
  - **Urinalysis** (by dipstick): specific gravity, pH, glucose, ketones, occult blood, protein, nitrate, leukocyte esterase, urobilinogen and bilirubin.

Technical formulas are described in Section 2.5.1.

2.1.5.4 Vital signs variables

Vital signs include: systolic and diastolic blood pressure in sitting position (mmHg), weight (kg) and heart rate (beats per minute).

2.1.5.5 Electrocardiogram variables

ECGs were recorded automatically by the device at the Investigator site.

ECG parameters include heart rate (beats per minute), PR interval (msec), QRS duration (msec), QT interval (msec) and corrected QTc (according to Bazett and Fridericia, see Section 2.5.1), ST deviation (mm), T-wave and U-wave morphologies (category as normal or abnormal).

2.1.6 Pharmacokinetic variables

All measurements (scheduled or unscheduled) will be assigned to analysis windows, as defined in Section 2.5.4, in order to provide an assessment of the PK parameter at each time point planned to be collected as per protocol.

C_{trough}, $C_{trough,av,ss}$) and $C_{Follow\text{-}Up}$ will be defined as follows:

- C_{trough}: SAR156597 concentration sample within 2h before each administration (Day 1, Week 4, Week 8, Week 12 and Week 24).
- $C_{trough,av,ss}$: SAR156597 average trough concentrations ("$C_{trough,av,ss}$") will be calculated for each patient as the mean of "$C_{troughs}$" considered at steady state (ss). The occurrence of steady state will be assessed graphically, by plotting "C_{trough}" throughout the study visits over all patients.
- $C_{Follow\text{-}up}$: SAR156597 concentration sample taken in Week 35, 12 weeks after last injection.

2.1.7 Immunogenicity variables

All measurements (scheduled or unscheduled) will be assigned to analysis windows defined in Section 2.5.4, in order to provide an assessment of the Anti-SAR156597 antibodies (ADA) parameter at each time point planned to be collected as per protocol.

ADA positive patients are patients with at least 1 treatment-induced or treatment-boosted ADA positive sample during the TEAE period, where:

- Treatment induced ADAs: ADAs developed de novo (seroconversion) following administration of the biotherapeutic (ie, formation of ADAs any time after the initial drug administration in a subject without pre-existing ADAs). If the baseline ADA sample is missing or non-reportable and at least one reportable ADA sample is available during the treatment (including follow-up period) the baseline sample will be considered as "negative" for data analysis. This is considered being a conservative approach for ADA assessment.

- Treatment boosted ADA: Pre-existing ADAs that were boosted to a higher level following administration of biotherapeutic (ie, any time after the initial drug administration) the ADA titer is significantly higher than the baseline titer. A low serial dilution schema (2-fold or 3 fold) should be applied during titration. A difference in titer values of two titer steps between an on treatment or follow-up sample and its baseline sample is considered significant. For examples, at least a 4-fold increase in titers for 2-fold serial dilution schema (or 9–fold increase in titers for 3- fold serial dilution schema). If no titer could be determined for a positive sample, the titer will be reported as the MRD of the assay.

The rest will be classified as ADA negative or inconclusive patients.

The following variables will be described:

- ADA response (Positive, Negative or inconclusive). For ADA positive:
  - Titer levels,
  - Neutralizing status (Positive or inconclusive).
- Pre-existing positive ADA defined as patients with positive ADA response at baseline with less than a specific increase (depending on titer calculation, eg, 4-fold increase in titer for 2-fold serial dilution schema, 9-fold increase in titer for 3-fold serial dilution schema) in the post-baseline period.
- Treatment-emergent positive ADA response defined as 1) Patients with no ADA positive response at baseline but with any positive response in the post-baseline period (up to follow-up visit) or 2) Patients with a positive ADA response at baseline and at least a specific increase (depending on titer calculation, eg, 4-fold increase in titer for 2-fold serial dilution schema, 9-fold increase in titer for 3-fold serial dilution schema) in titer in the post-baseline period (up to follow-up visit). For treatment-emergent positive ADA, the following categories for ADA duration will be applied:
  - A persistent positive response is a treatment-emergent ADA positive response detected in at least 2 or more post-baseline samples separated by at least a 16-week period (irrespective of any negative sample in between),
  - An indeterminate duration positive response is defined as ADA present only at the last sampling time point (and all previous samples negative) or with two last samples are positives but separated by a period less than 16 weeks.
- A transient positive response is defined as
  - Treatment induced ADA detected only at one sampling time during treatment or follow-up observation period (excluding last sampling time),
  - Treatment induced ADA detected at two or more sampling time during treatment where the first and last ADA positive sample are separated by a period less than 16 weeks and last sampling time is negative.
- Duration of ADA defined as longevity of treatment induced ADA (in days): date of last treatment induced ADA sample minus date of first treatment induced or treatment boosted ADA sample + 1. This will be calculated only for subjects with at least 2 positive ADA samples.
- Time to onset of treatment-emergent ADA positive response (in days): date of first ADA positive - date of first IMP administration.

2.1.8 Biomarker endpoints

All measurements (scheduled or unscheduled) will be assigned to efficacy analysis windows defined in Section 2.5.4, in order to describe all biomarkers (BM) assessments by time points even in case of premature treatment discontinuation.

Peripheral blood samples are planned to be collected for measurement of protein biomarkers at the study Visits 2 (baseline visit), Visit 6 (Week 12), Visit 7 (Week 24-EOT), Visit 8 (Week 35-EOS).

The protein biomarkers include, but are not limited to:

- Cartilage Oligomeric Matrix Protein (COMP) analyzed by Enzyme Linked Immuno Sorbent assay (ELISA) method.
- Chemokine (C-C Motif) Ligand 2 (CCL2) analyzed by ELISA method.
- Chemokine (C-C Motif) Ligand 17 - TARC analyzed by ELISA method.
- Periostin analyzed by ELISA method.
- Eotaxin-3 analyzed by Electrochemiluminescence (ECL).

2.1.9 Quality-of-life endpoints

See Section 2.1.3.2.

2.1.10 Health economic endpoints

Health economic endpoints other than quality of life will be described in a separate dedicated health economics and value assessment (HEVA) statistical analysis plan.

2.2 DISPOSITION OF PATIENTS

This section describes patient disposition for both patient study status and the patient analysis populations.

Screened patients are defined as any patients who met the inclusion criteria and signed the informed consent.

Randomized patients consist of all patients with a signed informed consent form who have had a treatment kit number allocated and recorded in the IRT database, regardless of whether the treatment kit was used.

For patient study status, the total number of patients in each of the following categories will be presented in the clinical study report using a summary table:

- Screened patients.
- Screen failure patients and reasons for screen failure.
- Non-randomized but treated patients.
- Randomized patients.
- Randomized but not treated patients and reason for not being treated.
- Randomized and treated patients.
- Patients who complete and who did not complete the study treatment period as per protocol.
- Patients who discontinued study treatment sorted by main reason for permanent treatment discontinuation.
- Patients who complete and did not complete the study follow-up period as per protocol.
- Patients who discontinued study sorted by main reason for study discontinuation.
- Status at last study contact.

For all categories of patients (except for the screened and non-randomized categories) percentages will be calculated using the number of randomized patients as the denominator. Reasons for treatment and study discontinuation will be supplied in tables giving numbers and percentages by treatment group.

The cumulative incidence of premature treatment discontinuation (irrespective of the reason) and premature treatment discontinuation due to AEs will be presented graphically by treatment arm using Kaplan-Meier method. Not treated patients will be considered with event at Day 1 (day of randomization). All completers will be considered as right-censored observations. For randomized and treated patients, the time-to-event or censoring variable is defined as: last IMP administration date - randomization date + 7 days.

The cumulative incidence of premature study discontinuation will be presented graphically by treatment arm using Kaplan-Meier method. All completers will be considered as right-censored observations. The time-to-event of censoring variable is defined as:
end of study date - randomization date + 1 day, where end of study date is the date of last successful contact or visit, whichever comes last, or, for patients who died before this date, the date of death.

A patient is considered lost to follow-up at the end of the study if he/she is not assessed at the last protocol-planned visit and if the time from the last successful contact or visit, whichever comes last, to the last protocol-planned visit is greater than 2 days.

All critical or major deviations potentially impacting efficacy analyses, randomization, and drug-dispensing irregularities, and other major or critical deviations will be summarized in tables giving numbers and percentages of deviations by treatment group.

Additionally, the analysis populations for safety, efficacy, anti-SAR156597 antibody (ADA) and pharmacokinetics will be summarized in a table by number of patients on the randomized population.

- Efficacy population: intent-to-treat (ITT) population.
- Safety population.
- Anti-SAR156597 antibody population.
- Pharmacokinetics population.
- Biomarkers populations.

2.2.1 Randomization and drug dispensing irregularities

Randomization and drug-dispensing irregularities occur whenever:

1. A randomization is not in accordance with the protocol-defined randomization method, such as
   a) An ineligible patient is randomized,
   b) A patient is randomized based on an incorrect stratum, or
   c) A patient is randomized twice.

   OR
2. A patient is dispensed an IMP kit not allocated by the protocol-defined randomization, such as
   a) A patient at any time in the study is dispensed a different treatment kit than as randomized (which may or may not contain the correct-as-randomized IMP), or
   b) A non-randomized patient is treated with IMP reserved for randomized patients.

Randomization and drug-dispensing irregularities will be monitored throughout the study and reviewed on an ongoing basis.

All randomization and drug-dispensing irregularities will be documented in the clinical study report. If the number of irregularities is large enough to make a tabular summary useful, the irregularities will be categorized and summarized among randomized patients (number and percentages). Non-randomized, treated patients will be described separately.

Randomization and drug-dispensing irregularities to be prospectively identified include but are not limited to:

Randomization and drug allocation irregularities
Kit dispensation without IRT transaction
Erroneous kit dispensation (IMP number actually administered to the subject is different from the IMP number allocated by IRT)
Kit not available
Stratification error (Wrong stratum of randomization)
Patient switched to another site
Randomization of a non-existent subject

2.3 ANALYSIS POPULATIONS

Patients treated without being randomized will not be considered randomized and will not be included in any efficacy population. The safety experience of these patients will be reported separately, and these patients will not be in the safety population.

The randomized population includes any patient who has been allocated to a randomized treatment regardless of whether the treatment kit was used.

For any patient randomized more than once, only the data associated with the first randomization will be used in any analysis population. The safety experience associated with any later randomization will be assessed separately.

The safety experience of patients treated and not randomized will be reported separately, and these patients will not be in the safety population.

2.3.1 Efficacy populations

2.3.1.1 Intent–to-treat population

The intent-to-treat population is defined as all randomized patients. Patients in the ITT population will be analyzed according to the treatment group allocated by randomization.

2.3.2 Safety population

The safety population is defined as randomized patients who actually received at least 1 dose or part of a dose of the IMP, analyzed according to the treatment actually received.

In addition:

- Randomized patients for whom it is unclear whether they took the IMP will be included in the safety population as randomized.
- For patients receiving more than 1 IMP during the trial, the treatment group allocation for as-treated analysis will be the one received in the majority of injections.

2.3.3 Anti-SAR156597 antibody (ADA) population

The anti-SAR156597 antibody analysis will be performed on all randomized and treated patients (safety population) with at least one baseline and post dose ADA sample with a reportable result. Patients will be analyzed according to the treatment actually received.

2.3.4 Pharmacokinetics population

The PK population will include all randomized and treated patients (safety population) with at least one post-dose, non-missing plasma concentration value.

2.3.5 Biomarker population

Baseline biomarker population includes the randomized and treated patients with a baseline sample successfully analyzed.

Pharmacodynamics biomarker population includes the randomized and treated patients with a baseline sample and at least one post first double-blind IMP injection sample successfully analyzed.

2.4 STATISTICAL METHODS

Continuous data will be summarized using the number of available data, mean, SD, median, minimum, Q1, Q3 and maximum for each treatment group. Categorical and ordinal data will be summarized using the number and percentage of patients in each treatment arm.

2.4.1 Demographics and baseline characteristics

Parameters will be summarized on the randomized population analyzed in the treatment arm to which they were randomized. Analyses for the safety population will be included in the appendices if the size of the safety population is different (>10%) from the size of that in the primary analysis population for any treatment group.

Parameters described in Section 2.1.1 will be summarized by treatment group and overall treatment groups using descriptive statistics.

The disease duration, mRSS, FVC and DLCO at baseline will be also summarized by stratification factor (as per IRT).

Medical/surgical history will be summarized in each treatment arm by primary SOC and PT. Events will be sorted by SOC internationally agreed order and decreasing frequency of PT based on the incidence in the overall treatment arm.

P-values on demographic and baseline characteristic data will not be calculated.

No specific description of the safety parameters will be provided at baseline. If relevant, the baseline values will be described along with each safety analysis.

No specific description of the efficacy parameters will be provided at baseline. If relevant, the baseline values will be described along with each efficacy analysis.

2.4.2 Prior or concomitant medications

The prior and concomitant medications will be presented for the randomized population.

Medications will be summarized by treatment group according to the WHO-DD, considering the first digit of the anatomic category (ATC) class (anatomic category) and the first 3 digits of the ATC class (therapeutic category). All ATC codes corresponding to a medication will be summarized, and patients will be counted once in each ATC category (anatomic or therapeutic) linked to the medication. Therefore patients may be counted several times for the same medication.

The table for prior medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the overall incidence across treatment groups. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

Patients who received background therapy are defined as all patients who took a medication from the anatomic classes of “antineoplastic/immunomodulating agents” and/or “systemic hormonal preparations, excl. sex hormones and insulins” (see Appendix J) before and at baseline.

The second definition of patients who received background therapy is defined as all patients taken “methotrexate, mycophenolate mofetil, azathioprine, ciclosporin, and cyclophosphamide” (see Appendix J) before and at baseline.

The tables for concomitant and post-treatment medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the incidence in the SAR156597 group. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

The patients who took a rescue medication during the study are defined according to the deviation “Prohibited use of immunosuppressive therapies as concomitant medications”.

2.4.3 Extent of investigational medicinal product exposure and compliance

The extent of IMP exposure and compliance will be assessed and summarized by actual treatment within the safety population (Section 2.3.2).

2.4.3.1 Extent of investigational medicinal product exposure

The extent of IMP exposure will be assessed by the duration of IMP exposure.

Duration of IMP exposure in weeks is defined as (last dose date + 7 - first dose date)/7 days, regardless of unplanned intermittent discontinuations (see Section 2.5.3 for calculation in case of missing or incomplete data).

Duration of IMP exposure will be summarized descriptively as a quantitative variable (number, mean, SD, median, minimum, and maximum) expressed in weeks. In addition, duration of treatment exposure will also be summarized categorically by numbers and percentages for each of

the following categories and cumulatively according to these categories: ≥1 day and <2 weeks, ≥2 weeks and <4 weeks, ≥4 weeks and <6 weeks, ≥6 weeks and <12 weeks, ≥12 weeks and <16 weeks, ≥16 weeks and <20 weeks, ≥20 weeks and <24 weeks, ≥24 weeks and <36 weeks. Non-integer values will be rounded to 1 decimal place.

Additionally, the cumulative duration of treatment exposure will be provided, defined as the sum of the duration of treatment exposure for all patients, and will be expressed in patient-years.

2.4.3.2 Compliance

A given administration will be considered non-compliant if the patient did not take the planned dose of treatment as required by the protocol. No imputation will be made for patients with missing or incomplete data.

Overall percentage of compliance for a patient will be defined as 100 - (% days with below-planned dosing + % days with above-planned dosing), considering that injections should be performed every week ±2 days as per protocol.

Above-planned dosing percentage for a patient will be defined as 100 * the number of days with more than 1 injection administration within the 5 days before divided by the duration of IMP injection exposure in days.

Below-planned dosing percentage for a patient will be defined as 100 * the number of days with no injection administration within the previous 9 days divided by the duration of IMP injection exposure in days.

Treatment compliance, above-planned, and under-planned dosing percentages will be summarized descriptively as quantitative variables (number, mean, SD, median, minimum, and maximum). The percentage of patients whose compliance is <80% will be summarized. In addition, numbers and percentages of patients with [0;5],]5;10],]10;20], and >20% of days with above or under-planned dosing will also be provided.

In addition, the following parameters will be summarized by treatment arm:

- The number of advanced injections (≤3 days apart from the previous injection, 4 days apart from the previous injection) and the number of delayed injections (more than 9 days between two consecutive injections).
- The number of patients with at least one advanced injection and the number of patients with at least one delayed injection. A patient could be counted in several categories.
- The frequency of IMP injections (mean (SD), median, Min:Max description) will be defined for each patient as the average number of days between two consecutive injections: (last injection date - first injection date)/(number of injections -1) for patients with at least 2 injections.

Cases of overdose are reported in the AE e-CRF pages as AESI if symptomatic or AE if asymptomatic. The reported cases of overdose will be described in the AE analysis (see Section 2.1.5 and Section 2.4.5.1).

2.4.4 Analyses of efficacy endpoints

All efficacy measurements, scheduled or unscheduled, will be assigned to time points according to analysis windows as defined in Table 3.

All endpoints will be described at each visit on the ITT population.

2.4.4.1 Analysis of primary efficacy endpoint(s)

2.4.4.1.1 Primary efficacy analysis

The primary estimand will be the difference in mean change from baseline to 24 weeks in mRSS in all randomized patients, regardless of whether or not patients completed the treatment period. This estimand corresponds to a "treatment policy strategy".

A spaghetti plot presentation of individual data by treatment group will be presented.

The change in mRSS from baseline to Week 24 will be analyzed in the ITT population using a MMRM approach. All post-baseline data available from Week 4 to Week 24 analysis windows will be included in the analysis, regardless of adherence to treatment. The lower the mRSS is, the better it is for the patient.

Missing data will be accounted for by the MMRM. The model includes the fixed categorical effects of treatment group (placebo, SAR156597), randomization strata (as per IRT, SSc-ILD: Yes/No), time point (Week 4, Week 8, Week 12, Week 24 as defined in Section 2.5.4), randomization strata-by-time point interaction and treatment-by-time point interaction, as well as the continuous fixed covariates of baseline mRSS value and baseline value-by-time point interaction.

This model will be run using SAS Mixed procedure. The repeated-measures analysis will be based on the restricted maximum likelihood method (Newton-Raphson algorithm) assuming an unstructured correlation matrix to model the within-patient errors. Denominator degrees of freedom will be estimated using Kenward-Roger approximation. This model will provide baseline adjusted LSmeans estimates at Week 24 for both treatment groups with their corresponding standard errors (SEs) and 95% confidence intervals (CIs). To compare SAR156597 to the placebo group, an appropriate contrast statement will be used to test the difference of these estimates at the 5% one-sided alpha level. The 95% and 90% confidence intervals of the difference will be provided. Let $\mu 0$ and $\mu 1$ be the population means of the change from baseline in mRSS at Week 24 under placebo and SAR156597, respectively. The null hypothesis that will be tested is "H0: $\mu 0 = \mu 1$" versus "H1: $\mu 0 > \mu 1$".

The MMRM model relies on the "missing-at-random" (MAR) assumption. As we can never exclude the possibility for a not-missing-at-random (NMAR) missingness mechanism, if more than 5 patients have no value for mRSS data at Week 24, a sensitivity analysis to explore the impact of non-ignorable missingness on the primary efficacy analysis will be conducted (see control-based pattern imputation model defined in Section 2.4.4.1.7).

2.4.4.1.2 Analysis of residuals

The analysis of the residuals of the MMRM will be primarily based on studentized residuals. The studentized residuals will be presented graphically using histogram and QQ-plot.

2.4.4.1.3 On-treatment analysis

A secondary estimand will be the difference in mean change from baseline to 24 weeks in mRSS estimated during the on-treatment period (from the first administration of the IMP to the last administration of the IMP + 7 days). This estimand corresponds to a "while on treatment strategy". This estimand will be considered for describing the effect of treatment as long as patients adhere to their randomized treatment.

As a secondary analysis, the change in mRSS from baseline to Week 24 will also be analyzed in the randomized and treated population using the same MMRM model as described above, including only post-baseline data measured during the on-treatment period (from the first administration of the IMP to the last administration of the IMP + 7 days), analyzed according to the treatment group allocated by randomization.

2.4.4.1.4 Sensitivity to error in randomization strata

In order to assess the robustness of the primary analysis to randomization stratum mistakes (ie, the stratum recorded in IRT differs from the actual one), if more than 15% of patients have been randomized to the wrong stratum, the MMRM model will be re-run including the actual stratum as per the eCRF instead of the stratum recorded in IRT.

2.4.4.1.5 Sensitivity analysis on rater

In order to assess the impact of different raters evaluating a same patient from baseline, the MMRM model will be re-run using only mRSS values from all visits evaluated by the same rater as baseline. All other values evaluated by different rater from baseline will be considered missing at random in this sensitivity analysis.

2.4.4.1.6 *Sensitivity* analysis on high mRSS

The MMRM model will be re-run using only patients with mRSS score superior or equal to 15 at baseline.

2.4.4.1.7 Control-based pattern imputation

In addition to the MMRM method, the control-based pattern imputation method will be used to address missing values, in the randomized population, followed by the testing of treatment arms using an analysis of covariance (ANCOVA) model, with the intent to evaluate the robustness of the primary analysis using a different statistical method.

Imputation rules

The control-based pattern imputation assumes:

- After treatment discontinuation (defined for this analysis as last injection + 7 days), patients from the experimental treatment arm (no longer receiving the experimental treatment) will exhibit the same future evolution of the disease as patients on the control treatment (who are also not exposed to the experimental treatment). In this experimental arm, missing values before treatment discontinuation will be considered missing at random and will be imputed using a model estimated using only samples from the experimental arm collected before treatment discontinuation.
- Missing on-treatment mRSS values in the experimental arm will be imputed from other on-treatment measurements in the experimental arm assuming MAR, using SAS® MI procedure. Only mRSS values collected during the on-treatment period will be included in the imputation model. This way, missing mRSS values during the on-treatment period will be imputed based solely on observed on-treatment mRSS values. The imputation model will include the randomization strata (as per IRT, SSc-ILD: Yes/No), baseline mRSS value, and all time points (Week 4, Week 8, Week 12, Week 24). Since the pattern of missing data will be non-monotone, a Monte-Carlo Markov Chain (MCMC) method will be used. A minimum value of 0 and a maximum value of 51 will be specified in order to respect the range value for mRSS.
- Whatever the timing of treatment discontinuation, patients who discontinue from the control arm are assumed to evolve in the same way as control patients who remain in the study. Therefore, their missing values will be considered as missing at random and imputed using a model estimated using all samples collected from the control arm.

Imputation method

To account for the uncertainty, missing values will be imputed 100 times to generate 100 complete datasets, using the MI SAS procedure. The change from baseline at Week 24 will be then derived from observed and imputed mRSS at this time point.

The 100 complete datasets will be then analyzed using an ANCOVA model with treatment group and randomization strata (as per IRT, as defined in Section 1.1) as fixed effects, and the baseline mRSS value as continuous covariate, and change in mRSS from baseline to Week 24 as explained variable. The MIANALYZE SAS procedure will be used to generate valid statistical inferences by combining results from the 100 analyses using Rubin's formulae.

The number of imputations (100) will be informally verified by replicating sets of 100 imputations and checking whether the combined results are stable. If not stable, the number of imputations will be increased and informally checked as above, and thus until stable estimates are obtained.

Non continuous variables included in the imputer's model (ie, treatment group, randomization strata) are not expected to be missing.

Only imputed values during the on-treatment period will be kept in the final datasets that will be analyzed using ANCOVA. Imputed values during the post-treatment period will be discarded and replaced by imputed values described in the next paragraph.

Imputation programming

In order to implement the control-based pattern imputation, we will break the imputation process into a sequence of multiple calls to PROC MI, where each call is intended to impute missing values at one time-point only. The general logic of such a strategy is as follows.

1. With each call to PROC MI, impute only one time-point using a regression-based MI method for monotone missingness. Variables included in the regression model are randomization strata and mRSS values from previous time-points (including baseline).
2. When imputing missing values for time-point t, the input dataset should include all control patients, but only those patients from the experimental arm who have values at time-point t missing (only those who need imputation at time-point t). Since patients from the experimental arm with non-missing values at time-point t are not included in the input dataset, they will not contribute to the estimation of an imputation model for time-point t. The imputation model will then be estimated using control patients only, while this call to PROC MI will impute missing data at time-point t for all patients who need imputation at that time-point. This way, patients from experimental arm will be imputed based on the control patients' model. Note that treatment arm should not be included as an effect in this model.
3. Repeat (2.) for all other time-points sequentially. Patients whose missing values were imputed in the last call to PROC MI will be included in the input dataset for the next call to PROC MI except for patients in the experimental arm having non-missing data at time-point t+1 (patients with non monotone missingness). Thus data for time-point t, imputed during the last call, will be used as predictor variables in the next call to PROC MI (for time-point t+1) (6).

This imputation allows the patient to continue from the level achieved under the active treatment as though it had been achieved under the reference treatment. One consequence is that if a patient on active is above the reference mean then this positive residual will feed through into subsequent observations, to a degree determined by the correlation pattern in the referenced arm. Hence, the patient's profile will slowly decay back toward the mean for reference at later times (7).

The imputed mRSS data will be between 0 and 51 (range value for mRSS score).

2.4.4.1.8 Subgroup analyses

To assess the homogeneity of the treatment effect across various subgroups, treatment-by-subgroup factor and time point-by-subgroup factor interaction terms and a subgroup factor term will be added in the primary MMRM model. Within each subgroup, baseline adjusted LSmeans estimates at Week 24 will be provided for both treatment groups with their corresponding SEs and 95% CIs, as well as their difference with the corresponding 95% and 90% Cis. Subgroups of interest are:

- Randomization stratum (as per IRT).
- Gender (male/female).
- Both Background therapy (yes/no) definitions defined in Section 2.4.2.
- Disease duration of non-Raynaud's phenomenon (<20 months vs ≥20 months).
- Auto antibodies (Centromere Antibodies, RNA Polymerase III Antibody, Scl-70 Antibody, [if size allows] and ANA staining pattern [Centromere vs. Non-Centromere staining pattern]).
- Geographical region (Estonia, Romania, Russia, Poland and Ukraine vs Rest of the World).

2.4.4.2 Analyses of secondary efficacy endpoints

All secondary endpoints are described in Section 2.1.3.2 and will be analyzed using the ITT population.

Change in continuous secondary efficacy endpoints (HAQ-DI composite score, observed FVC, observed DLco [corrected for hemoglobin]) from baseline to Week 24 will be analyzed using the same MMRM model as for the primary endpoint. Specifically, the model will contain fixed categorical effects of treatment group (placebo, SAR156597), randomization strata (as per IRT, SSc-ILD: Yes/No), time point (Week 4 [only for HAQ-DI], Week 8 [only for HAQ-DI], Week 12 and Week 24), randomization strata-by-time point interaction and treatment-by-time point interaction, as well as the continuous fixed covariates of corresponding baseline value and baseline-by-time point interaction. All post-baseline data available from Week 4 for HAQ-DI, Week 12 for FVC and DLco, to Week 24 analysis windows will be included in the analysis, regardless of adherence to treatment. Missing data will be accounted for by the MMRM. The studentized residuals of the MMRM will be analyzed as for the primary endpoint.

This model will provide baseline adjusted LSmeans estimates at Week 24 for both treatment groups with their corresponding standard errors (SEs) and 95% CIs. The SAR156597 to placebo difference of these estimates will be provided with its corresponding 95% and 90% CIs and p-value.

Within each randomization stratum (as per IRT, SSc-ILD: Yes/No), baseline adjusted LSmeans estimates at Week 24 will also be provided for both treatment groups with their corresponding SEs and 95% CIs, as well as their difference with the corresponding 95% and 90% CIs.

In order to assess the robustness of the analyses to randomization stratum mistakes (ie, the stratum recorded in IRT differs from the actual one), if more than 15% of patients have been randomized to the wrong stratum, the MMRM model will be re-run including the actual stratum as per the eCRF instead of the stratum recorded in IRT.

In addition, the observed FVC at Week 24 will also be analyzed through a linear mixed effects model. The model will assume random intercept and random slope to take into account between-subject heterogeneity and assume a linear evolution of the observed FVC over time (time will be included as a continuous variable in the model). The analysis will be also adjusted for fixed categorical effects of treatment group (placebo, SAR156597), randomization strata (as per IRT, SSc-ILD: Yes/No). The following interaction terms will be added to the model: treatment-by-time and randomization strata-by-time. All the visits are taken into account in this analysis.

2.4.4.3 Multiplicity issues

No adjustment will be made. For secondary and exploratory efficacy endpoints, p-values will be provided for descriptive purpose only.

2.4.4.4 Analyses of exploratory efficacy endpoints

All exploratory endpoints are described in Section 2.1.4 and will be analyzed using the ITT population unless otherwise specified.

Change in continuous exploratory efficacy endpoints from baseline to Week 24 will be analyzed using the same MMRM model as for the primary and continuous secondary efficacy endpoints. Specifically, the model will contain fixed categorical effects of treatment group, randomization strata (as per IRT), time point, randomization strata-by-time point interaction and treatment-by-time point interaction, as well as the continuous fixed covariates of corresponding baseline value and baseline-by-time point interaction. All post-baseline data available from Week 4 to Week 24 analysis windows will be included in the analysis, regardless of adherence to treatment. Missing data will be accounted for by the MMRM. This model will provide baseline adjusted LSmeans estimated at Week 24 in both treatment groups with their corresponding 95% CIs. The SAR156597 to placebo differences of these estimates will be provided with its corresponding 95% and 90% CIs and p-value.

Change in continuous exploratory efficacy endpoints from baseline to Week 35 will be analyzed using the same model. All post-baseline data available from Week 4 to Week 35 analysis windows will be included in the analysis, regardless of adherence to treatment.

2.4.4.4.1 mRSS

Responder rates (proportion of patients with an improvement from baseline in mRSS of at least 20%, 40% and 60%) at Week 24 and Week 35 will be analyzed using a logistic regression with the categorical effects of treatment group and randomization strata (as per IRT) and the continuous effect of mRSS at baseline. This model will provide the SAR156597 to placebo odds ratio estimate and its 95% and 90% confidence intervals. The p-value will be obtained from the

Wald Chi-square test. Patients with missing value at Week 24, respectively Week 35, will be considered as non-responders. Cumulative distribution function (CDF) graphs will be also presented (percentage of patients according to the relative improvement from baseline in mRSS) at W24 and W35.

Furthermore, the change from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

Three supplementary definitions for responder rate at Week 24 will be used and analyzed using a logistic regression with the categorical effects of treatment group and randomization strata (as per IRT) and the continuous effect of mRSS at baseline:

- Responder rate defined as the proportion of patients with an improvement from baseline in mRSS of at least 25% at Week 24.
- Responder rate defined as the proportion of patients with an improvement from baseline in mRSS of at least 25% at Week 24 and no intake of rescue medication.
- Overall responder rate defined as the proportion of patients with an improvement from baseline in mRSS of at least 25% and without CRISS event and no intake of rescue medication and alive at W24 are defined as responder.

2.4.4.4.2 VAS from SHAQ

Change in VAS from SHAQ from baseline to Week 24 and from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

CDF graphs will be also presented for each VAS (percentage of patients according to the relative improvement from baseline in VAS) at Week 24 and Week 35.

2.4.4.4.3 UCLA SCTC GIT 2.0 score

Change in UCLA SCTC GIT 2.0 score from baseline to Week 24 and from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

2.4.4.4.4 TJC28

Change in TJC28 from baseline to Week 24 and from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

2.4.4.4.5 Digital ulcer count

Change in digital ulcer count (sum of active or new plus indeterminate/healing digital ulcers secondary to SSc) from baseline to Week 24 and from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

The digital ulcer count, as well as the subcategories: number of active/new ulcers and number of indeterminate/healing ulcers, will also be descriptively summarized by treatment group and by time point.

2.4.4.4.6 EQ-5D-5L

Change in EQ-5D-5L (VAS and index) from baseline to Week 24 and from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4. The EQ-5D-5L (index and VAS) will be summarized descriptively at each time point.

CDF graphs will be also presented for EQ-5D-5L index and EQ-5D-5L VAS (percentage of patients according to the relative improvement from baseline in EQ-5D-5L value [index or VAS]) at Week 24 and Week 35.

2.4.4.4.7 HAQ-DI

Change in HAQ-DI total score from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

CDF graphs will be also presented (percentage of patients according to the relative improvement from baseline in HAQ-DI) at Week 24 and Week 35.

2.4.4.4.8 FVC

Due to the non-normality of the distribution of % predicted FVC, the change from baseline at Week 24, respectively Week 35, will be analyzed using a rank-based analysis of covariance (rank analysis of covariance [ANCOVA]) model adjusted for baseline and stratified on randomization stratum (as per IRT). This model will use the ranks of absolute change from baseline to Week 24 (respectively Week 35) in % predicted FVC as the outcome and the ranks of baseline % predicted FVC as a covariate stratified on the stratification factor (with/without ILD). In case of ties, the mean of the corresponding ranks will be assigned.

The ranking will be done within each stratum of the stratification factor (with/without ILD) in ITT population (both treatment groups combined).

- Missing data due to death will be ranked based on a time-to-death, with shortest time until death as the worst (lowest) rank (from 1 to $n_{stratum,\ death}$).
- Absolute change from baseline at Week 24 (respectively Week 35) in % predicted FVC values will be calculated from Week 24 (respectively Week 35) % predicted FVC values and then ranked (from $n_{stratum,\ death} + 1$ to $n_{stratum}$).
- Standardization of ranks using SAS PROC RANK.

Furthermore, the adjustment on baseline % predicted FVC covariate will be done. The residuals from the linear regression of response ranks on baseline ranks will be calculated within each stratum of the stratification factor (with/without ILD).

Finally, treatment effect will be tested using the residuals as scores. The active treatment group will be compared to placebo using a Mantel-Haenszel mean score chi-square test at a one-sided Type-I error level of 5%. This test will be stratified on the stratification factor.

In case of more than 20% of missing data on the % predicted FVC value at Week 24, missing data will be imputed following methods described in Section 2.5.2. Otherwise, the data will not be imputed.

Furthermore, the change in observed FVC from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

The number of patients with an absolute decreased % predicted FVC from baseline at W24 superior or equal to 5% and respectively 10% will be presented by treatment group.

2.4.4.4.9 DLco

Same rank ANCOVA as for % predicted FVC will be done for the change in % predicted DLco from baseline at Week 24 and Week 35, using % predicted DLco as the outcome and the ranks of baseline % predicted DLco as a covariate stratified on the stratification factor.

Furthermore the change in observed DLco (corrected for hemoglobin) from baseline to Week 35 will be analyzed using the MMRM model described in Section 2.4.4.4.

The number of patients with an absolute decreased % predicted DLCO from baseline at W24 superior or equal to 5% and respectively 10% will be presented.

2.4.4.4.10 CRISS

The distribution of the predicted probability of improving obtained using the CRISS at Week 24 and Week 35 will be compared between SAR156597 and the placebo group using a Van Elteren's test stratified on randomization strata.

In case of more than 20% of missing data on the CRISS value at Week 24, missing data will be imputed following methods described in Section 2.5.2. Otherwise, the data will not be imputed.

In addition, rate of patients for whom the predicted probability of improving obtained using the CRISS at Week 24 and Week 35 is ≥60% will be provided.

2.4.5 Analyses of safety data

The summary of safety results will be presented by treatment group.

General common rules

All safety analyses will be performed on the safety population as defined in Section 2.3.2, unless otherwise specified, using the following common rules:

- Safety data in patients who do not belong to the safety population (eg, exposed but not randomized) will be listed separately.
- The baseline value is defined as the last available value before the first administration of the IMP.
- The potentially clinically significant abnormality (PCSA) values are defined as abnormal values considered medically important by the Sponsor according to predefined criteria/thresholds based on literature review and defined by the Sponsor for clinical laboratory tests, vital signs, and ECG (PCSA version dated May 2014 [Appendix A]).

- PCSA criteria will determine which patients had at least 1 PCSA during the TEAE period, taking into account all evaluations performed during the TEAE period, including unscheduled or repeated evaluations. The number of all such patients will be the numerator for the treatment-emergent PCSA percentage.
- The treatment-emergent PCSA denominator by group for a given parameter will be based on the number of patients assessed for that given parameter in the TEAE period by treatment group on the safety population.
- For quantitative safety parameters based on central laboratory/reading measurements, descriptive statistics will be used to summarize results and change from baseline values by visit (baseline and post-baseline time points) and treatment group. Summaries will include the last and the worst on-treatment value. Only data sampled before or on the day of the last IMP administration will be included in the analysis. The worst on-treatment value is defined as the nadir and/or the peak post-baseline before or on the day of the last IMP administration according to the direction (minimum or maximum) of the abnormality as defined in the PCSA list.
- All measurements (scheduled or unscheduled) will be assigned to time points according to analysis windows as defined in Section 2.5.4.
- The analysis of the safety variables will be essentially descriptive and no systematic testing is planned. Relative risks versus placebo and their 95% confidence intervals may be provided, if relevant.

2.4.5.1 Analyses of adverse events

Generalities

The primary focus of adverse event reporting will be on TEAEs. Pre-treatment and post-treatment adverse events will be described separately.

If an adverse event date/time of onset (occurrence, worsening, or becoming serious) is incomplete, an imputation algorithm will be used to classify the adverse event as pre-treatment, treatment-emergent, or post-treatment. The algorithm for imputing date/time of onset will be conservative and will classify an adverse event as treatment emergent unless there is definitive information to determine it is pre-treatment or post-treatment. Details on classification of adverse events with missing or partial onset dates are provided in Section 2.5.3.

Adverse event incidence tables will present by SOC, HLGT, HLT, and PT, sorted in alphabetical order for each treatment group, the number (n) and percentage (%) of patients experiencing an adverse event. Multiple occurrences of the same event in the same patient will be counted only once in the tables within a treatment phase. The denominator for computation of percentages is the safety population within each treatment group.

Sorting within tables ensures the same presentation for the set of all adverse events within the observation period (pre-treatment, treatment-emergent, and post-treatment). For that purpose, the table of all TEAEs presented by SOC and PT sorted by the internationally agreed SOC order and decreasing frequency of PTs within SOCs will define the presentation order for all other tables unless otherwise specified. Sorting will be based on results for the experimental treatment arm.

Analysis of all treatment-emergent adverse events

The following TEAE summaries will be generated for the safety population.

- Overview of TEAEs, summarizing number (%) of patients with any
  - TEAE,
  - Serious TEAE,
  - TEAE leading to death,
  - TEAE leading to permanent treatment discontinuation.
- All TEAEs by primary SOC, HLGT, HLT, and PT, showing number (%) of patients with at least 1 TEAE sorted by the SOC internationally agreed order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order.
- All TEAEs by primary SOC and PT, showing the number (%) of patients with at least 1 TEAE, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC. This sorting order will be applied to all other tables, unless otherwise specified.
- All TEAEs regardless of relationship and related to IMP by primary SOC, HLGT, HLT and PT, showing the number (%) of patients with at least 1 TEAE, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order.
- All TEAEs by maximal severity, presented by primary SOC and PT, showing the number (%) of patients with at least 1 TEAE by maximal severity (ie, mild, moderate, or severe), sorted by the sorting order defined above.

Analysis of all treatment emergent SAEs

- All treatment-emergent SAEs by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients with at least 1 serious TEAE, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order.
- All treatment-emergent SAEs by primary SOC and PT, showing the number (%) of patients with at least 1 serious TEAE, sorted by the sorting order defined above.
- All treatment-emergent SAEs regardless of relationship and related to IMP, by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients with at least 1 treatment-emergent SAE, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order.
- Listings will be provided for all SAEs by treatment arm during the TEAE period.

Analysis of all TEAEs leading to treatment discontinuation

- All TEAEs leading to treatment discontinuation, by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order.
- All TEAEs leading to treatment discontinuation, by primary SOC and PT, showing the number (%) of patients with at least 1 TEAE, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC.
- Listing will be provided for all TEAE leading to permanent treatment discontinuation by treatment arm.

Analysis of treatment-emergent AESIs

All treatment-emergent AESIs, by AESI category and PT, showing the number (%) of patients by treatment group, sorted by decreasing incidence of PT within each AESI category. The AESIs categories and details of the MedDRA coding are provided in Section 2.1.5.1.

- Listing of all Treatment emergent AESIs experienced per patient and by treatment arm: Patient ID, AESI category, PT, time to onset of event, date of first and last IMP administration, duration, intensity, corrective treatment, seriousness, relationship to SAR.

Analysis of pre-treatment and post-treatment adverse events

If any pre-treatment or post-treatment AE lead to death or any pre-treatment AE lead to treatment discontinuation, a listing will be provided.

2.4.5.2 Deaths

The following summaries of deaths will be generated for the safety population by treatment group.

- Number (%) of patients who died by study period (on-study, on-treatment, post-study).
- Deaths in non-randomized patients or randomized but not treated patients.
- TEAEs leading to death (death as an outcome on the adverse event case report form page as reported by the Investigator) by primary SOC, HLGT, HLT, and PT showing number (%) of patients sorted by internationally agreed SOC order, with HLGT, HLT, and PT presented in alphabetical order within each SOC.
- Listings of deaths will be provided.

2.4.5.3 Analyses of laboratory variables

The summary statistics (including number, mean, median, standard deviation, minimum and maximum) of all laboratory variables (central laboratory values and changes from baseline) will be calculated for each visit (baseline and post-baseline time points) and presented by treatment group. This section will be organized by biological function as specified in Section 2.1.5.3.

For parameters creatinine, glucose, neutrophils, monocytes, basophils, eosinophils and bicarbonate, the last on-treatment and worst on-treatment value will also be provided.

[redacted]

The incidence of PCSAs (list provided in Appendix A) at any time during the TEAE period will be summarized by biological function and treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/missing.
- Abnormal according to PCSA criterion or criteria.

For parameters for which no PCSA criteria are defined, similar table(s) using the normal range will be provided.

Listing will be provided with flags indicating the out of range values as well as PCSA values.

Drug-induced liver injury

The liver function tests, namely AST, ALT, alkaline phosphatase, and total bilirubin, are used to assess possible drug-induced liver toxicity. The proportion of patients with PCSA values at any post-baseline visit by baseline status will be displayed by treatment group for each parameter.

A graph of distribution of peak values of ALT versus peak values of total bilirubin will also be presented. Note that the ALT and total bilirubin values are presented on a logarithmic scale. The graph will be divided into 4 quadrants with a vertical line corresponding to 3 x ULN for ALT and a horizontal line corresponding to 2 x ULN for total bilirubin.

The normalization (to ≤1 x ULN or return to baseline if baseline >ULN) of elevated liver function tests will be summarized for each parameter by categories of elevation (3 x, 5 x, 10 x, 20 x ULN for ALT and AST, 1.5 x ULN for alkaline phosphatase, and 1.5 x and 2 x ULN for total bilirubin), with the following categories of normalization: no worsening as compared to the baseline PCSA status, returned to baseline PCSA status before last IMP dose, returned to baseline PCSA status after last IMP dose, never returned to baseline PCSA status, no assessment after the worst on-treatment value. Note that a patient will be counted only under the maximum elevation category.

The incidence of liver-related adverse events will be summarized by treatment group. The selection of PTs will be based on the hepatic disorder SMQ.

2.4.5.4 Analyses of vital sign variables

The summary statistics (including number, mean, median, standard deviation, minimum and maximum) of all vital signs variables (values and changes from baseline) will be calculated for each visit (baseline and post-baseline time points) by treatment group. For parameters heart rate, systolic and diastolic blood pressures, the last on-treatment and worst on-treatment value will also be provided.

The incidence of PCSAs at any time during the TEAE period will be summarized by treatment arm irrespective of the baseline level and according to the following baseline status categories (only for SBP and DBP):

- Normal/missing.
- Abnormal according to PCSA criterion or criteria.

Listing will be provided with flags indicating the out of range values as well as PCSA values.

2.4.5.5 Analyses of electrocardiogram variables

The summary statistics (including number, mean, median, standard deviation, minimum and maximum) of all ECG variables (values and changes from baseline) will be calculated for each visit (baseline and post-baseline time points) by treatment group.

The incidence of PCSAs at any time during the TEAE period will be summarized by treatment group irrespective of the baseline level and according to the following baseline status categories:

- Normal/missing.
- Abnormal according to PCSA criterion or criteria.

Listing will be provided with flags indicating the out of range values as well as PCSA values.

2.4.6 Analyses of pharmacokinetic variables

C_{trough} concentrations of SAR156597 in plasma will be described on the PK population for each time point by treatment group using descriptive statistics (number of patients, arithmetic mean (SD), geometric mean, coefficient of variation, minimum and maximum).

In a separate table, $C_{trough,av,ss}$ and $C_{Follow\text{-}up}$ will be described on the PK population by treatment group using descriptive statistics (number of patients, arithmetic mean (SD), geometric mean, coefficient of variation, minimum and maximum).

The graphical representation of time profiles for C_{trough} concentration (mean ±SD or Median, as appropriate) will be also provided by treatment group. Average steady state achievement will be performed by visual inspection of mean plots.

According to study results, additional plots will be prepared, as deemed necessary (eg, to explore the relationship with some safety or efficacy endpoints of interest).

The results of population PK modeling will be reported separately from the study report.

2.4.7 Analyses of immunogenicity data

The following summaries will be performed on the ADA population, taking into account all samples regardless of timing in relation to injections:

- ADA results (negative, positive or inconclusive) by time point and by treatment arm.
- Neutralizing status (inconclusive or positive) for positive ADA by time point and by treatment arm.

- ADA titers using descriptive statistics (median, Q1, Q3, minimum and maximum) for positive ADA by time point and by treatment arm.
- Number (%) of patients with pre-existing ADA and number (%) of patients with treatment-emergent ADA positive response by treatment arm.
- Number (%) of patients with persistent/transient/indeterminate treatment-emergent ADA positive response by treatment arm.
- Time to onset of treatment-emergent ADA positive response (in days) using descriptive statistics by treatment arm and also according to peak titer.
- Duration of ADA (in days, only for subjects with at least 2 positive ADA samples) using descriptive statistics by treatment arm.

Correlations between ADA parameters (eg, titers, treatment-emergent ADA positive status, neutralizing status) and PK, safety and/or main efficacy endpoint(s) will be explored (eg, scatter plot) according to the study results.

2.4.8 Analyses of Biomarker data

The biomarkers will be summarized by treatment arm and time-point: number of values below Lower Limit of Quantification (LLOQ), mean, SD. The biomarkers may be log-transformed based on their distribution. A biomarker could be discarded from the analysis if too many values are missing or below LLOQ. In addition, the values below LLOQ will be imputed by LLOQ/2.

To assess the effect of SAR156597 on each biomarker, the change from baseline to Week 24 will be analyzed using a mixed model for repeated measures with treatment, visit and treatment-by-visit interaction as fixed effects, baseline biomarker value as fixed covariates, and assuming an unstructured covariance structure.

The biomarkers at baseline will be tested one-by-one for a potential prognostic and predictive effect for mRSS endpoint. The change from baseline of mRSS to Week 24 will be assessed using biomarker main effect (prognostic effect), treatment effect, randomization stratum (as per IRT) and interaction between biomarker and treatment (predictive effect). As a second step, the biomarkers will be dichotomized using median value.

2.4.9 Analyses of quality of life/health economics variables

See Section 2.4.4.4.6 and Section 2.4.4.4.7.

2.5 DATA HANDLING CONVENTIONS

2.5.1 General conventions

The following formulas will be used for computation of parameters.

Study day

Study day = assessment date - first IMP injection date + 1 (if assessment date ≥first IMP injection date), otherwise study day = assessment date - first IMP injection date, the day of first IMP injection being Day 1. For randomized but not treated patients, Day 1 is the day of randomization.

Demographic formulas

Age (years) = (date of informed consent - date of birth)/365.25

Renal function formulas

Creatinine clearance value will be derived using the equation of Cockroft and Gault:

For male:

$$CL_{CR}\ (mL/min) = \frac{(140 - age(years)) * weight(kg)}{0.814 * serumcreatinine(\mu mol / L)}$$

For Female: result above multiplied by 0.85.

Date of last dose of investigational medicinal product

The date of the last dose of IMP is equal to the last date of administration reported on the IMP administration case report form page, or missing if the last administration date is unknown.

Lipids variables, laboratory safety variables, hs-CRP

For data below the lower limit of quantification (LLOQ)/limit of linearity, half of the lower limit value (ie, LLOQ/2) will be used for quantitative analyses. For data above the upper limit of quantification (ULOQ) / limit of linearity, the upper limit value (ie, ULOQ) will be used for quantitative analyses.

For ANCA parameter: values "<1:20" are considered as negative; other results are considered as positive.

2.5.2 Data handling conventions for efficacy variables

Handling of CRISS missing data

Two different methods will be used to impute CRISS missing data, depending on the nature of the missing item.

If the missing item at a given time-point relates to organ involvement (Step 1 of CRISS derivation):

- If there was already an organ involvement at a previous post-baseline time-point (CRISS already evaluated as 0 at a previous time point), CRISS will be imputed the value of 0.
- Otherwise, if at the current time-point, one organ has worsened (CRISS evaluated as 0 due to new or worsening of another organ for which the evaluation is non missing), CRISS will be imputed the value of 0.
- Otherwise, if no other organ has worsened, as such an event should have been classified as a serious adverse event, it is assumed that no organ had worsened otherwise the event would have been recorded, so CRISS will be derived using Step 2. Each missing component (FVC, DLco, mRSS, PT-glob, MD-glob and HAQ-DI) at Week 24 will be imputed using the method explained below.

Handling of FVC, DLco and CRISS missing data

In order to impute the FVC, DLco and CRISS missing data at Week 24, the following steps will be performed:

- Patients in active arm and taking treatment at Week 24 (treatment discontinuation is defined for this analysis as last injection + 7 days): Missing values at Week 24 will be imputed using a model estimated from the subset of patients in the active arm who completed the 24-week treatment period and with a Week 24 value. The imputation model will be based on a MMRM with fixed categorical effect of stratification (as per IRT, with/without ILD) and repeated categorical effect for time point (baseline, Week 4 (except for % predicted DLco and % predicted FVC), Week 8 (except for % predicted DLco and % predicted FVC), Week 12 and Week 24). Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Kenward-Roger's approximation. Missing values at Week 24 will be imputed using the Empirical Best Linear Unbiased Prediction (EBLUP).
- Other patients (patients in active arm and not taking treatment at Week 24; or patients in placebo arm): Missing values at Week 24 will be imputed similarly but using a model estimated from the subset of patients in the placebo arm who completed the 24-week treatment period and with a Week 24 value.
- When running the imputation model:
  - If the mRSS value is inferior to 0, then it will be set to 0. If the mRSS value is superior to 51, it will be set to 51,
  - If the % predicted FVC value is inferior to 0, it will be set to 0,
  - If the % predicted DLco value is inferior to 0, it will be set to 0,
  - If the PT-glob value is inferior to 0 or superior to 10 it will be set to 0 and respectively to 10,
  - If the MD-glob value is inferior to 0 or superior to 10, it will be set to 0 and respectively to 10,
  - If the HAQ-DI value is inferior to 0 or superior to 3, it will be set to 0 and respectively to 3.

2.5.3 Missing data

For categorical variables, patients with missing data are not included in calculations of percentages unless otherwise specified. When relevant, the number of patients with missing data is presented.

Handling of computation of treatment duration if investigational medicinal product end of treatment date is missing

For the calculation of the treatment duration, the date of the last dose of IMP is equal to the date of last administration reported on the end-of-treatment case report form page. If this date is missing, the exposure duration should be left as missing.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of medication missing/partial dates

No imputation of medication start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior or concomitantly, it will be considered as prior, concomitant, and post-treatment medication.

Handling of adverse events with missing or partial date/time of onset

Missing or partial adverse event onset dates and times will be imputed so that if the partial adverse event onset date/time information does not indicate that the adverse event started prior to treatment or after the treatment-emergent adverse event period, the adverse event will be classified as treatment-emergent. No imputation of adverse event end dates/times will be performed. These data imputations are for categorization purpose only and will not be used in listings. No imputation is planned for date/time of adverse event resolution.

Handling of adverse events when date and time of first investigational medicinal product administration is missing

When the date and time of the first IMP administration is missing, all adverse events that occurred on or after the day of randomization should be considered as treatment-emergent adverse events. The exposure duration should be kept as missing.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of missing assessment of relationship of adverse events to investigational medicinal product

If the assessment of the relationship to IMP is missing, then the relationship to IMP has to be assumed and the adverse event considered as such in the frequency tables of possibly related adverse events, but no imputation should be done at the data level.

Handling of missing severity of adverse events

If the severity is missing for 1 of the treatment-emergent occurrences of an adverse event, the maximal severity on the remaining occurrences will be considered. If the severity is missing for all the occurrences, a "missing" category will be added in the summary table.

Handling of potentially clinically significant abnormalities

If a patient has a missing baseline he will be grouped in the category "normal/missing at baseline."

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing; eg, for eosinophils the PCSA is >0.5 GIGA/L or >ULN if ULN ≥0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

2.5.4 Windows for time points

Data analyzed by time point (including efficacy, safety, PK, BM and ADA) will be summarized using the analysis windows below. The analysis windows will be exhaustive so that data recorded at any time point (including unscheduled visits) have the potential to be summarized. These analysis windows will be applicable for all analyses, and they are defined to provide more homogeneous data for time point-specific analyses.

- For efficacy, if multiple valid values of a variable exist within an analysis window, the nearest from the targeted study day will be selected. In case of ties, the worst value will be displayed.
- For PK, BM and ADA, if multiple valid values of a variable exist within an analysis window, the value before the injection will be selected.
- For safety, if multiple valid values of a variable exist within an analysis window, the nearest from the targeted study day will be selected. In case of ties, the last value will be displayed (except for ANCA).
- For ANCA and ANA, if multiple valid values of a variable exist within an analysis window, the nearest from the targeted study day will be selected. In case of ties, the positive value with the highest titer will be displayed.

Table 3 - Analysis windows definition for efficacy and safety

Time point	Targeted study day	Analysis window in study days for efficacy (except CRISS) and safety	Analysis window in study days for CRISS
Baseline	1	Last value before first IMP*	
Week 2	15	2 to 21**	
Week 4	29	2 to 42***	
Week 8	57	43 to 70***	
Week 12	85	71 to 126	71 to 126
Week 24 (EOT)	169	127 to 207	127 to 207
Week 35 (EOS)	246	>207	>207

Note: Study days are calculated from the day of first IMP injection, the day of first IMP injection being Day 1. For randomized but not treated patients, Day 1 is the day of randomization.

*except for ANCA.

**only for vital signs.

***except for FVC, DLco, EQ-5D-5L.

The BM, PK and ADA will be analyzed according to the date of corresponded injection. A time windows superior or inferior to 2 days will be accepted. If the date of the previous injection is unknown, the SAR156597 concentration will not be considered for the analysis.

2.5.5 Unscheduled visits

For all analyses, unscheduled visit measurements may be used to provide a measurement for a time point.

2.5.6 Pooling of centers for statistical analyses

Not applicable.

2.5.7 Statistical technical issues

Not applicable.

3 INTERIM ANALYSIS

As there is no major enrollment issue, no futility analysis is planned.

Early analysis:

The primary analysis of efficacy and safety will be performed on the data collected during the 24-week treatment period. The results of this analysis will not be used to change the conduct of the follow-up in any aspect.

The analysis will be conducted in two steps:

- First step: Main efficacy and safety analyses.
- The first analysis will be conducted when all patients have been randomized and have at least all their data up to Week 24 collected and validated, and will consist in the final analysis of the efficacy endpoints up to Week 24. The safety analysis will be performed on all safety data collected and validated at the time of first analysis. The results of the first analysis will not be used to change the conduct of the ongoing study in any aspect. The 24-week data will not be reanalyzed at the end of the study. Individuals who have access to patients' source documents will remain blinded with regards to the treatment arm of individual patients throughout the study. Treatment arm will not be disclosed on Data Surveillance reports.
- Second step: Final analysis.
- The second analysis will be conducted at the end of the study and will consist in the final analysis of Week 35 efficacy endpoints and final safety analysis.

4 DATABASE LOCK

The database is planned to be locked 28 days after last patient last visit.

There are two database locks planned for this study as follows:

- First database lock (for early analysis of the main efficacy and safety analyses evaluation at Week 24): 4 weeks after last patient last efficacy visit (Visit 7 Week 24).
- Final database lock (for final analysis of efficacy endpoints and final safety analysis at Week 35): 4 weeks after last patient last visit (Visit 8 Week 35).

5 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS Version 9.0 or higher.

6 REFERENCES

1. Sampson HA, Munoz-Furlong A, Campbell RL, Adkinson NF Jr, Bock SA, Branum A, et al. Second symposium on the definition and management of anaphylaxis: Summary report Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. J Allergy Clin Immunol. 2006;117:391-7.

2. Khanna D, Furst DE, Hays RD, Park GS, Wong WK, Seibold JR, et al. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. Ann Rheum Dis. 2006;65:1325-9.

3. Amjadi S, Maranian P, Furst DE, Clements PJ, Wong WK, Postlethuaite AE, et al. Course of Modified Rodnan Skin Score in systemic sclerosis clinical trials: analysis of three large multicenter, randomized clinical trials. Arthritis Rheum. 2009;60(8):2490-8.

4. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095-10

5. Van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. Value Health 2012;15(5):708-15.

6. Ratitch B and O'Kelly M. Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures. PharmaSUG2011 - Paper SP04.

7. Kenward G. Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. Clin Invest. (Lond) 2015;5(3):311-20.

7 LIST OF APPENDICES

Appendix A Potentially clinically significant abnormalities (PCSA) criteria

Appendix B Summary of statistical analyses

Appendix C mRSS

Appendix D SHAQ

Appendix E UCLA SCTC GIT 2.0 Questionnaire

Appendix F EQ-5D-5L Questionnaire (UK version)

Appendix G Patient and physician global assessments of overall health using the 0-10 Likert scale

Appendix H TJC28

Appendix I Digital ulcer count

Appendix J Codelists for background therapy definitions, AESIs and related hepatic disorder

Appendix A Potentially clinically significant abnormalities criteria

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for Phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
Clinical Chemistry		
ALT	By distribution analysis: >3 ULN >5 ULN >10 ULN >20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI - FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. Categories are cumulative. First row is mandatory. Rows following one mentioning zero can be deleted.
AST	By distribution analysis: >3 ULN >5 ULN >10 ULN >20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI - FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. Categories are cumulative. First row is mandatory. Rows following one mentioning zero can be deleted.
Alkaline Phosphatase	>1.5 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI - FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008.
Total Bilirubin	>1.5 ULN >2 ULN	Must be expressed in ULN, not in µmol/L or mg/L. Categories are cumulative. Concept paper on DILI - FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008.
Conjugated Bilirubin	>35% Total Bilirubin and TBILI>1.5 ULN	Conjugated bilirubin dosed on a case-by-case basis.
ALT and Total Bilirubin	ALT >3 ULN and TBILI >2 ULN	Concept paper on DILI - FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. To be counted within a same treatment phase, whatever the interval between measurement.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for Phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
CPK	>3 ULN	FDA Feb 2005.
	>10 ULN	Am J Cardiol April 2006.
		Categories are cumulative.
		First row is mandatory. Rows following one mentioning zero can be deleted.
CLcr (mL/min)	<15 (end stage renal disease)	FDA draft Guidance 2010.
(Estimated creatinine clearance based on the Cokcroft-Gault equation)	≥15 - <30 (severe decrease in GFR)	Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling.
	≥30 - <60 (moderate decrease in GFR)	
	≥60 - <90 (mild decrease in GFR)	
	≥90 (normal GFR)	
eGFR (mL/min/1.73m^2)	<15 (end stage renal disease)	FDA draft Guidance 2010.
(Estimate of GFR based on an MDRD equation)	≥15 - <30 (severe decrease in GFR)	Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling.
	≥30 - <60 (moderate decrease in GFR)	
	≥60 - <90 (mild decrease in GFR)	
	≥90 (normal GFR)	
Creatinine	≥150 µmol/L (Adults)	Benichou C., 1994.
	≥30% change from baseline	
	≥100% change from baseline	
Uric Acid		Harrison- Principles of internal Medicine 17th Ed., 2008.
Hyperuricemia	>408 µmol/L	
Hypouricemia	<120 µmol/L	
Blood Urea Nitrogen	≥17 mmol/L	
Chloride	<80 mmol/L	
	>115 mmol/L	
Sodium	≤129 mmol/L	
	≥160 mmol/L	
Potassium	<3 mmol/L	FDA Feb 2005.
	≥5.5 mmol/L	
Total Cholesterol	≥7.74 mmol/L	Threshold for therapeutic intervention.
Triglycerides	≥4.6 mmol/L	Threshold for therapeutic intervention.
Lipasemia	≥3 ULN	

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for Phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
Amylasemia	≥3 ULN	
Glucose		
Hypoglycaemia	≤3.9 mmol/L and <LLN	ADA May 2005.
Hyperglycaemia	≥11.1 mmol/L (unfasted); ≥7 mmol/L (fasted)	ADA Jan 2008.
HbA1c	>8%	
Albumin	≤25 g/L	
CRP	>2 ULN or >10 mg/L (if ULN not provided)	FDA Sept 2005.
Hematology		
WBC	<3.0 Giga/L (Non-Black); <2.0 Giga/L (Black)	Increase in WBC: not relevant.
	≥16.0 Giga/L	To be interpreted only if no differential count available.
Lymphocytes	>4.0 Giga/L	
Neutrophils	<1.5 Giga/L (Non-Black);<1.0 Giga/L (Black)	International Consensus meeting on drug-induced blood cytopenias, 1991. FDA criteria.
Monocytes	>0.7 Giga/L	
Basophils	>0.1 Giga/L	
Eosinophils	>0.5 Giga/L or >ULN (if ULN≥0.5 Giga/L)	Harrison- Principles of internal Medicine 17th Ed., 2008.
Hemoglobin	≤115 g/L (Male); ≤95 g/L (Female) ≥185 g/L (Male); ≥165 g/L (Female) Decrease from Baseline ≥20 g/L	Criteria based upon decrease from baseline are more relevant than based on absolute value. Other categories for decrease from baseline can be used (≥30 g/L, ≥40 g/L, ≥50 g/L).
Hematocrit	≤0.37 v/v (Male) ; ≤0.32 v/v (Female) ≥0.55 v/v (Male) ; ≥0.5 v/v (Female)	
RBC	≥6 Tera/L	Unless specifically required for particular drug development, the analysis is redundant with that of Hb. Otherwise, consider FDA criteria.
Platelets	<100 Giga/L ≥700 Giga/L	International Consensus meeting on drug-induced blood cytopenias, 1991.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for Phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
Urinalysis		
pH	≤4.6	
	≥8	
Vital signs		
HR	≤50 bpm and decrease from baseline ≥20 bpm	To be applied for all positions (including missing) except STANDING.
	≥120 bpm and increase from baseline≥20 bpm	
SBP	≤95 mmHg and decrease from baseline ≥20mmHg	To be applied for all positions (including missing) except STANDING.
	≥160 mmHg and increase from baseline ≥20 mmHg	
DBP	≤45 mmHg and decrease from baseline ≥10 mmHg	To be applied for all positions (including missing) except STANDING.
	≥110 mmHg and increase from baseline ≥10 mmHg	
Orthostatic Hypotension		
Orthostatic SDB	≤-20 mmHg	
Orthostatic DBP	≤-10 mmHg	
Weight	≥5% increase from baseline	FDA Feb 2007.
	≥5% decrease from baseline	
ECG		Ref.: ICH E14 guidance (2005) and E14 Q&A (2012), and Cardiac Safety Research Consortium White Paper on PR and QRS (Nada et al. Am Heart J. 2013;165(4):489-500)
HR	<50 bpm	Categories are cumulative.
	<50 bpm and decrease from baseline ≥20 bpm	
	<40 bpm	
	<40 bpm and decrease from baseline ≥20 bpm	
	<30 bpm	
	<30 bpm and decrease from baseline ≥20 bpm	
	>90 bpm	Categories are cumulative.
	>90 bpm and increase from baseline ≥20bpm	
	>100 bpm	
	>100 bpm and increase from baseline ≥20bpm	
	>120 bpm	
	>120 bpm and increase from baseline ≥20 bpm	

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES

for Phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
PR	>200 ms	Categories are cumulative.
	>200 ms and increase from baseline ≥25%	
	>220 ms	
	>220 ms and increase from baseline ≥25%	
	>240 ms	
	>240 ms and increase from baseline ≥25%	
QRS	>110 ms	Categories are cumulative.
	>110 msec and increase from baseline ≥25%	
	>120 ms	
	>120 ms and increase from baseline ≥25%	
QT	>500 ms	
QTc	Absolute values (ms)	To be applied to any kind of QT correction formula. Absolute values categories are cumulative.
	>450 ms	
	>480 ms	QTc >480 ms and ΔQTc >60 ms are the 2 PCSA categories to be identified in individual subjects/patients listings.
	>500 ms	
	Increase from baseline	
	Increase from baseline]30-60] ms	
	Increase from baseline >60 ms	

Appendix B Summary of statistical analyses

EFFICACY ANALYSIS

Criteria	Analysis population	Primary analysis	Supportive analysis	Subgroup analysis	Other analyses
Primary endpoint					
mRSS	*ITT*	*Change from baseline to W24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, mRSS at baseline, mRSS at baseline by time point interaction*	*Analysis on safety population including only post-baseline data measured during the on-treatment period* *Analysis replacing the planned stratum with the actual stratum (if more than 15% of patients randomized to the wrong stratum)* *Analysis on patients with same rater* *Control-based pattern multiple imputation*	*Randomization stratum (as per IRT)* *Gender (male/female)* *Background therapy (yes/no)*	*Change from baseline to Week 35 (exploratory endpoint): same MMRM model as primary analysis*
Secondary endpoints					
HAQ-DI	*ITT*	*Change from baseline to Week 24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, HAQ-DI at baseline, HAQ-DI at baseline by time point interaction*	*Analysis replacing the planned stratum with the actual stratum (if more than 15% of patients randomized to the wrong stratum)*	*Randomization stratum (as per IRT)*	*Change from baseline to Week 35 (exploratory endpoint): same MMRM model as primary analysis* *Cumulative distribution function at Week 24 and Week 35*

Criteria	Analysis population	Primary analysis	Supportive analysis	Subgroup analysis	Other analyses
Observed FVC	*ITT*	*Change from baseline to Week 24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, FVC at baseline, FVC at baseline by time point interaction*	*Analysis replacing the planned stratum with the actual stratum (if more than 15% of patients randomized to the wrong stratum)* *Linear mixed effects model with random intercept and random slope with following covariates: treatment group, randomized srata, time, randomization strata by time interaction, treatment by time interaction, FVC at baseline, FVC at baseline by time interaction*	*Randomization stratum (as per IRT)*	*Change from baseline to Week 35 (exploratory endpoint): same MMRM model as primary analysis*
Observed DLco	*ITT*	*Change from baseline to Week 24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, Dlco at baseline, Dlco at baseline by time point interaction*	*Analysis replacing the planned stratum with the actual stratum (if more than 15% of patients randomized to the wrong stratum)*	*Randomization stratum (as per IRT)*	*Change from baseline to Week 35 (exploratory endpoint): same MMRM model as primary analysis*
Exploratory endpoints					
mRSS: responder rates	*ITT*	*Logistic regression for improvement of at least 20%, 40% and 60% at Week 24 with following covariates: treatment group, randomization strata, mRSS at baseline*	*No*	*No*	*Same logistic regression for improvement of at least 20%, 40% and 60% at Week 35 as primary analysis* *Cumulative distribution function at W24 and W35* *Same logistic regression as primary analysis for 3 supplementary definitions for responder rate at Week 24*

Criteria	Analysis population	Primary analysis	Supportive analysis	Subgroup analysis	Other analyses
Each of the 6 VAS from SHAQ (pain, breathing, vascular, gastrointestinal, digital ulcers, global assessment)	*ITT*	*Change from baseline to W24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, VAS at baseline, VAS at baseline by time point interaction*		*No*	*Change from baseline to W35: same MMRM model as primary analysis* *Cumulative distribution function at Week 24 and W35*
UCLA SCTC GIT 2.0 score	*ITT*	*Change from baseline to W24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, UCLA SCTC GIT 2.0 score at baseline, UCLA SCTC GIT 2.0 score at baseline by time point interaction*	*No*	*No*	*Change from baseline to W35: same MMRM model as primary analysis*
TJC28	*ITT*	*Change from baseline to W24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, TJC28 at baseline, TJC28 at baseline by time point interaction*	*No*	*No*	*Change from baseline to W35: same MMRM model as primary analysis*
Digital ulcer count	*ITT*	*Change from baseline to W24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction,digital ulcer count at baseline, digital ulcer count at baseline by time point interaction*	*No*	*No*	*Change from baseline to W35: same MMRM model as primary analysis* *Descriptive summary by treatment group and by time point of the digital ulcer count, and its subcategories: number of active/new ulcers and number of indeterminate/healing ulcers*

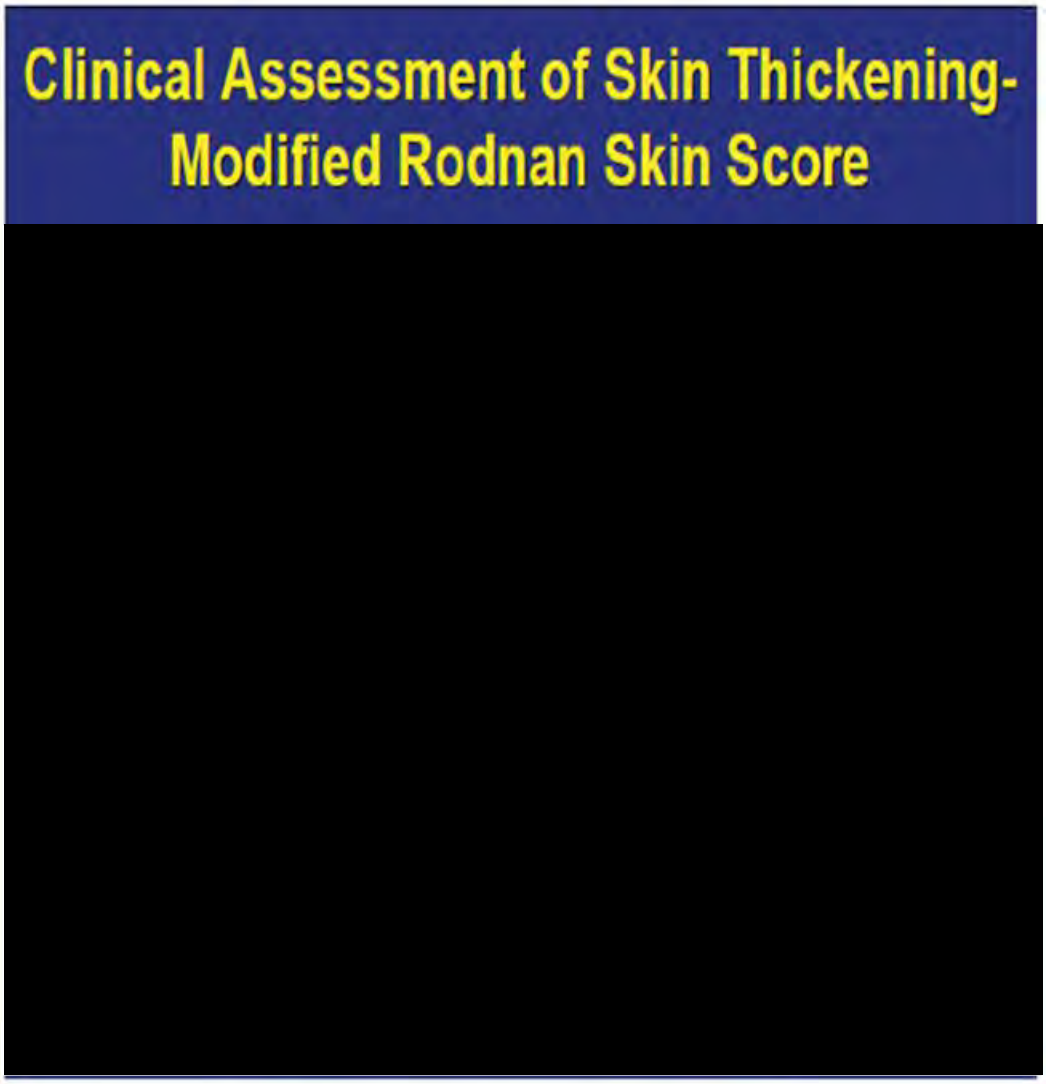
Criteria	Analysis population	Primary analysis	Supportive analysis	Subgroup analysis	Other analyses
EQ-5D-5L index and VAS	*ITT*	*Change from baseline to W24: MMRM with following covariates: treatment group, randomization strata, time point, randomization strata by time point interaction, treatment by time point interaction, EQ-5D-5L value at baseline, EQ-5D-5L value at baseline by time point interaction*	*No*	*No*	*Change from baseline to W35: same MMRM model as primary analysis* *Cumulative distribution function at Week 24 and Week 35*
% predicted FVC	*ITT*	*Change from baseline at W24: Rank ANCOVA adjusted for baseline and stratified on randomization stratum*	*No*	*No*	*Change from baseline at W35: same Rank ANCOVA as primary analysis* *Number (%) of patients with absolute decrease in % predicted FVC ≥5% or absolute increase in % predicted FVC ≥10% from baseline to W24*
% predicted DLco	*ITT*	*Change from baseline to W24: Rank ANCOVA adjusted for baseline and stratified on randomization stratum*	*No*	*No*	*Change from baseline at W35: same Rank ANCOVA as primary analysis* *Number (%) of patients with absolute decrease in % predicted DLco ≥5% or absolute increase in % predicted DLco ≥10% from baseline to W24*
CRISS	*ITT*	*Van Elteren's test stratified on randomization strata at Week 24*	*No*	*No*	*Van Elteren's test stratified on randomization strata at Week 35* *Rate of patients with predicted probability of improving obtained using the CRISS at W24 and W35 is ≥60%*

mRSS: modified Rodnan Skin Score, FVC: forced vital capacity, MMRM: mixed-effect model with repeated measures, ITT: Intent-to-treat population, ANCOVA: Analysis of Covariance.

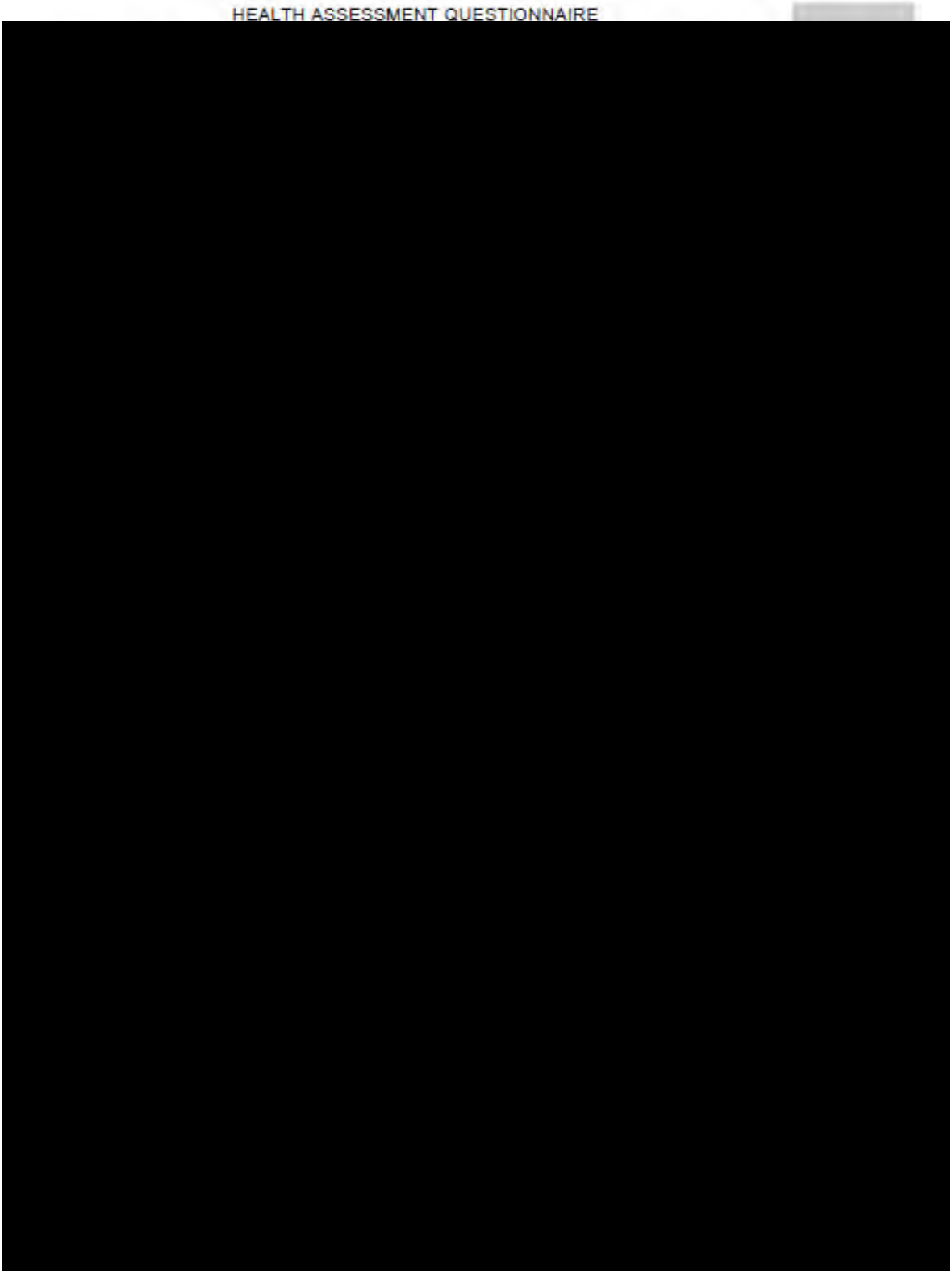
SAFETY ANALYSES

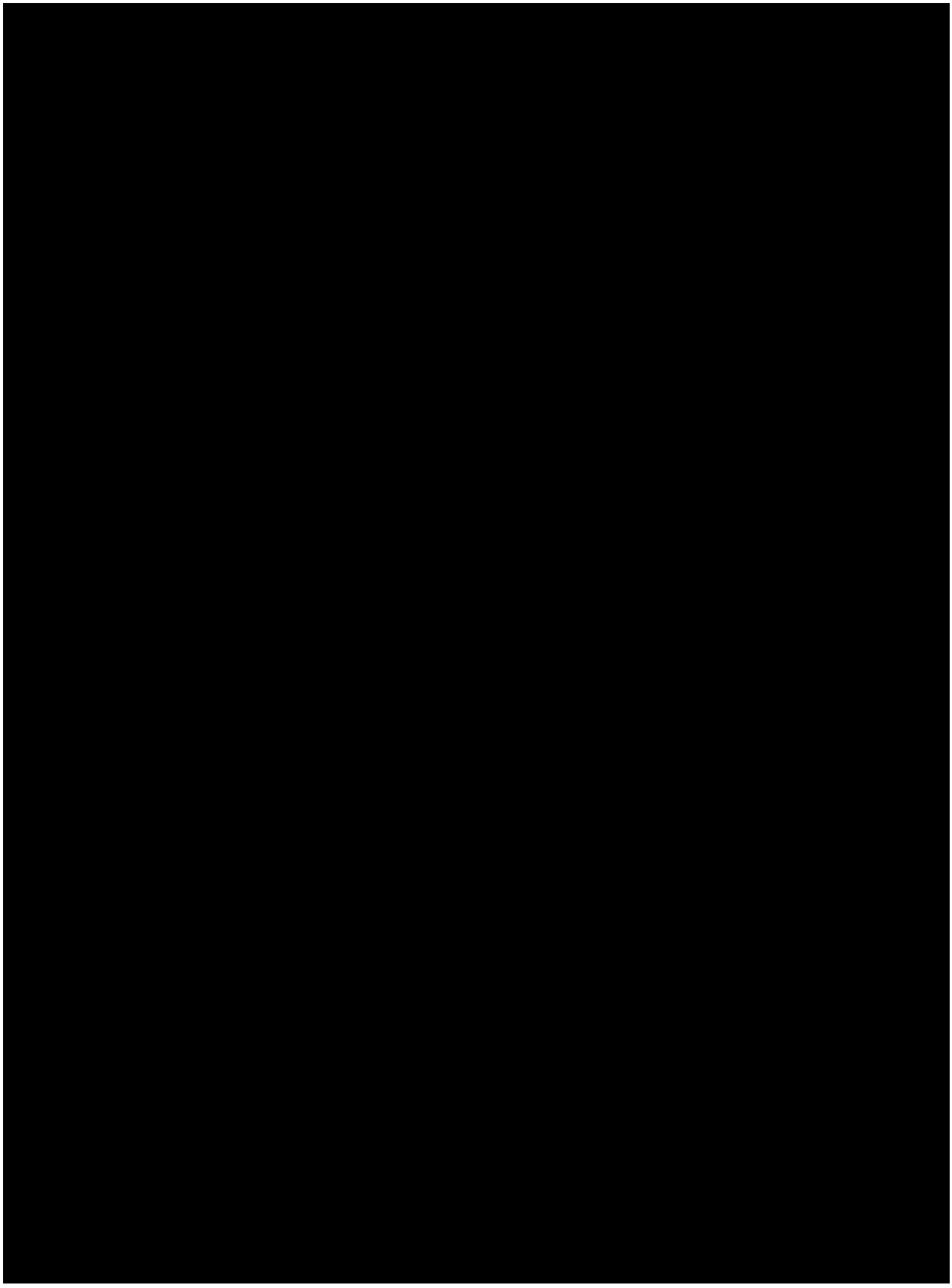
Endpoint	*Analysis population*	*Primary analysis*	*Supportive analysis*	*Subgroup analysis*	*Other analyses*
Adverse events	*Safety*	*Description of Adverse Events*	*No*	*No*	*No*
Vital signs	*Safety*	*Description of vital signs*	*No*	*No*	*No*
Laboratory variables	*Safety*	*Description of laboratory variables*			
ECG	*Safety*	*Description of ECG*			

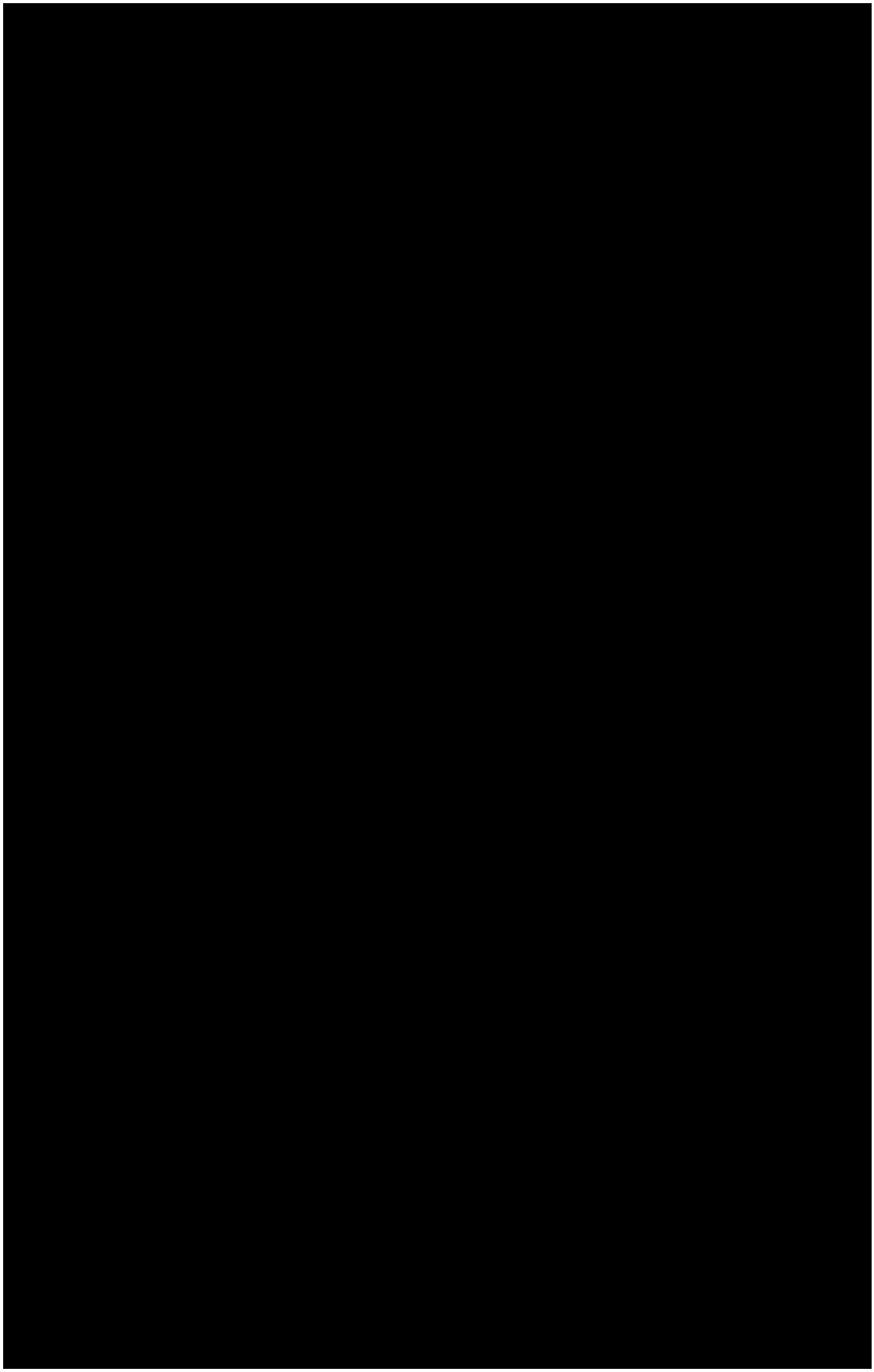
Appendix C mRSS



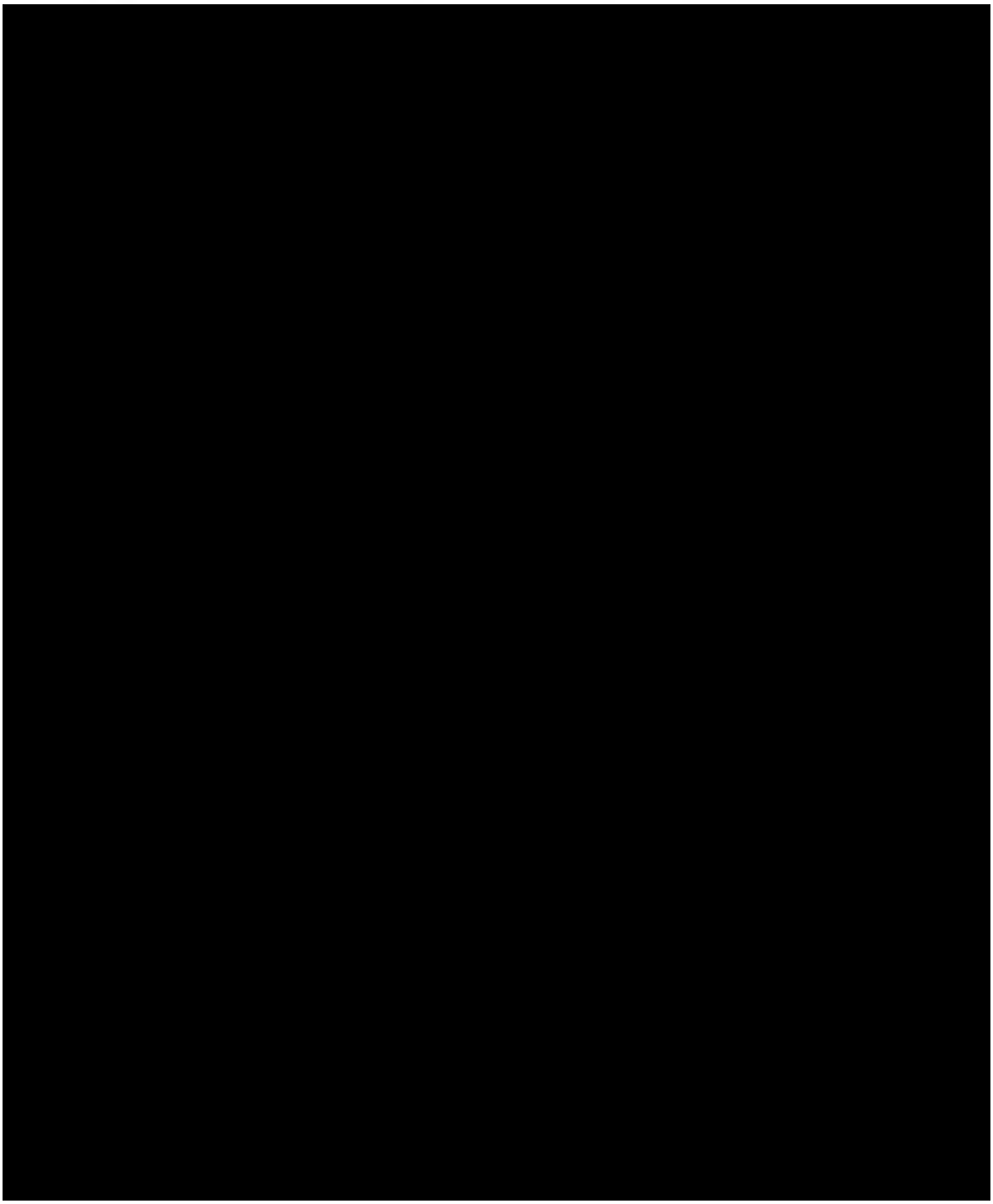
Appendix D SHAQ

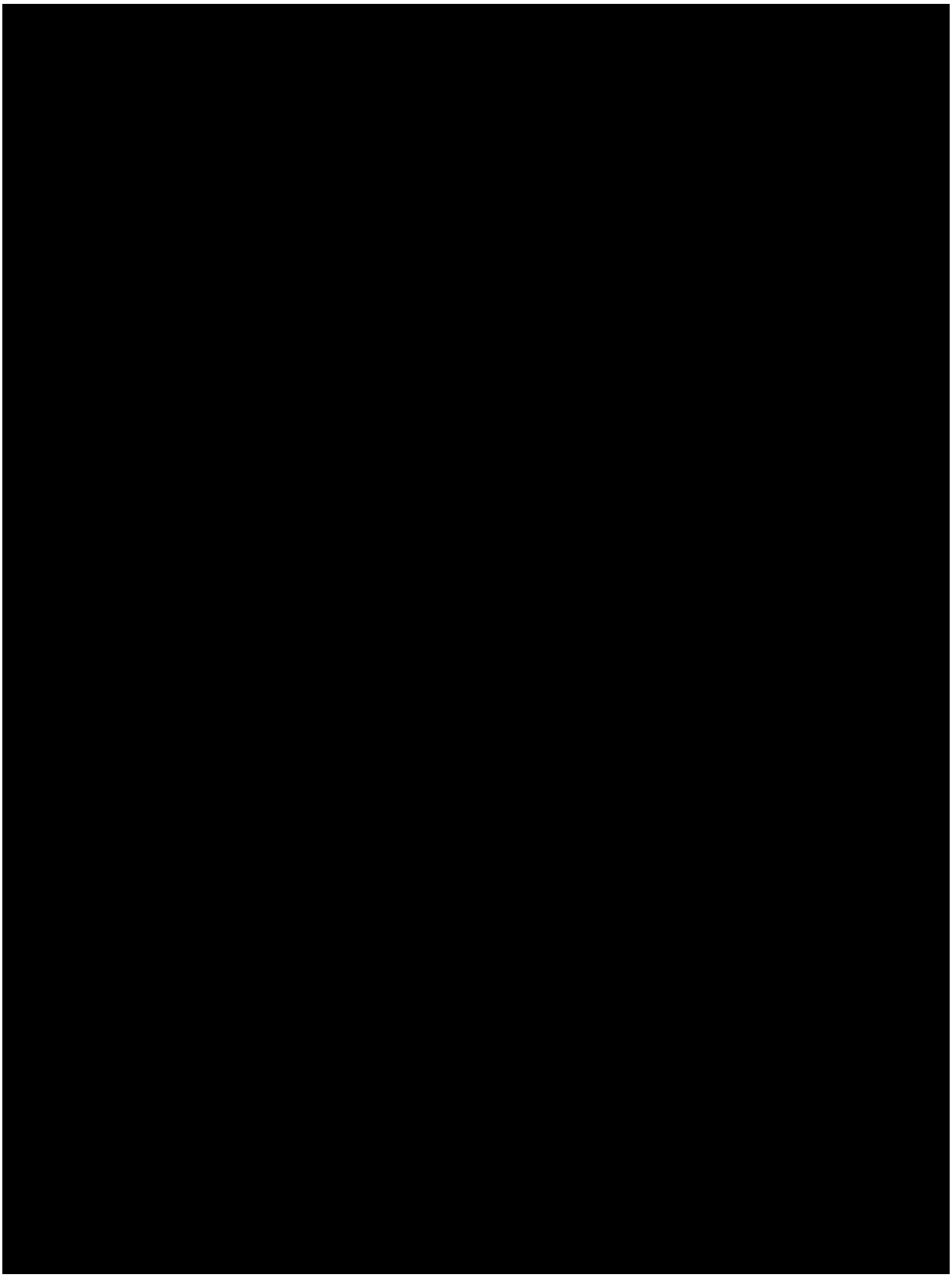


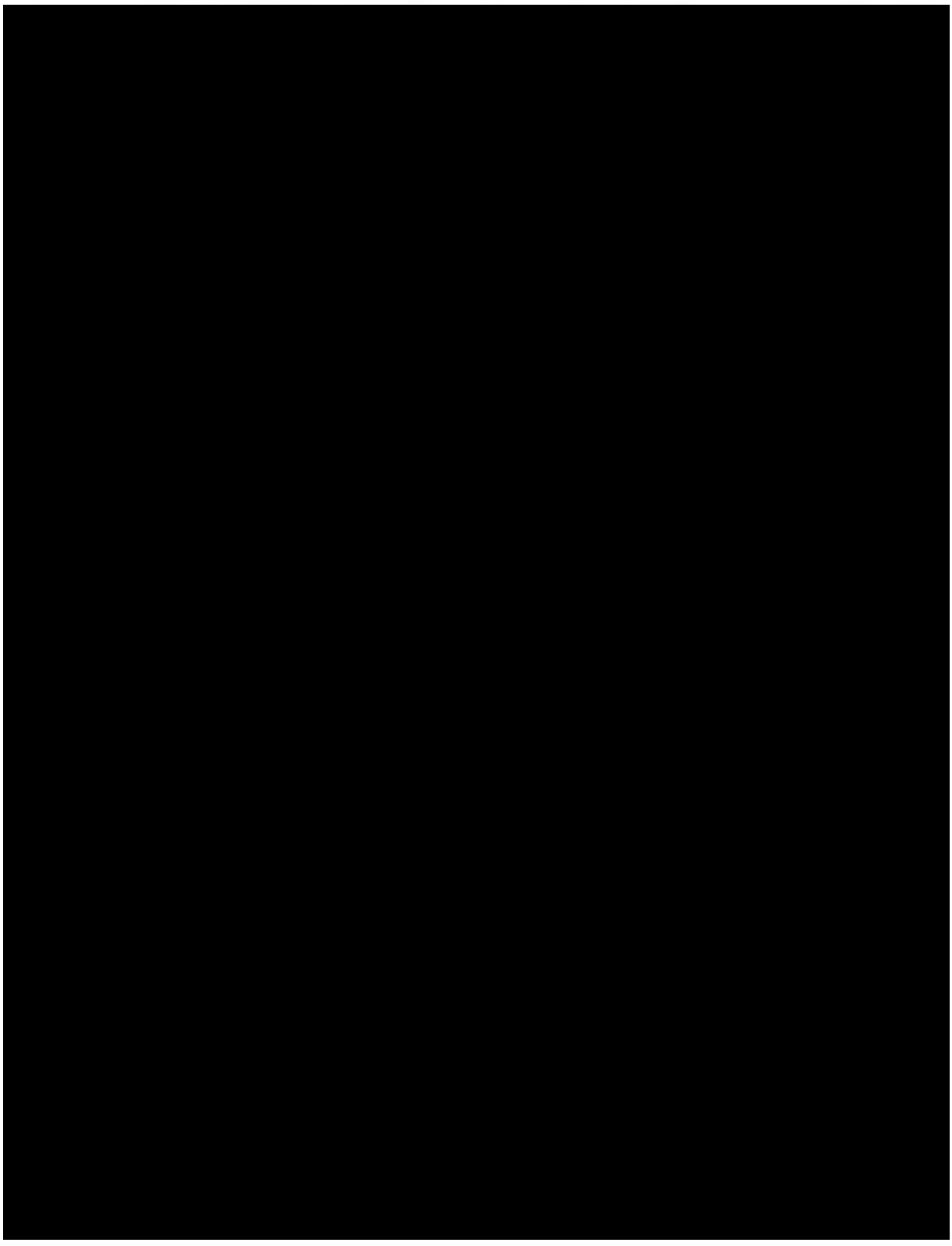


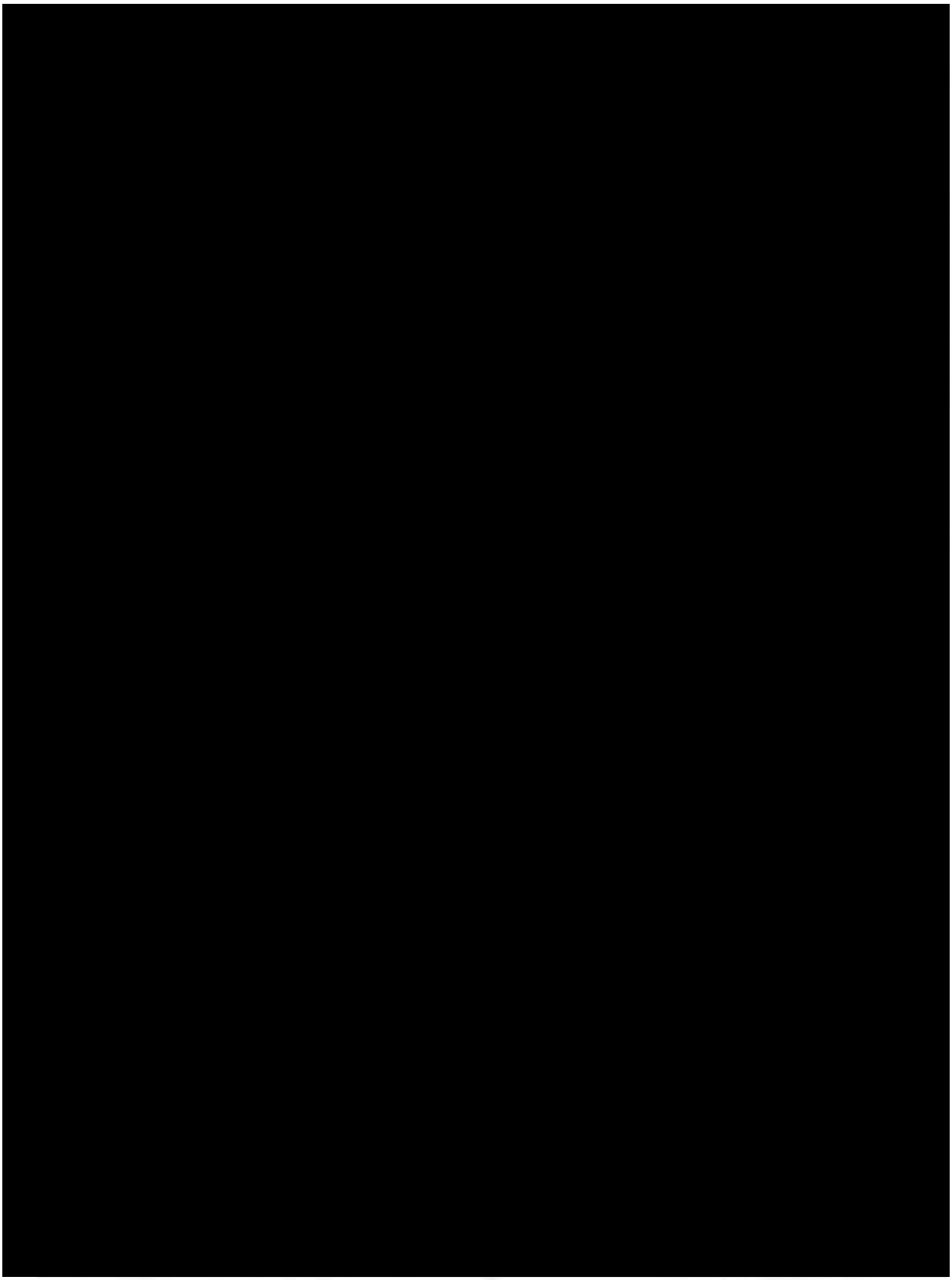


Appendix E UCLA SCTC GIT 2.0 Questionnaire

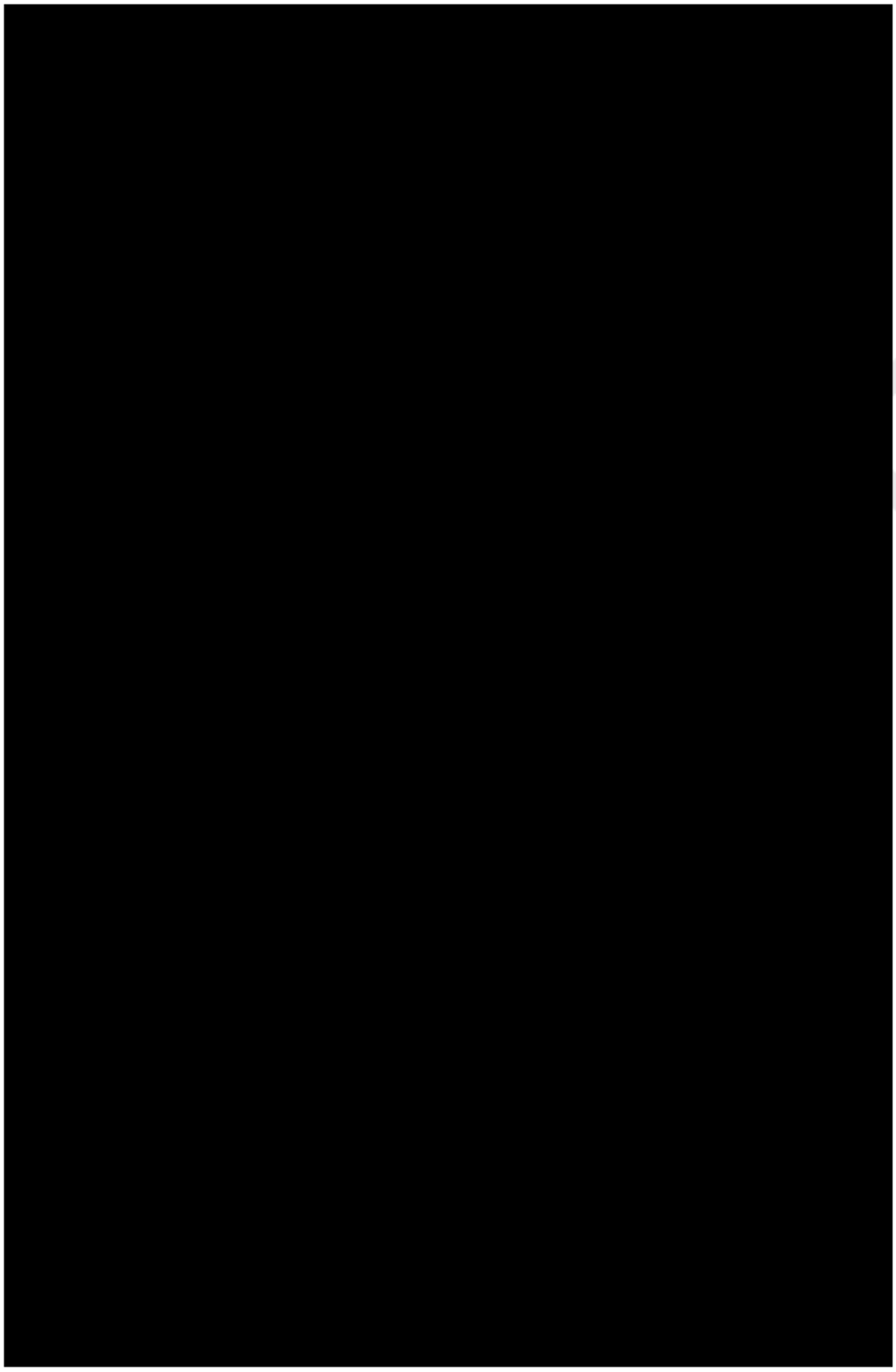


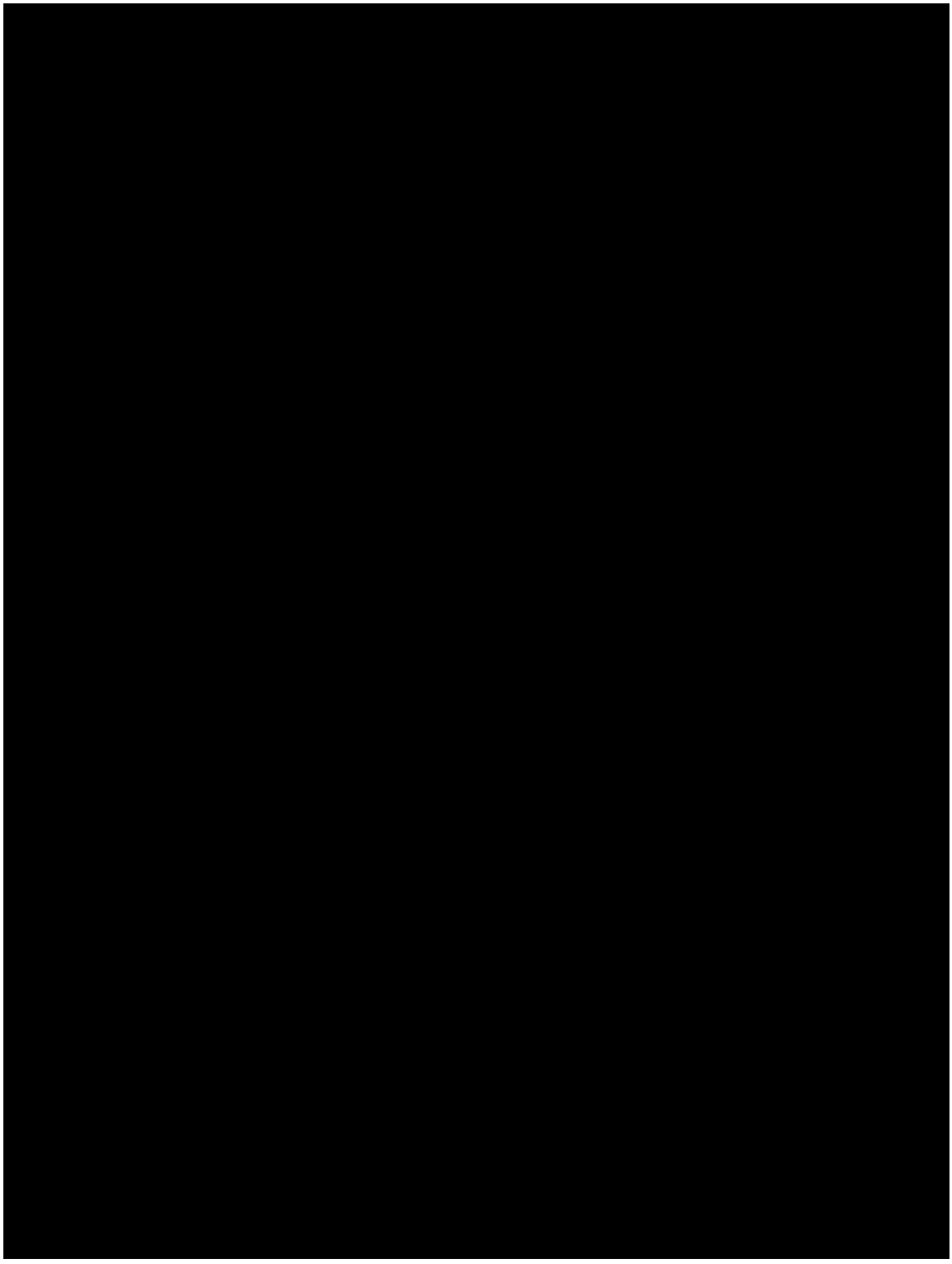






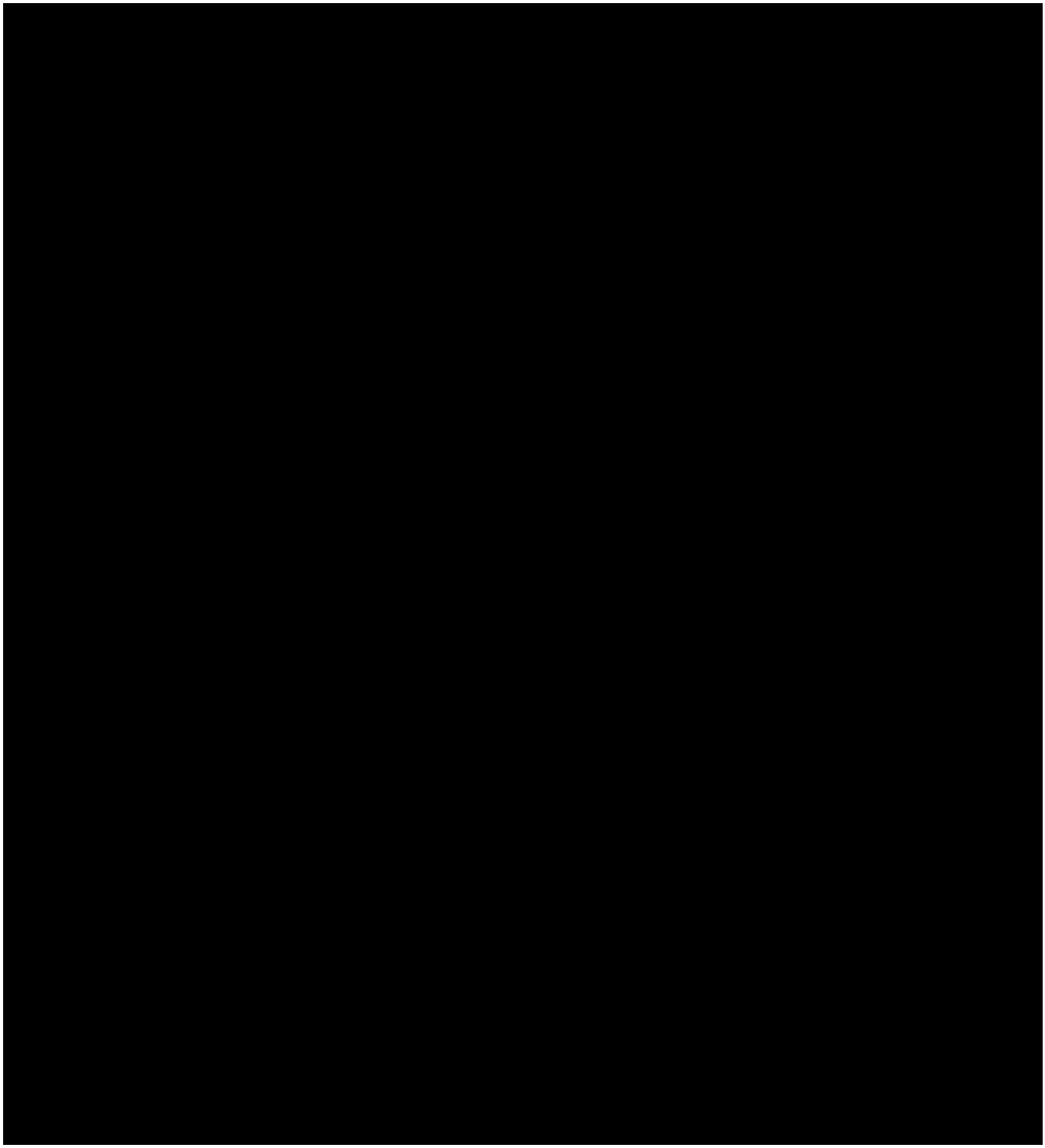
Appendix F EQ-5D-5L Questionnaire (UK version)





Appendix G Patient and physician global assessments of overall health using the 0-10 Likert scale

Appendix H TJC28



Appendix I Digital ulcer count

ACT14604/SAR156597
Digital Ulcer Count Source Document

Patient No. (12 digits): ☐☐☐ ☐☐☐☐ ☐☐☐☐☐ (Country No. Centre No. Subject No.)

Investigator Initials: _ _ _

Visit Name: V2/Baseline

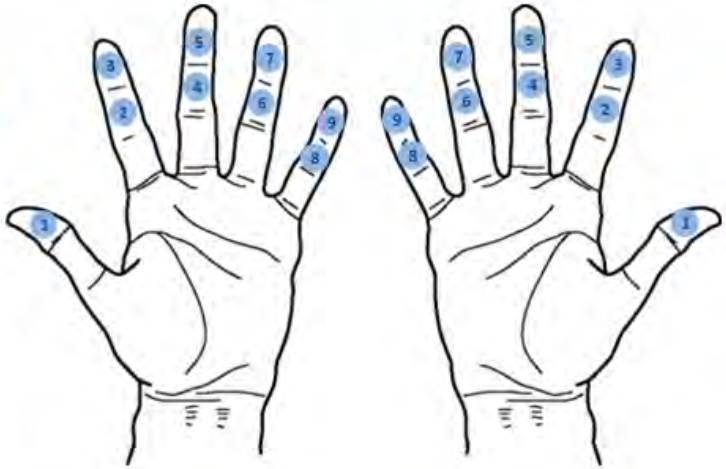
Visit Date: ☐☐/☐☐☐/☐☐☐☐ (DD MMM YYYY)

V2/BASELINE

LEFT HAND

Ulcer Location*	Ulcer Status	
	Active	Indeterminate / Healing
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
Total Number of Ulcers for CRF		

*Please enter the number corresponding to ulcer location



RIGHT HAND

Ulcer Location*	Ulcer Status	
	Active	Indeterminate / Healing
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
	☐	☐
Total Number of Ulcers for CRF		

*Please enter the number corresponding to ulcer location

For an ulcer that encompasses two adjacent locations (e.g., Right hand, locations 2 and 3), please select one of the two locations where the majority of the ulcer is present and record it under this location. (Such an ulcer should not be recorded separately in two different locations.) For all subsequent assessment of the same ulcer, please continue to record the ulcer using this same location.

ACT14604/SAR156597
Digital Ulcer Count Source Document

Patient No. (12 digits): ☐☐☐ ☐☐☐☐ ☐☐☐☐☐ (Country No. / Centre No. / Subject No.)

Investigator Initials: _ _ _

Visit Name: _____

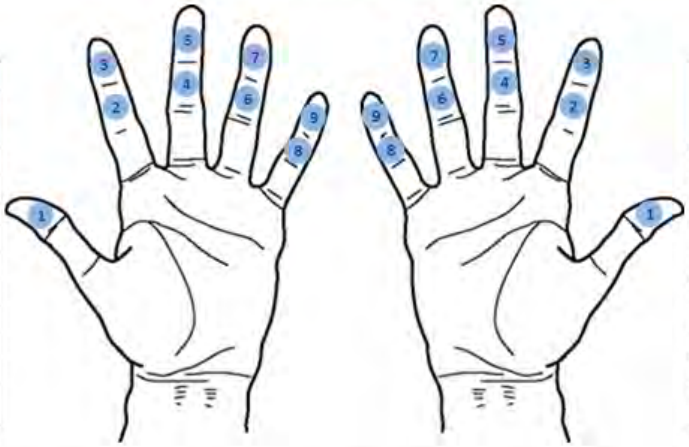
Visit Date: ☐☐/☐☐☐/☐☐☐☐ (DD / MMM / YYYY)

ALL VISITS EXCEPT V2/BASELINE

LEFT HAND

Ulcer Location*	Ulcer Status		
	Active(1)	New(2)	Indeterminate / Healing(3)
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
Total Number of Ulcers for CRF			

*Please enter the number corresponding to ulcer location



RIGHT HAND

Ulcer Location*	Ulcer Status		
	Active(1)	New(2)	Indeterminate / Healing(3)
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
	☐	☐	☐
Total Number of Ulcers for CRF			

*Please enter the number corresponding to ulcer location

For an ulcer that encompasses two adjacent locations (e.g., Right hand, locations 2 and 3), please select one of the two locations where the majority of the ulcer is present and record it under this location. (Such an ulcer should not be recorded separately in two different locations.) For all subsequent assessment of the same ulcer, please continue to record the ulcer using this same location.

1) **Active ulcers: already present at last visit**
2) **New ulcers: not present at last visit, regardless of their current status (active, indeterminate/healing or healed)**
3) **Indeterminate/Healing: already present at last visit**

Appendix J Codelists for background therapy definitions, AESIs and related hepatic disorder

Requested list	Type (cmq, smq, others)	Name or Code (cmqname, smqcd, others)	Description (cmqdesc or SMQ_ShortName, others)	Scope (smq) / prim_sec (soc)
Reqlist	**Type**	**Codename**	**Desc**	**Scope**
Background therapy definition 2	CDG	CDG00103	METHOTREXATE_mono_and_multiingredients	
Background therapy definition 2	CDG	CDG20046	MYCOPHENOLATE_MOFETIL_mono_and_multi_ingredients	
Background therapy definition 2	CDG	CDG00039	AZATHIOPRINE_mono_and_multi_ingredients	
Background therapy definition 2	CDG	CDG20011	CICLOSPORIN_mono_and_multi_ingredients	
Background therapy definition 2	CDG	CDG00040	CYCLOPHOSPHAMIDE_mono_and_multi_ingredients	
Background therapy definition 1	CDG	CDG00014	ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS	
Background therapy definition 1	CDG	CDG10109	SYSTEMIC HORMONAL PREPARATIONS, EXCL. SEX HORMONES AND INSULINS	
Acute renal failure	SMQ	Acute renal failure (SMQ)	Acute renal failure Narrow	Narrow
Tuberculosis	CMQ	GLB_TUBERCULOSIS	Tuberculosis	
Anaphylactic reactions or acute allergic reactions that require immediate treatment	CMQ	Anaphylactic reaction Narrow	Anaphylactic reaction Narrow	
[redacted]				
Hepatic disorder	SMQ	Hepatic disorders (SMQ)	Hepatic disorders Narrow	Narrow

Signature Page for VV-CLIN-0196174 v2.0
act14604-16-1-9-sap

Approve & eSign	[redacted] 07-Mar-2019 16:13:03 GMT+0000

Approve & eSign	[redacted] 07-Mar-2019 16:21:52 GMT+0000