ASSET Statistical Analysis Plan

TRIAL FULL TITLE	A phase II study to evaluate subcutaneous abatacept vs. placebo in diffuse cutaneous systemic sclerosis—a double-blind, placebo-controlled, randomized controlled trial
SAP VERSION	1.0
SAP VERSION DATE	06APR2018
TRIAL STATISTICIAN	Cathie Spino, DSc
Protocol Version (SAP associated with)	Protocol Version 4.0 December 8, 2015
TRIAL PRINCIPAL INVESTIGATOR	Dinesh Khanna, MD, MS
SAP AUTHOR(s)	Cathie Spino, DSc

1   SAP Signatures

I give my approval for this SAP entitled "ASSET Statistical Analysis Plan," dated 06APR2018.

Statistician (Author)
Name: Cathie Spino, ScD

Signature:
Date: 06APR2018

Statistician Reviewer (As applicable)
Name: Robert Parker, ScD

Signature:

Date: 07 April 2018

Principal Investigator
Name: Dinesh Khanna, MD, MS

Signature:

Date: 06APR2018

2   Table of Contents

3   Abbreviations and Definitions

AE	Adverse Event
ALT	Alanine Transaminase
AST	Aspartate Transaminase
CRF	Case Report Form
CMH	Cochran-Mantel-Haenszel test
CRISS	Combined Response Index in Systemic Sclerosis
DCC	Data Coordinating Center
dcSSc	Diffuse cutaneous systemic sclerosis
FDA	Food and Drug Administration
FVC	Percent Predicted forced vital capacity
GGT	Gamma-Glutamyltransferase
HAQ	Health Assessment Questionnaire
HAQ-DI	Health Assessment Questionnaire – Disability Index
HIV	Human Immunodeficiency Virus
HRCT	High resolution computer tomography
HRQOL	Health-related quality of life
ILD	Interstitial lung disease
IU	International Unit
IVIG	Intravenous Immunoglobulin
mITT	Modified intention to treat analysis population
mRSS	Modified Rodnan Skin Score
PROMIS	Patient-Reported Outcomes Measurement Information System
PROs	measure patient-reported outcomes
PRO-SRSS	PRO for Scleroderma-related Skin Symptoms
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SHAQ-DI	Scleroderma health assessment questionnaire-disability index
SSc	Systemic sclerosis
UCLA SCTC GIT	UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Instrument
VAS	Visual analogue scale

4   Introduction

4.1   Preface

Systemic sclerosis (Scleroderma, SSc) is one of the most fatal rheumatic diseases, and is associated with substantial morbidity[1] and many detrimental effects on health-related quality of life. Recent years have seen a revolution in the development and validation of outcome measures[2-4] and refinement of trial methodology in SSc.[5] This is paralleled by an increased understanding of the pathogenesis of SSc[6, 7] and development of targeted therapies.[8, 9] Modified Rodnan Skin Score (mRSS), a measure of skin thickness,[10, 11] has been used as the primary outcome measure in clinical trials of diffuse cutaneous SSc (called diffuse SSc or dcSSc). The Food and Drug Administration considers improvement in mRSS as an approvable end point.

Several observations support the role of activated T cells in the pathogenesis of SSc. Skin biopsies obtained from SSc patients early in their disease demonstrate a perivascular, mononuclear cell infiltrate comprised of T cells and macrophage. T cell activation is a prominent feature in SSc, as demonstrated by the presence of increased numbers of T cells bearing activation markers. T and B lymphocyte interactions are important in the pathogenesis of SSc, and T cells have been shown to be

essential for the production of autoantibodies in this disease. Finally, treatments directed against activated T cells, such as cyclosporine A, or depletion of T cells have resulted in skin softening in patients with SSc.

A recent pilot study which evaluated the effect of blockade of T cell co-stimulation with intravenous abatacept in patients with dcSSc showed efficacy and not serious or unanticipated adverse events. Therefore, we aim to perform a phase II, multi-center double-blind randomized controlled trial of subcutaneous abatacept vs. placebo in patients with early dcSSc.

4.2 Scope of the analyses

These analyses will assess the efficacy and safety of abatacept in comparison with placebo during the 12-month double-blind period, addressing the primary, secondary and exploratory objectives of study through the double-blind period. The ancillary objective of validating the new PRO for Scleroderma-related Skin Symptoms (PRO-SRSS), correlative studies (e.g., skin and blood biomarkers), and objectives associated with the open-label extension period are not included in this SAP.

5 Study Objectives and End points

5.1 Study Objectives

Primary Objectives
- To assess the safety of treatment with abatacept 125 mg SC versus placebo SC given every week
- To assess the efficacy of treatment with abatacept 125 mg SC versus placebo SC given every week on skin fibrosis using the modified Rodnan Skin Score (mRSS)

Secondary Objectives
To assess the efficacy of treatment with abatacept 125 mg SC versus placebo SC given every week on:
- Joint tenderness as measured by 28-tender joint count
- Joint swelling as measured by 28-swollen joint count
- Patient's and physician's global assessment on a Likert scale
- Health-related quality of life (HRQOL) using PROMIS-29 2.0
- Physical function as assessed by the scleroderma health assessment questionnaire-disability index (SHAQ-DI)
- Fatigue as assessed by the PROMIS Fatigue scale
- Sleep as assessed by the PROMIS sleep disturbance and impairment scale
- Gastrointestinal symptoms as assessed by UCLA SCTC GIT 2.0
- Combined Response Index in Systemic Sclerosis (CRISS)
- Percent predicted forced vital capacity (FVC)

Exploratory Objectives
To assess the efficacy of treatment with abatacept 125 mg SC versus placebo SC given every week on a core set of items developed for a composite index in early dcSSc:
- Patient interference with the skin involvement in the past month on a Likert scale
- Proportion of participants with new or worsened clinically significant heart disease, considered secondary to dcSSc, including congestive heart failure requiring hospitalization, new onset pulmonary hypertension requiring treatment, pericardial disease requiring intervention or exhibiting clinical decompensation, and arrhythmias and/or conduction defects requiring treatment
- Proportion of participants with new renal crisis
- Percent predicted carbon monoxide diffusing capacity (DLCO), corrected for hemoglobin
- FVC (in ml)

- Proportion of subjects with significant ILD defined by a decline in forced vital capacity (FVC)% predicted ≥15% (relative), high resolution computer tomography (HRCT) to confirm interstitial lung disease (ILD; if previous high resolution computer tomography of chest did not show ILD) and FVC% predicted below 80% predicted
- Change from baseline in body mass index
- Digital ulcer net burden as assessed by the investigator during the trial (baseline to 12 months)
- Pain intensity due to dcSSc over the past week on a 0-150 mm VAS

5.2 End points

5.2.1 Primary Efficacy End point
Change from baseline to month 12 in mRSS is the primary end point.

5.2.2 Secondary Efficacy End points
- Change from baseline to month 1, 3, 6, and 9 in mRSS
- Change from baseline to month 1, 3, 6, 9 and 12 in:
  - 28-tender joint count
  - 28-swollen joint count
- Change from baseline to month 3, 6 and 12 in:
  - Patient global assessment for overall disease
  - Physician global assessment for overall disease
  - PROMIS-29 Profile v2.0 measures in the following domains: physical function, anxiety, depression, fatigue, sleep disturbance, pain interference, and ability to participate in social roles and activities, and a single item on pain intensity
  - HAQ-DI overall (HAQ-DI) and 8 categories: Dressing and Grooming, Hygiene, Arising, Reach, Eating, Grip, Walking, Common Daily Activities (IADL)
  - Scleroderma-HAQ-DI visual analogue scales (VAS) assessing burden of digital ulcers, Raynaud's, gastrointestinal involvement, breathing, and overall disease
  - PROMIS fatigue measure
  - PROMIS sleep disturbance measure
  - PROMIS sleep impairment measure
  - UCLA SCTC GIT 2.0 final composite score and 7 scales (Reflux, Distension/Bloating, Fecal Soilage, Diarrhea, Social Functioning, Emotional Wellbeing, and Constipation)
- Change from baseline to 6 and 12 months in:
  - Combined Response Index in Systemic Sclerosis (CRISS)
  - Percent predicted FVC

5.2.3 Exploratory Efficacy End points
- Change from baseline to months 3, 6 and 12 in patient interference with the skin involvement
- The proportion of participants with cardiac involvement at 12 months, defined as new or worsened clinically significant heart disease considered secondary to SSc
- The proportion of participants with new renal crisis at 12 months
- Change from baseline to months 6 and 12 in percent predicted DLCO, corrected for hemoglobin
- Change from baseline to months 6 and 12 in FVC (in ml)
- The proportion of participants with development of significant ILD
- Change from baseline in body mass index at 12 months
- Change from baseline to month 12 in digital ulcer net burden
- Change from baseline to months 3, 6 and 12 in pain intensity due to SSc over the past week on a 0-150 mm VAS

6   Study Methods

6.1   General Study Design and Plan

This study is a randomized placebo-controlled double-blind phase 2 trial of patients with dcSSc. Eligibility for the study was assessed during a one-month screening period. Eligible participants were randomized in a 1:1 ratio to either 125 mg SC abatacept or matching placebo, stratified by duration of dcSSc disease duration (<18 months vs >18 to <36 months). Study participants were treated for 12 months on double-blind study medication, followed by an additional 6 months of open-label SC abatacept therapy and a 30 day follow-up phone call upon completion of the study.

86 patients were to be randomized from approximately 35 centers in the US, Canada and Europe, with the goal of analyzing 74 participants. Our study was designed to test whether abatacept is statistically superior to placebo in reducing the mRSS at month 12, relative to baseline, and explore the ability of abatacept to prevent or reverse progression in patients with early disease duration and lower mRSS scores, and reverse established disease in patients with longer disease duration and higher MRSS scores.

Escape Therapy

Starting at Month 6, participants with worsening of skin disease (defined as > 5 units worsening of mRSS) had the opportunity to add escape therapy to their randomized study medication (weekly abatacept or placebo SC). In addition, worsening of ILD as defined by absolute decline in FVC% predicted by $\geq 10\%$ or absolute decline in DLCO% predicted by ≥ 15 (confirmed by repeat pulmonary function testing within 1 month) had the opportunity to add escape therapy. Other indications included: active inflammatory polyarthritis or inflammatory myositis. The decision to initiate escape therapy was based on investigator discretion in eligible participants. Escape therapy included methotrexate, mycophenolate mofetil, cyclophosphamide, hydroxychloroquine, azathioprine or intravenous immunoglobulin (IVIG). Other biologic therapies were not acceptable as escape therapy.

Should a participant worsen at 3 months, the PI could decide that escape therapy should be initiated immediately. If this occurred, the participant was to be withdrawn from study medication. If the subject agreed to continue study follow up (through month 12), participating in visits and procedures, and complying with blood/tissue collection, the subject could participate in the open label phase.

The schema below describes the main elements of the study design:
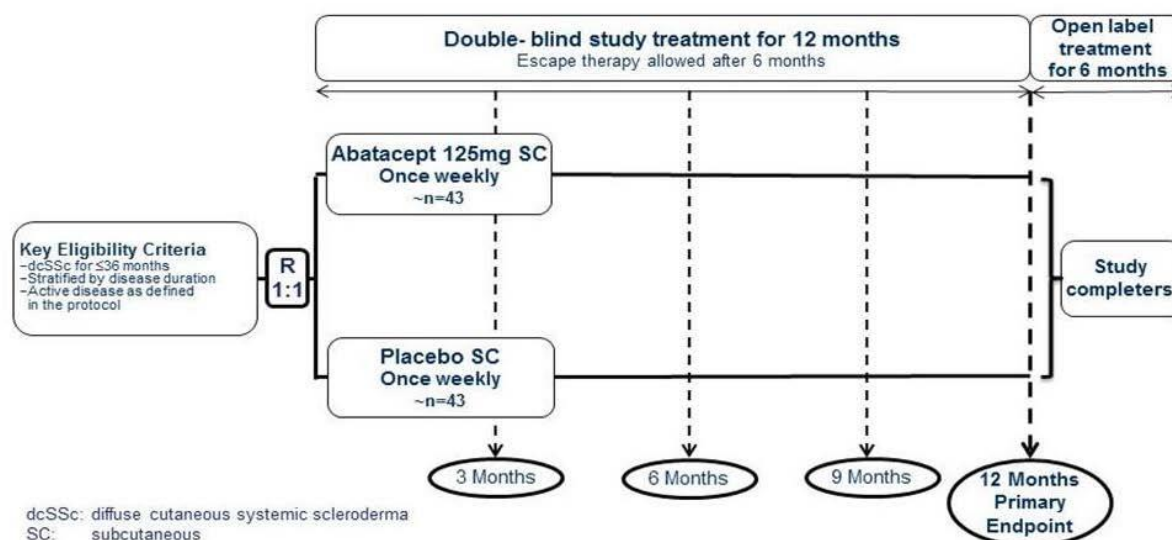
6.2   Inclusion-Exclusion Criteria and General Study Population

6.2.1   Inclusion Criteria
1. Signed Written Informed Consent
2. Diagnosis of SSc, as defined using the 2013 American College of Rheumatology/ European Union League Against Rheumatism classification of SSc
3. dcSSc as defined by LeRoy and Medsger
4. Disease duration of ≤ 36 months (defined as time from the first non−Raynaud phenomenon manifestation)
    For disease duration of ≤ 18 months
    - ≥ 10 and ≤ 35 mRSS units at the screening visit
    For disease duration of >18-36 months
    - ≥ 15 and ≤ 45 mRSS units at the screening visit and one of the following:
        1. Increase ≥ 3 in mRSS units compared with the last visit within previous 1–6 months
        2. Involvement of one new body area with ≥ 2 mRSS units compared with the last visit within the previous 1–6 months
        3. Involvement of two new body areas with ≥ 1 mRSS units compared with the last visit within the previous 1–6 months
        4. Presence of 1 or more Tendon Friction Rub
5. Age ≥ 18 years at the screening visit
6. If female of childbearing potential, the patient must have a negative pregnancy test at screening and baseline visits
7. Oral corticosteroids (≤ 10 mg/day of prednisone or equivalent) and NSAIDs are permitted if the patient is on a stable dose regimen for ≥ 2 weeks prior to and including the baseline visit
8. ACE inhibitors, calcium-channel blockers, proton-pump inhibitors, and/or oral vasodilators are permitted if the patient is on a stable dose for ≥ 2 weeks prior to and including the baseline visit

6.2.2   Exclusion Criteria
1. Rheumatic disease other than dcSSc; it is acceptable to include patients with fibromyalgia and scleroderma-associated myopathy
2. Limited cutaneous SSc or sine scleroderma at the screening visit
3. Major surgery (including joint surgery) within 8 weeks prior to screening visit
4. Any infected ulcer prior to randomization
5. Treatment with any investigational agent within ≤ 4 weeks (or 5 half-lives of the investigational drug, whichever is longer) of the baseline visit
6. Severe (MRSS 3+) skin on the inner aspects of thighs, upper arms, and abdomen
7. Previous treatment with cell-depleting therapies, including investigational agents, including but not limited to, CAMPATH, anti-CD4, anti-CD5, anti-CD3, anti-CD19, and ABA
8. Anti-CD20, and cyclophosphamide within 12 months prior to baseline visit
9. Use of Intravenous Immunoglobulin (IVIG) within12 weeks prior to baseline visit
10. Previous treatment with chlorambucil, bone marrow transplantation, or total lymphoid irradiation
11. Immunization with a live/attenuated vaccine within ≤ 4 weeks prior to the baseline visit
12. Treatment with methotrexate, hydroxychloroquine, cyclosporine A, azathioprine, mycophenolate mofetil rapamycin, colchicine, D-penicillamine, within ≤ 4 weeks prior to the baseline visit
13. Treatment with etanercept within ≤ 2 weeks, infliximab, certolizumab, golimumab, ABA or adalimumab within ≤ 8 weeks, anakinra within ≤ 1 week prior to the baseline visit
14. Pulmonary disease with FVC ≤ 50% of predicted, or DLCO (uncorrected for hemoglobin) ≤ 40% of predicted at the screening visit
15. Pulmonary arterial hypertension (PAH) as determined by right heart catheterization or on PAH approved medications for PAH. It is acceptable to use PDE-5 inhibitors for Raynaud's and digital ulcers.
16. Subjects at risk for tuberculosis (TB). Specifically excluded from this study will be participants with a history of active TB within the last 3 years, even if it was treated; a history of active TB

greater than 3 years ago, unless there is documentation that the prior anti-TB treatment was appropriate in duration and type; current clinical, radiographic, or laboratory evidence of active TB; and latent TB that was not successfully treated ($\geq$ 4 weeks).

17. Positive for hepatitis B surface antigen prior to the baseline visit
18. Positive for hepatitis C antibody, if the presence of hepatitis C virus was also shown with polymerase chain reaction or recombinant immunoblot assay prior to baseline visit
19. Any of the following prior to the baseline visit:
    - Hemoglobin <8.5 g/dL;
    - WBC < 3,000/mm3 (<3 x 109/L);
    - platelets < 100,000/mm3 (<3 x 109/L);
    - serum creatinine > 2 x ULN; or
    - serum ALT or AST > 2 x ULN
20. Any other laboratory test results that, in the opinion of the investigator, might place a participant at unacceptable risk for participation in the study.
21. The following medical history and concurrent diseases:
    - Subjects who are impaired, incapacitated, or incapable of completing study-related assessments.
    - Subjects with active vasculitis of a major organ system.
    - Subjects with current symptoms of severe, progressive, or uncontrolled renal, hepatic, hematologic, gastrointestinal, pulmonary, cardiac, neurologic, or cerebral disease, whether or not related to SSc and which, in the opinion of the investigator, might place a participant at unacceptable risk for participation in the study.
    - Subjects with a history of cancer in the last 5 years, other than non-melanoma skin cell cancers cured by local resection or carcinoma in situ. Existing non-melanoma skin cell cancers should be removed, the lesion site healed, and residual cancer ruled out before administration of the study drug.
    - Subjects who currently abuse drugs or alcohol.
    - Subjects with evidence (as assessed by the investigator) of active or latent bacterial or viral infections at the time of potential enrollment, including participants with evidence of human immunodeficiency virus (HIV) detected during screening.
    - Subjects with herpes zoster or cytomegalovirus (CMV) that resolved less than 2 months prior to screening.
    - Subjects with any serious bacterial infection within the last 3 months, unless treated and resolved with antibiotics, or any chronic bacterial infection (e.g., chronic pyelonephritis, osteomyelitis, or bronchiectasis).
22. Patients with a history of anaphylaxis to abatacept

6.3 Randomization and Blinding

Patients were randomized after all screening assessments were completed and the investigator verified that eligibility criteria were met. At the time of randomization, patients were assigned a unique randomization number; no participant was to begin treatment prior to randomization. Eligible participants were randomized to abatacept or placebo in a 1:1 manner, stratified by dcScc disease duration (<18 months vs >18 to <36 months). The DCC prepared the randomization schedule, using computer-generated block randomization with the block size(s) known only by the DCC. A secure web-based application was built for use by the coordinators to enter participant information (e.g., participant ID, stratification factor(s)) and to obtain the randomization number. The information was printed and sent and/or emailed directly to the site pharmacists.

6.4 Study Assessments

The following table provides the Schedule of Evaluations used in the study:

Study Period	Screening	Double-Blind Treatment						Open-Label Treatment			End of Study[l] (Phone call)
		Baseline	Month 1	Month 3	Month 6	Month 9	Month 12[j]	Month 14	Month 16	Month 18	
Study Week	≤ -4	0	4	12	24	36	48	56	64	72	76
Study Day	-28 to -1	0	28	84	168	252	336	392	448	504	534
Window (in days)		0	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)
Informed Consent	X										
Eligibility (Inclusion/Exclusion) Assessment	X										
Demographics	X										
Medical History	X										
SAFETY ASSESSMENTS											
Physical Exam	X		X	X	X	X	X	X	X	X	
Skin examination for cancer [k]	X		X	X	X	X	X	X	X	X	
Vital Signs	X	X	X	X	X	X	X	X	X	X	
Laboratory Tests[a]											
CBC, Differential, Comp Panel	X			X	X	X	X	X	X	X	
ESR	X			X			X			X	
PPD/QuantiFERON/TSpot	X										
Pregnancy Test[b]	X	X	X	X	X	X	X	X	X	X	
Hepatitis B and C	X										
Blood Collection for biomarkers (50 mL)[c]		X	X	X	X						
Skin Biopsy[d]		X		X	X						
Echocardiogram and Chest HRCT[e]	X *if available						X *if available				
Pulmonary Function Tests[f]	X				X		X			X	
Concomitant Medications	X	X	X	X	X	X	X	X	X	X	
Adverse Events		X	X	X	X	X	X	X	X	X	X
EFFICACY ASSESSMENTS											

Study Period	Screening	Double-Blind Treatment						Open-Label Treatment			End of Study[l] (Phone call)
		Baseline	Month 1	Month 3	Month 6	Month 9	Month 12[j]	Month 14	Month 16	Month 18	
Study Week	≤ -4	0	4	12	24	36	48	56	64	72	76
Study Day	-28 to -1	0	28	84	168	252	336	392	448	504	534
Window (in days)		0	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)	(±10)
mRSS	X[h]	X	X	X	X	X	X	X	X	X	
Joint Count, digital ulcer assessment, joint contractures, tendon friction rubs	X	X	X	X	X	X	X	X	X	X	
Patient Reported Outcomes[g]		X		X	X		X			X	
Exploratory outcomes[i]		X		X	X		X			X	
CLINICAL DRUG SUPPLIES											
Randomization		X									
Study Drug Supplied		X	X	X	X	X	X	X			
Study Drug Adherence			X	X	X	X	X	X	X	X	

[a] All laboratory samples will be analyzed at local lab

[b] For women of child-bearing potential

[c] Out of the 45 mL, 10 mL will be used for proteomics analysis, 5 mL will be collected in PAXgene tubes for RNA analysis, 29.5 mL will be used for flow cytometry, and 0.5 mL for autoantibody measurement

[d] Two 3-mm skin biopsies will be collected.

[e] Echocardiogram HRCT of Chest is part of standard of care assessments to be performed at the discretion of the PI. The data should be abstracted from the patient records at Screening and month 12. Test results from 6 months prior are acceptable.

[f] Spirometry with DLCO

[g] All Patient-reported outcomes include patient's and physician's global assessment, PROMIS-29 2.0, SHAQ-DI, PROMIS fatigue scale, PROMIS sleep disturbance and impairment scales, UCLA GIT 2.0 and the PRO-SRSS

[h] mRSS will be assessed as one of the inclusion criteria

[i] Patient interference with the skin involvement and pain intensity

[j] Should a subject terminate early the subject will be brought in to complete all assessments included at the Month 12 visit. Study drug however would not be provided but returned at that time.

[k] This skin evaluation will be an assessment of a need for referral to dermatology for formal cancer evaluation and will be assessed alongside every physical exam.

[j] **End of Study:** A Follow up Phone Call should occur 30 days post completion of OLE (18 month visit) or 30 days post End of double blind (12 month study) should participant decides not to enter OLE. Additionally, when a participant discontinues drug, a 30-day post last dose Early Term visit is scheduled; if the participant is unable or unwilling to return, a 30 day follow up phone call will suffice.

- Given flexibility around subject and care provider scheduling, visits are not required to have occurred on a specific date, but rather within a defined window.
- Where analyses reference the timing of outcomes and/or covariates, the nominal visit or time point as collected in the database will be used. For instances when there are scheduled visits and unscheduled visits in the same analysis time window, scheduled visits will be selected over unscheduled visits. If there are multiple observations for scheduled visits within a window, the one closest to the visit target date will be utilized. Where two observations are equi-distant from the target date the later will be utilized.
- For determination of escape therapy prior to 3 months, the month 3 visit date will be used; i.e., if escape therapy starts prior to the month 3 visit date (not the lower bound, target date, or upper bound of the analysis time windows defined below), then the participant will be considered to have started escape therapy prior to 3 months.

Analysis Time Windows

Visit Study Month (Day)	Lower bound of Window (Day)	Target Date (Day) and Per protocol window	Upper bound of Window (Day)
Screening	-28	-28 to -1	-1
Baseline (0)	0	0	0
Month 1 (28)	1	28 ± 10	56
Month 3 (84)	57	84 ± 10	126
Month 6 (168)	127	168 ± 10	210
Month 9 (252)	211	252 ± 10	294
Month 12 (336)	295	336 ± 10	372
Month 14 (392)*	373	392 ± 10	420
Month 16 (448)*	421	448 ± 10	476
Month 18 (504)*	477	504 ± 10	532

*note that these visits are not included in the analyses described in this SAP; they will be used in the analyses of the open-label extension period.

6.4.1 Primary Efficacy Assessment

The primary efficacy end point is based on the Modified Rodnan skin score, a validated physical examination method for estimating skin induration. It is correlated with biopsy measures of skin thickness and reflects prognosis and visceral involvement, especially in early disease. It is scored on a 0 (normal) to 3+ (severe induration) ordinal scales over 17 body areas, with a maximum score of 51 and is used to categorize severity of SSc. It has been extensively used as primary/ secondary outcome in RCTs. This will be collected at every study visit.

6.4.2 Secondary Efficacy Assessments

- 28-Tender joint count: Investigator will assess tenderness of the joints and score them as positive or negative. The joints include: proximal interphalangeal joints, metacarpophalangeal joints, wrists, elbows, shoulders, and knees. It is performed at every visit.
- 28-Swollen joint count: Investigator will assess swelling of the joints and score them as positive or negative. The joints include: proximal interphalangeal joints, metacarpophalangeal joints, wrists, elbows, shoulders, and knees. It is performed at every visit.

- Patient global assessment for overall disease: This assessment represents the patient's assessment of the patient's global scleroderma on a 0-10 Likert scale. "On a scale of 0-10, how was your overall health in the last week? 0=Excellent; 10=Extremely Poor. It is assessed at baseline and months 3, 6, 12 and 18.

- Physician global assessment for overall disease: This assessment represents the physician's assessment of the patient's current disease activity on a 0-10 Likert scale. "On a scale of 0-10, how was your patient's overall health in the last week? 0=Excellent; 10=Extremely Poor" . Assessed at baseline and months 3, 6, 12 and 18.

- PROMIS-29 Profile v2.0 measure: The National Institutes of Health (NIH) Patient-Reported Outcomes Measurement Information System (PROMIS®) Roadmap initiative (www.nihpromise.org) is a cooperative research program designed to develop, evaluate, and standardize item banks to measure patient-reported outcomes (PROs) across different medical conditions as well as the US population. PROMIS-29 Profile v2.0 measure contains 29 items, which includes four items each from physical function, anxiety, depression, fatigue, sleep disturbance, pain interference, and satisfaction with social roles domains, and a single item on pain intensity. With the exception of physical function which does not include a time frame, all item banks reference the past 7 days. Three scores are available for each PROMIS domain: cumulative score, instrument score and transformed score; the transformed score (Tscore) will be the score analysed in this study. It is assessed at baseline and months 3, 6, 12 and 18.

- SHAQ-DI: The SHAQ consists of the HAQ-DI (8 domains and an overall score) and 6 visual analogue scales assessing the burden of pain, digital ulcers, Raynaud's, gastrointestinal involvement, breathing, and overall disease.[12] The HAQ-DI is a disease-targeted, musculoskeletal-targeted measure intended for assessing functional ability in arthritis. It is a self-administered 20-question instrument that assesses a patient's level of functional ability and includes questions that involve both upper and lower extremities. The score for each question ranges from 0 (no disability) to 3 (severe disability). There are 8 categories and an overall score (HAQ-DI). It has a 7-day recall period and has been extensively used in SSc.[13] It is assessed at baseline and months 3, 6, 12 and 18.

- PROMIS fatigue measure: Apart from assessing the PROMIS-29 measure that assesses overall HRQOL, we will assess fatigue as it is one of the common symptoms of patients with SSc. We will administer the 8-item short form with 1-week recall (available at www.nihpromis.org). It is assessed at baseline and months 3, 6, 12 and 18.

- PROMIS sleep disturbance and sleep impairment measures: Sleep disturbances are rated as the one of the top complaints from the patients with SSc and will be assessed using 4 items each with 1 week recall. (available at www.nihpromis.org). They are assessed at baseline and months 3, 6, 12 and 18.

- UCLA SCTC GIT 2.0: This validated instrument assesses scleroderma-related gastrointestinal symptoms. It has 7 scales and a final composite score (available at http://uclascleroderma.researchcore.org. It is assessed at baseline and months 3, 6, 12 and 18.

- Combined Response Index in Systemic Sclerosis (CRISS): CRISS is a composite measure for early dcSSc[13].[14] It is determined in a 2-step process that assesses the probability of deterioration (step 1) and of improvement (step 2), where each probability ranges from 0.0 to 1.0. The first step assesses whether the patient has had a significant decline in renal or cardiopulmonary involvement. The second step assesses the probability of improvement by incorporating changes in the modified Rodnan skin score, percent predicted forced vital capacity (FVC), patient and physician global assessments, and SHAQ-DI over 1 year.   It is assessed at 6 and 12 months.

- Percent predicted FVC and FVC are assessed at Screening and months 6, 12 and 18. The calculation of percent predicted FVC is based on equations from Hankinson[15] (see table below).

Result name	Formula
Age	(Date of PFT - Date of birth)/365.25 and then rounded down (always down) to the nearest integer.
FVC Predicted (Reference)	Where H = height in centimeters and Age = age at last birthday

FVC Predicted (Reference):

Caucasian
Males≥20: $FVC(L) = 0.00018642*H^2 + 0.00064*Age - 0.000269*Age^2 - 0.1933$
Males<20: $FVC(L) = 0.00018642*H^2 - 0.20415*Age + 0.010133*Age^2 - 0.2584$
Females≥18: $FVC(L) = 0.00014815*H^2 + 0.01870*Age - 0.000382*Age^2 - 0.3560$

African-American
Males≥20: $FVC(L) = 0.00016643*H^2 - 0.01821*Age - 0.1517$
Males<20: $FVC(L) = 0.00016643*H^2 - 0.15497*Age + 0.007701*Age^2 - 0.4971$
Females≥18: $FVC(L) = 0.00013606*H^2 + 0.00536*Age - 0.000265*Age^2 - 0.3039$

NOTE 1: Those subjects who indicate that they are both African American and Mexican-American or Hispanic will use the African American reference equations; those subjects who indicate that they are both Caucasian and Mexican-American or Hispanic will use the Caucasian reference equations.
NOTE 2: Those subjects who indicate that they are Asian will use 0.88x the Caucasian values.
NOTE 3: If Race is "Unknown or Not Reported", then use Caucasian values.

Result name	Formula
FVC Percent Predicted	(FVC Observed / FVC Predicted) * 100

6.4.3 Exploratory Efficacy Assessments

- Patient interference with the skin involvement in the past month on a 0-10 Likert scale. On a scale of 0-10, in the last month how much has your skin involvement interfered with your daily activities? 0=Does not limit activity; 10= Very severe limitation. This will be assessed at baseline and month 3, 6, 12 and 18.
- Cardiac involvement at 12 months: New/worsened clinically significant heart disease considered secondary SSc, including any of the following: heart failure requiring hospitalization, new onset pulmonary hypertension requiring specific treatment, pericardial disease requiring intervention or clinical decompensation, and arrhythmias and/or conduction defects requiring treatment).
- New renal crisis at 12 months.
- Significant ILD defined by a decline in forced vital capacity (FVC)% predicted ≥15% (relative), high resolution computer tomography (HRCT) to confirm interstitial lung disease (ILD; if previous high resolution computer tomography of chest did not show ILD) and FVC% predicted below 80% predicted
- Percent predicted DLCO, corrected for hemoglobin. It is assessed at screening, and months 6, 12, and 18.
  The equation used for adjusting predicted DLCO for hemoglobin is as follows:

  For adult males, the equation (expressing Hb in gm•dL-1) is:
  DL,CO,predicted × (1.7Hb / (10.22+Hb))
  In adult women, the equation is:
  DL,CO,predicted × (1.7Hb / (9.38+Hb))

- Height and weight are assessed at baseline and 12 months. Body mass index is calculated as weight (kg) / height x height (m^2).
- Digital ulcer net burden as assessed by the investigator during the trial (after randomization to 12 months): Digital ulcer net burden is defined as the number of overall digital ulcers as assessed by the investigator. A digital ulcer is defined as an ulcer at or distal to the

metacarpophalangeal joint with loss of surface epithelialization. This does not include fissures, cracks or calcium extrusion from calcinosis cutis.

- Pain intensity due to SSc over the past week on a 0-150 mm VAS assessed at baseline, and 3, 6, 12 and 18 months.

6.5   Imputation of Dates

If partial dates occur, the convention for replacing missing dates for the purpose of calculating derived variables is as follows:

For partial original SSc diagnosis dates: (a) if only the day is missing, and the month and year match the first dose date, then the day is assigned the first day of the month (01); otherwise the day assigned is 15; and (b) if both the day and month are missing then the day/month assigned is the first day of July (01JUL), as long as the date is before the first dose date; otherwise, the day/month assigned is the first day of January (01JAN).

If start dates are entirely missing for adverse events or medications, then adverse events will be classified as treatment-emergent and medications will be classified as concomitant. For partial AE or concomitant medication start dates or end dates, the table below describes the date imputation.

Missing	Condition – START DATE	Condition – END DATE	Imputation	Classification
Start Day	Start month & year match those of the first dose date	End date ≥ first dose date or ongoing	Set start date = first dose date	AE = treatment-emergent & med = concomitant
		End date < first dose date	Set start date = 01	AE = not treatment emergent & med = prior
	Start year < first dose year	End date ≥ first dose date or ongoing	Set start date = 15	AE = not treatment-emergent & med = concomitant
		End date < first dose date	Set start date = 01	AE = not treatment emergent & med = prior
	Start year = first dose year and start month < first dose month	End date ≥ first dose date or ongoing	Set start date = 15	AE = not treatment-emergent & med = concomitant
		End date < first dose date	Set start date = 01	AE = not treatment emergent & med = prior
	Start year = first dose year and start month > first dose month	All cases	Set start date = 15	AE = treatment-emergent & med = concomitant
Start Day & Start Month	Start year matches first dose date	End date ≥ first dose date or ongoing	Set start date = first dose date	AE = treatment-emergent & med = concomitant
		End date < first dose date	Set start date = 1 and start month = end date month	AE = not treatment emergent & med = prior
	Start year < first dose year	All cases	Set start date = 1 and start month = JAN	
	Start year > first dose year	All cases	Set start day =1 and start month = JAN	AE = treatment-emergent & med = concomitant
End Day	NA	All	Set end date = min(last day of month, day of visit date)	NA
End Day & End Month	NA	All	Set end day =31 and end month = DEC, or day and month of visit date if earlier	NA

6.6   Laboratory Reporting

In general, for quantitative laboratory values reported as "<" or "≤" the lower limit of quantitation (LLOQ) or limit of detection (LOD), one-half of the reported value (i.e., LLOQ/LOD) will be used for analysis.. For quantitative laboratory values reported as ">" or "≥" the upper limit of quantitation (ULOQ), the reported value (i.e., ULOQ) will be used for analysis. The number and percentage of

subjects with values $\leq$ or $\geq$ limits of quantitation or detection will also be provided.

For analysis purposes, repeat laboratory test results will not be used unless the original laboratory value is missing or indicated as invalid, in which case the first non-missing repeat laboratory value will be used for data analysis.

7   Sample Size

Previous randomized controlled studies in early dcSSc provide the basis for us to characterize the magnitude of treatment differences we could detect with sample sizes of 74-86 participants with a two-sided Type I error of 5%, 80% power, and a drop-out rate of 15%. The phase 2 randomized controlled trial of recombinant human relaxin vs. placebo in participants studied moderate-to-severe dcSSc patients who are similar to those considered for this study. Khanna et al.[13] found no statistically significant differences in the change from baseline to week 24 in mRSS between relaxin and placebo. The pooled standard deviation (SD) of the change in mRSS is approximately 7 points. We conservatively selected a larger estimate of the SD for our sample size calculations.

This phase 2 study was primarily sized based on practical considerations, rather than a desired power for a pre-specified difference. We planned to screen approximately 121 participants in order to randomize 86 participants to achieve 74 analyzable participants (assuming a 15% attrition rate). With the proposed sample of 74 participants in the study (37 per treatment group), there is at least 80% power to detect a 24% treatment difference in proportions of participants with adverse events with a two-sided Type I error of 5% and a placebo rate of 70% (two-sample test of binomial proportions, East 5.4). For continuous outcomes, there is at least 80% power to detect an effect size of at least 0.66 with a two-sided Type I error of 5% with this sample size (two-sample t test, East 5.4). This effect size (treatment difference / pooled SD) translates into a treatment difference in change from baseline to month12 in mRSS of 5.3 with a SD of 8 points. If the pooled SD or drop-out rate is smaller, then the given sample size would allow for smaller treatment differences to be detected with the same power.

8   General Analysis Considerations

8.1   Timing of Analyses

The final analysis of the double-blind period will be performed after all randomized participants have completed their 12-month visit or dropped out prior to their 12-month visit, all corresponding data have been entered, cleaned, locked and unblinded as per SABER SOPs. This SAP document was finalized and approved prior to the double-blind database lock and unblinding.

8.2   Blinded Data Review

Prior to database lock and the start of any formal analyses, blinded data reviews will be completed, including summary statistics of key variables. This will allow the data for key variables to be examined to identify unusual values that need to be queried and patterns of missing values.   In addition, the data reviews will allow the protocol writing committee to assess the format of the data presentations. Note that blinded data reviews incorporate real data but random treatment assignment (i.e., investigators do not receive data summarized by actual treatment group, rather they review data on two randomly formed groups). All decisions will be made and documented in this SAP document prior to unblinding and database lock.

8.3   Analysis Populations

All randomized participants will be used in the analyses of subject disposition.

8.3.1   Modified Intention to Treat Population

The main analysis set for efficacy is the modified intention to treat (mITT) population, defined as all participants randomized and receiving at least one dose of study medication. Participants are analyzed by assigned treatment. Membership in the mITT analysis population was determined before

study unblinding.

8.3.2   Per Protocol Population

A per protocol (PP) population is used for sensitivity analyses of the primary end point. It is defined as the mITT population, excluding all participants who have major protocol deviations. Major protocol deviations are defined as eligibility criteria violations for which no exemption was granted, study drug compliance <80% and >120% , and receipt of escape medication prior to month 3. Membership in the PP analysis population will be determined before study unblinding.

8.3.3   Safety Population

The Safety Population is defined as all participants who are randomized and receive at least one dose of study drug. The Safety Population will be used for all safety analyses, as well as demographic and baseline analyses. Subjects will be analyzed by the treatment received if they received the wrong treatment for the entire duration of the double-blind period of the study.

8.4   Covariates and Subgroups

There are a limited number of covariates that will be incorporated in statistical models in our analyses because of the relatively small sample size in each treatment group: the stratification factor duration of dcSSc diseaseand baseline outcome measure. Baseline values of mRSS are available for all subjects. We will not impute missing values for other baseline covariates in secondary and exploratory analyses in the mITT analysis set.

The primary end point will be summarized by the following subgroups: (1) baseline mRSS $\leq$ 22 vs >22, (2) baseline SCL-70 positive vs other autoantibodies; and (3) RNA pol 3 positive vs other autoantibodies.

8.4.1   Multi-center Studies

Given that dcSSc is a relatively rare disease, many centers were required to obtain the required sample size. Study centers will not be incorporated as stratification into the analyses.   Descriptive statistics of the primary end point by treatment group, separately by center (Michigan, Toronto [the two largest enrolling sites] and the remainder) will be provided.

8.5   Missing Data

We will summarize the extent of missing data over time for the primary end point. We will investigate the missing data mechanism (missing at random, not missing at random), which is important for the validity of our analytic approaches, through exploratory analysis. Exploratory analyses will include plots of the mean profile of mRSS at months 0, 3, 6, 9 and 12 by treatment group for those who have complete data throughout the study and those who don't, as well as plots of the mean change from baseline at months 3, 6, 9, and 12 in mRSS in the two treatments within each group (completers and non-completers). If the plots reveal consistent differences between completers and non-completers within each of the treatment groups, then there is evidence that data are not missing at random.

The primary analysis of the primary end point (see section 10.1) assumes a missing at random mechanism. If data are not missing at random, we will use a multiple imputation approach within the pattern-mixture model framework.[16] The imputation models, applied sequentially for each missing data pattern, will include baseline mRSS, treatment group, stratification factor (SSc disease duration), and demographic variables (age and gender), allowing for the dependence of later time points on earlier time points. The analysis model will include the same covariates used in the primary analysis of the primary end point and will incorporate the uncertainty due to imputation in the calculation of the standard error, as described by Rubin.[17]

8.6   Interim Analyses and Data Monitoring

No formal interim analyses were planned nor carried out for this study. The study was overseen by a

Data and Safety Monitoring Board (DSMB) that reviewed the pooled and by-treatment subject disposition, study conduct and safety data approximately every 6 months.

8.7   Multiple Testing
Two-sided p-values will be reported, and no adjustments for multiplicity will be made. Thus, p-values for secondary and exploratory outcomes will be interpreted with caution. Confidence intervals will be provided to summarize treatment differences for efficacy end points.

9   Summary of Study Data
Descriptive summary statistics will be derived for all data at baseline, separately by treatment group and overall. For efficacy, exploratory and safety data, data will be presented by treatment group. Treatment group will be characterized as "Abatacept" and "Placebo"; for pooled summaries, "Overall" will be used as the column heading. All tables will be annotated with the total population size relevant to that table/treatment, including any missing observations.

For continuous variables, mean, standard deviation, median, interquartile range, minimum and maximum will be reported. For categorical variables, number and percentages will be reported (excluding missing values). Graphical methods will be heavily used in this Phase 2 study to assess the pattern of response over time for key variables and to assess the relationships among variables.

9.1   Subject Disposition
The number of participants approached for study participation, the number consented and the number who did not consent (including reasons: screen failures, refusals) will be summarized in a CONSORT diagram. The number of participants who dropped out prior to randomization, and the reasons for dropout, will be summarized. The number randomized and treated and the number who dropped out by months 1, 3, 6, 9 and 12 will be provided, as well as the number in each of the analysis populations (i.e., mITT, PP, Safety). Reasons for post-treatment dropout will be provided.

9.2   Protocol Deviations
Major protocol deviations that exclude a patient from the Per Protocol Population are described in section 8.3.2. A listing of protocol deviations that exclude participants from the Per Protocol population will be provided. A listing of participants who receive exemptions for study eligibility will also be provided.

9.3   Demographic and Baseline Variables
Demographic variables include: age at consent (defined as a continuous variable, e.g., 52.6 years), age by category (18 to 35 years, >35 to 55 years, >55 to 75 years, and >75 years), gender, race and ethnicity.

Baseline is defined as pre-treatment measures, either at screening (if a measure was only assessed at screening) or at baseline (if a measure was assessed at baseline even if also assessed at screening). Baseline variables include:
- dcSSc disease duration ($\leq$ 18 vs >18 to $\leq$36 months), dcSSc disease duration (analyzed as a continuous variable), disease duration since first non-Raynaud's sign or symptom (years), disease duration since Raynaud's phenomenon (years)
- mRSS, percent predicted FVC, percent predicted DLCO corrected for hemoglobin, HAQ-DI, SHAQ VAS measures (pain from illness, intestinal problems interfere with daily activities, breathing problems interfere with daily activities, Raynaud's interfere with daily activities, finger ulcers interfere with daily activities), physician global assessment, patient global assessment, autoantibodies, use of biologics, use of immunosupressives, use of prednisone, dose of prednisone, proportions of participants with tendon friction rubs, proportions of participants with large joint contractures, swollen joint count, proportion of participants with swollen joint count > 0, tender joint count, and proportion of participants with tender joint count >0

- height (cm), weight (kg), creatinine (mg/dL), and hemoglobin (g/dL).

9.4   Treatment Compliance

Compliance with study medication (injections) will be calculated, for each participant, as the proportion of time (weeks) that a participant took the full or partial contents of the syringe. Specifically, the percent compliance is calculated as 100 x the ratio of the number of weeks during the double-blind period when the participant took the full or partial amount of the syringe divided by the number of weeks during the double-blind period during which the participant was expected to take study medication. Participants were expected to take study medication unless it was temporarily discontinued due to an AE or permanently discontinued due to escape therapy prior to month 3. The study medication log (Form 027), adverse event form (Form 044), serious adverse event form (Form 045) and final status form (Form 035) are used to derive the compliance measure.

The summary statistics will be produced in accordance with section 9.

10   Efficacy Analyses

10.1   Primary Efficacy Analysis

The primary efficacy end point is the change from baseline to month 12 in mRSS scores. The goal of the primary analysis will be to test the null hypothesis that the difference between treatments in the primary end point is zero. The analysis will be performed on the mITT population. For the primary analysis, we will use a linear mixed model[18] with month in the study (3, 6, 9 and 12) as the unit of analysis and the change from baseline in mRSS as the outcome, with terms for treatment group, month, the interaction of treatment group with month, baseline mRSS, and duration of dcSSc disease (stratification factor). Study participant will be treated as a random effect to account for both heterogeneity among participants and correlation among measurements taken on the same participant. An autoregressive (AR(1)) variance-covariance will be assumed. Given that the incorporation of escape therapy after month 3 is an indication of treatment failure, we censor primary end point data after initiation of escape therapy.

Predicted mean change from baseline to month 12 for an exemplary participant by treatment group will be provided, as well as the estimate of the treatment effect at month 12, adjusted for baseline covariates, and corresponding 95% confidence interval and p-value for the treatment effect. Parameter estimates will be calculated using restricted maximum likelihood (REML) methods, and the Kenward-Rogers method will be used to calculate the degrees of freedom.

10.2   Secondary Efficacy Analyses

10.2.1   Secondary Analyses of Primary Efficacy End point

Several sensitivity analyses will be performed to assess how alternative approaches to missing data and model assumptions affect the conclusions of the analysis of the primary outcome:

- Analysis of the primary efficacy variable as described above will also be performed on the PP Population.
- Analysis of the primary efficacy variable as described above in the mITT Population, except that all primary endpoint data will be incorporated (i.e., no censoring after escape therapy).
- Analysis of the primary efficacy variables using time-in-study as a continuous variable (expressed as a fraction of a year) instead of the discrete version used in the primary analysis.
- If mRSS does not change linearly as a function of time-in-study, we will extend the linear mixed model to include: a polynomial of time-in-study (expressed as a fraction of a year) of an appropriate degree, the interaction between treatment and the polynomial of time-in-study, duration of dcSSc disease (stratification factor), and patient-specific random effects, Given the limited sample size of this phase 2 study, we will carefully assess to the fit of these longitudinal models.

- Analysis of the primary efficacy variable as described above in the mITT Population, except that we will account for the possibility of escape therapy by adding a variable that indicates the time (in months) the escape therapy was added to the participant's randomized study medication. The model will account for escape therapy by including, at the time points following the beginning of the participant's escape therapy, the time escape therapy began and the interaction term between treatment and time escape therapy began. The model will generate adjusted estimates of change in the mRSS score from baseline for each treatment group and month. To test the impact of treatment on mean changes from baseline to month 12, a linear contrast of the effect of treatment and the interaction of treatment X month will be assessed at month 12. The model is summarized below.

Change in $mRSS_{it} =$
$$b_0 + b_{1t} \times TREATMENT_i + b_2 \times MONTH_{it} + b_3 \times [TREATMENT_i * MONTH_{it}] +$$
$$b_4 \times SSc\_DURATION_i + b_5 \times BL\_mRSS_i +$$
$$b_6 \times ESCAPE_{it} + b_7 \times [TREATMENT_i * ESCAPE_{it}] +$$

Fixed effects

$$b_{oi} + b_{1i} \times TREATMENT_i + e_{it}$$

Random effects

for i=1, …, n participants, t= 1, 2, 3, 4 (corresponding to 3, 6, 9, 12 months) and $ESCAPE_{it}$ = the number of months after escape therapy started. We assume that residual errors e_{it} (t = 1, 2, 3, 4) for participant i are normally distributed with zero mean and 4x4 AR(1) variance-covariance matrix.

10.2.2 Analyses of Secondary End points
Analysis for secondary outcome measures will be performed using the same approach to that for the primary end point, that is, a linear mixed model with censoring of secondary end point data after initiation of escape therapy.

10.3 Exploratory Efficacy Analyses
Graphical methods will be used to explore the distributions of exploratory outcomes by treatment group (using boxplots), and the inter-relationships among the exploratory outcomes and the primary efficacy end point and the following secondary end points (using scatterplots): HAQ-DI, percent predicted FVC, patient global assessment, physician global assessment and CRISS scores. Only exploratory outcomes assessed at 12 months will be formally tested for treatment differences. For continuous outcomes, ANCOVA models (or non-parametric alternatives) will be used with covariates duration of dcScc (stratification factor) and baseline score. For binary outcomes, Cochran-Mantel-Haenszel (CMH) tests will be used, stratified by duration of dcScc. Analyses for the exploratory efficacy variables will be performed on the mITT population.

11 Safety Analyses
Safety data, including AEs, clinical laboratory tests, vital signs, physical examinations, and concomitant medication usage will be summarized descriptively by treatment group for the Safety Population; select parameters will also be summarized for the entire population (overall).

11.1 Extent of Exposure
Total participant weeks of exposure to study medication through the month 12 visit will be summarized descriptively by treatment group. Both full and partial amount of drug received will count in the assessment of exposure.

11.2 Adverse Events
Treatment-emergent adverse events are AEs that start on or after the first study day treatment is administered. The causal relationship of the AE to the study drug is determined by the site PI investigator as "not related" or "related". Adverse event severity grades are reported according to the CTCAE Version 4.0. If the CTCAE does not have a grading for a particular adverse event, the

severity of the event is reported by the investigator as mild, moderate, severe, or very severe. In the case of multiple occurrences of the same AE within the same subject, AEs will be summarized according to the maximum severity reported for each body system and overall.

Descriptive summary statistics for treatment-emergent adverse events (AEs) will be reported. The number of treatment-emergent AEs and the frequencies (number and percentage) of participants with one or more treatment-emergent AE will be summarized by treatment group, overall, by severity, and by body system. Coding of adverse events into body system was performed by the study chair for adverse events and by the medical monitors for serious adverse events. All treatment-emergent AEs related to study drug will be summarized, as will the frequencies of participants with one or more treatment-emergent AE related to study drug. Similarly, all treatment-emergent AEs causing study discontinuation, and frequencies of participants experience these, will be summarized.

A subject listing of all treatment-emergent AEs and treatment-emergent AEs causing study discontinuation will be presented. A subject listing prednisone and rescue (escape) drug use in relation to adverse events will be presented.

In addition, the number of treatment-emergent AEs potentially casually related to abatacept will be summarized by treatment group. These were identified in the Investigator's Brochure and through correspondence with BMS. These include: (1) any infection, local or systemic, that requires oral or systemic treatment; (2) any new malignancy or reoccurrence, including skin cancers and cancer in situ; (3) injection site reaction; (4) new autoimmune disorders, including psoriasis, cutaneous vasculitis, Sjogren's/sicca, E. nodosum, episcleritis; and (5) COPD exacerbation in participants with COPD.

In accordance with clincaltrial.gov reporting requirements, the following table summarizing adverse events is required and will be provided:
  ● Other (Not Including Serious) Adverse Events: A table of anticipated and unanticipated events (not included in the serious adverse event table) that exceed 5% within either treatment group, grouped by organ system, with number and frequency of such events in each treatment group.

Adverse events that occurred after consent and before treatment will be listed.

The summary statistics will be produced in accordance with section 9.

11.3 Deaths, Serious Adverse Events and other Significant Adverse Events

Descriptive summary statistics for treatment-emergent serious adverse events (SAEs) will be reported. The number of treatment-emergent SAEs and the frequencies (number and percentage) of participants with one or more treatment-emergent SAE will be summarized by treatment group, overall and by body system. Coding into body system was performed by the medical monitors for SAEs. All treatment-emergent SAEs related to study drug will be summarized, as will the frequencies of participants with one or more treatment-emergent SAE related to study drug. Similarly, all treatment-emergent SAEs causing study discontinuation, and frequency of participants experiencing these, will be summarized.

The number and proportion of participants with at least with one or more treatment-emergent infectious and non-infectious SAEs will be summarized. Similarly, the following body systems and corresponding SAEs will be summarized: infections, cardiovascular disorders, gastrointestinal disorders, musculoskeletal and connective tissue disorders, skin and subcutaneous tissue disorders, vascular disorders, blood and lymphatic system disorders, general disorders and administrative site conditions, nervous system disorders, renal and urinary disorders and psychiatric disorders.

A subject listing of all treatment emergent SAEs, SAEs causing study discontinuation, and deaths

(including the post-treatment follow-up period through month 12) will be presented.

In accordance with clincaltrial.gov reporting requirements, the tables below summarizing deaths and SAEs are required:
- All-Cause Mortality: A table of all anticipated and unanticipated deaths due to any cause, with number and frequency of such events in each treatment group.
- Serious Adverse Events: A table of all anticipated and unanticipated serious adverse events, grouped by organ system, with number and frequency of such events in each treatment group.

The summary statistics will be produced in accordance with section 9.

11.4 Pregnancies
A listing of all pregnancies occurring after the start of study medication will be provided.

11.5 Clinical Laboratory Evaluations
Clinical laboratory (complete blood count (CBC) with differential, platelets and ESR, and chemistry) test parameters will be summarized using descriptive statistics by study visit and treatment group, as both observed value at the time point of interest, change from baseline, and the percent change from baseline values. The summary statistics will be produced in accordance with section 9.

For laboratory and other safety parameters that are continuous, Wilcoxon rank sum tests will be performed to compare the two treatment groups for changes from baseline to month 12 (or last time point on study medication).

The number and proportion of participants with potential drug-induced liver injury (DILI) over the course of the 12-month double-blind period will be summarized in accordance with section 9. These include: (1) ALT or AST elevation > 3 times the upper limit of normal (ULN), (2) total bilirubin > 2 x ULN; (3) alkaline phosphatase > 2 x ULN; (4) hemoglobin < 8 mg/dL; (5) decrease in hemoglobin > 2 gm/dL; and (6) Hy's law.

The incidence of clinically-meaningful shifts from baseline for select laboratory measures will be provided by treatment group. These include ALT, AST, total bilirubin, alkaline phosphatase, haemoglobin, using the thresholds defined above to define clinically-meaningful shifts from baseline to month 12. In addition, serum creatinine will be summarized, with a 50% worsening from baseline to month 12 indicating a clinically meaningful outcome.

11.6 Prior and Concurrent Medications

The proportion of participants on prednisone medications prior to the start of study medication will be summarized in the baseline table by treatment group and overall, using summary statistics in accordance with section 9.

The proportion of participants who begin escape therapy (overall and by type) will be summarized by treatment group and by time point and overall, using summary statistics in accordance with section 9. A listing of participants who begin escape therapy and the reason for its initiation will be provided.

No medication coding dictionary was used in this study. The investigators characterized concomitant medications by name, and classified them as escape therapy or not.

11.7 Other Safety Measures
Vital signs (temperature, respiratory rate, systolic and diastolic blood pressure, and pulse), weight and BMI will be summarized using descriptive statistics by clinical visit (through month 12) and treatment group, as both observed value at the time point of interest and the change from baseline values.

12 Other Analyses

Analyses of the following outcomes will be summarized (in accordance with section 9) and allow for further interpretation of the study results.

- Responder analysis of the percent change from baseline to month 12 in mRSS, dichotomized using several different thresholds to define improvement: 20%, 40% and 60%. Treatment differences will be tested use CMH, stratified by duration of dcSSc.
- Incidence of clinically meaningful changes in change from baseline to month 12 in mRSS, where improvement is defined as change > 5, worsening as change <-5, and no change as -5 $\leq$ change $\leq$ 5.
- Incidence of clinically meaningful changes in change from baseline to month 12 in mRSS, where improvement is defined as change > 4, worsening as change <-4, and no change as -4 $\leq$ change $\leq$ 4.
- Incidence of clinically meaningful changes in change from baseline to month 12 in mRSS, where improvement is defined as change $\geq$25%, worsening as change $\leq$-25%, and no change as -25% < change < 25%.
- Changes from screening to months 6 and 12 in FEV1 and FEV1/FVC ratio.
- Number and proportion of participants who improved, stayed the same, or worsened at month 12 in chest HRCT findings.
- Number and proportion of participants who improved, stayed the same, or worsened at month 12 in Doppler Echo results.
- The proportion of participants with tendon friction rubs at baseline and months 1, 3, 6, 9 and 12.
- The proportion of participants with large joint contractures at baseline and months 1, 3, 6, 9 and 12.
- The proportion of participants with small joint contractures at baseline and months 1, 3, 6, 9 and 12.

13   Reporting Conventions

P-values $\geq$0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as "<0.001". The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, or minimum and maximum will use the same number of decimal places as the original data. Estimated parameters, not on the same scale as raw observations (e.g., regression coefficients) will be reported to 3 significant figures.

14 Summary of Changes to the Protocol and/or SAP

The changes from the protocol-specified definitions of aims, outcomes and statistical analytic approaches are outlined below. These changes reflect advances in our knowledge of scleroderma since the design of the study in 2013-2014 that were not incorporated as protocol amendments, but were discussed during the formation of the Statistical Analysis Plan. These changes are documented herein and represent changes made prior to the database lock.

1. **Additional Secondary Objective**
There is a lack of consensus about the way to summarize FVC – using the original scale (ml) or using the percent predicted FVC. To allow for comparison with the literature that supports both approaches, we added a secondary objective to include FVC on the original scale.

PROTOCOL:
N/A

SAP:
Section 5.1 Study Objectives / Exploratory Objectives
To assess the efficacy of treatment with abatacept 125 mg SC versus placebo SC given every week on FVC (in ml).

Section 5.2.3 Exploratory Efficacy End points
Change from baseline to month 6 and 12 in FVC (in ml).

2. Approach to Handle Missing Data

We modify the approach to handle missing data for the primary end point using an alternative model that is valid under the missing at random mechanism and for which software are more readily available in the SAS system – namely, linear mixed effects model which is valid under MAR.

PROTOCOL:
Section 8.2.2 General Approach of the protocol states:
If data are missing at random, in the secondary analysis for the primary end point, we may fit a linear mixed model within a Bayesian framework. In a Bayesian framework, missing data are treated as an additional model parameter to be estimated during model fitting. Missing values are estimated through an iterative procedure by imputing them multiple times based on the data model established using the most current estimates of the remaining model parameters

SAP:
N/A

3. Primary Analysis of the Primary End point

We modify the approach originally specified for the primary analysis (analysis of covariance using last-observation carried forward) to a more contemporary approach using linear mixed effects models for repeated measures, which is valid under the missing at random assumptions.

PROTOCOL:
Section 8.2.4. Efficacy Analyses. Analysis of the Primary Efficacy End point states:
The primary efficacy end point is the change from baseline to month 12 in mRSS scores. For the primary analysis, changes in mRSS scores from baseline to month 12 will be compared in the two treatment groups using an ANCOVA model with terms for treatment group, duration of dcSSc disease (stratification factor) and baseline mRSS score. If the assumptions of this parametric model are not met, an alternative non-parametric model will be used. Given that the incorporation of escape therapy after month 3 is an indication of treatment failure, we will use a last-observation-carried-forward approach to reflect the impact of treatment on mRSS at the time just prior to escape therapy. Sensitivity analyses will be used to assess the impact of this approach.

SAP:
Section 10.1. Primary Efficacy Analysis states:
The primary efficacy end point is the change from baseline to month 12 in mRSS scores. The goal of the primary analysis will be to test the null hypothesis that the difference between treatments in the primary end point is zero. The analysis will be performed on the mITT population. For the primary analysis, we will use a linear mixed model[18] with month in the study (3, 6, 9 and 12) as the unit of analysis and the change from baseline in mRSS as the outcome, with terms for treatment group, month, the interaction of treatment group with month, baseline mRSS, and duration of dcSSc disease (stratification factor). Study participant will be treated as a random effect to account for both heterogeneity among participants and correlation among measurements taken on the same participant. An autoregressive (AR(1)) variance-covariance will be assumed. Given that the incorporation of escape therapy after month 3 is an indication of treatment failure, we censor primary end point data after initiation of escape therapy.

Predicted mean change from baseline to month 12 for an exemplary participant by treatment group will be provided, as well as the estimate of the treatment effect at month 12, adjusted for baseline covariates, and corresponding 95% confidence interval and p-value for the treatment effect. Parameter estimates will be calculated using restricted maximum likelihood (REML) methods, and the Kenward-Rogers method will be used to calculate the degrees of freedom.

4. **Secondary Analysis of the Primary End point**

We simplify the approach for the non-linear secondary analysis of the primary end point, eliminating the adjustment for escape therapy. We also include several other sensitivity analyses of the primary end point in the SAP that are not discussed in the protocol.

PROTOCOL:

Section 8.2.4. Efficacy Analyses. Analysis of the Primary Efficacy End point states:

If mRSS does not change linearly as a function of time-in-study, we will extend the linear mixed model to include: a polynomial of time-in-study (expressed as a fraction of a year) of an appropriate degree, the interaction between treatment and the polynomial of time-in-study, duration of dcSSc disease (stratification factor), patient specific random effects, and for time points following the beginning of the escape therapy, the interaction between treatment, escape therapy and the polynomial of time-in-study minus time-since-escape-therapy (both expressed as fractions of a year). In this case, to test whether there is a significant difference between the two groups in the way mRSS changes over time, we will simply test whether any of the coefficients in the interaction between treatment and the polynomial of time-in-study is significantly different from zero. Given the limited sample size of this phase 2 study, we will carefully assess to the fit of these longitudinal models.

SAP:

Section 10.2.1 Secondary Analyses of Primary Efficacy End point

- Analysis of the primary efficacy variable as described above will also be performed on the PP Population.
- Analysis of the primary efficacy variable as described above in the mITT Population, except that all primary endpoint data will be incorporated (i.e., no censoring after escape therapy).
- Analysis of the primary efficacy variables using time-in-study as a continuous variable (expressed as a fraction of a year) instead of the discrete version used in the primary analysis.
- If mRSS does not change linearly as a function of time-in-study, we will extend the linear mixed model to include: a polynomial of time-in-study (expressed a fraction of a year) of an appropriate degree, the interaction between treatment and the polynomial of time-in-study, duration of dcSSc disease (stratification factor), and patient-specific random effects, Given the limited sample size of this phase 2 study, we will carefully assess to the fit of these longitudinal models.
- Analysis of the primary efficacy variable as described above in the mITT Population, except that we will account for the possibility of escape therapy by adding a variable that indicates the time (in months) the escape therapy was added to the participant's randomized study medication. The model will account for escape therapy by including, at the time points following the beginning of the participant's escape therapy, the time escape therapy began and the interaction term between treatment and time escape therapy began. The model will generate adjusted estimates of change in the mRSS score from baseline for each treatment group and month. To test the impact of treatment on mean changes from baseline to month 12, a linear contrast of the effect of treatment and the interaction of treatment X month will be assessed at month 12. The model is summarized below.

Change in mRSS_{it} =

$b_0 + b_{1t} \times \text{TREATMENT}_i + b_2 \times \text{MONTH}_{it} + b_3 \times [\text{TREATMENT}_i * \text{MONTH}_{it}] +$
$b_4 \times \text{SSc\_DURATION}_i + b_5 \times \text{BL\_mRSS}_i +$ Fixed effects
$b_6 \times \text{ESCAPE}_{it} + b_7 \times [\text{TREATMENT}_i * \text{ESCAPE}_{it}] +$

$b_{oi} + b_{1i} \times \text{TREATMENT}_i + e_{it}$ Random effects

for i=1, …, n participants, t= 1, 2, 3, 4 (corresponding to 3, 6, 9, 12 months) and ESCAPE_{it} = the number of months after escape therapy started. We assume that residual errors e_{it} (t = 1, 2, 3, 4) for participant i are normally distributed with zero mean and 4x4 AR(1) variance-covariance matrix.

5. **Testing of Safety Outcomes**

We eliminate inferential statistics (hypothesis testing) and focus on descriptive statistics of safety outcomes, given the limited power to detect safety signals other than those of large magnitude.

PROTOCOL:

Section 8.2.3 Safety Analyses states:

The total number of adverse events of each grade occurring in the two treatment groups by month 12 will be compared using a Fisher's exact test. Poisson regression or comparable non-parametric methods will be used to compare the total number of serious adverse events during the double-blind 12-month period by treatment group. For laboratory and other safety parameters that are continuous, two-sample t-tests or Wilcoxon rank sum tests will be performed to compare the two treatment groups.

SAP:

N/A

6. **Analysis Populations**

There is an inconsistency in the protocol with respect to the definition of modified intention-to-treat population. The SAP uses the definition that is more closely aligned with the pure intention-to-treat approach.

PROTOCOL:

Protocol Synopsis states:

The main population for efficacy will be the modified intention-to-treat population (MITT), defined as all participants randomized and receive at least one dose of study drug.

Section 8.2.1 Analysis Populations states:

The main analysis set for efficacy will be the mITT population, defined as all participants randomized, receiving at least one dose of study medication, and having at least one post-baseline efficacy assessment

SAP:

Section 8.3.1. Modified Intention to Treat Population states:

The main analysis set for efficacy is the modified intention to treat (mITT) population, defined as all participants randomized and receiving at least one dose of study medication.

7. **Other Analyses**

Additional analyses of outcomes that were collected, but not explicitly noted in the protocol are included to provide further interpretation of study results.

PROTOCOL:

N/A

SAP:

See Section 12. Other Analyses.

15   References

1. Altman RD, Medsger TA, Jr., Bloch DA, Michel BA. Predictors of survival in systemic sclerosis (scleroderma). Arthritis and rheumatism. 1991;34(4):403-13. Epub 1991/04/11. PubMed PMID: 1901491.

2. Amjadi S, Maranian P, Furst DE, Clements PJ, Wong WK, Postlethwaite AE, Khanna PP, Khanna

D. Course of the modified Rodnan skin thickness score in systemic sclerosis clinical trials: analysis of three large multicenter, double-blind, randomized controlled trials. Arthritis and rheumatism. 2009;60(8):2490-8. Epub 2009/08/01. doi: 10.1002/art.24681. PubMed PMID: 19644851; PMCID: PMC2725229.

3. Furst DE, Khanna D, Mattucci-Cerinic M, Silman AJ, Merkel PA, Foeldvari I. Scleroderma--developing measures of response. The Journal of rheumatology. 2005;32(12):2477-80. Epub 2005/12/07. PubMed PMID: 16331791.

4. Khanna D, Merkel PA. Outcome measures in systemic sclerosis: an update on instruments and current research. Current rheumatology reports. 2007;9(2):151-7. Epub 2007/05/16. PubMed PMID: 17502046.

5. Khanna D, Brown KK, Clements PJ, Elashoff R, Furst DE, Goldin J, Seibold JR, Silver RM, Tashkin DP, Wells AU. Systemic sclerosis-associated interstitial lung disease-proposed recommendations for future randomized clinical trials. Clinical and experimental rheumatology. 2010;28(2 Suppl 58):S55-62. Epub 2010/09/02. PubMed PMID: 20576216.

6. Varga J. Scleroderma and Smads: dysfunctional Smad family dynamics culminating in fibrosis. Arthritis and rheumatism. 2002;46(7):1703-13. Epub 2002/07/19. doi: 10.1002/art.10413. PubMed PMID: 12124852.

7. Charles C, Clements P, Furst DE. Systemic sclerosis: hypothesis-driven treatment strategies. Lancet (London, England). 2006;367(9523):1683-91. Epub 2006/05/23. doi: 10.1016/s0140-6736(06)68737-0. PubMed PMID: 16714190.

8. Baroni SS, Santillo M, Bevilacqua F, Luchetti M, Spadoni T, Mancini M, Fraticelli P, Sambo P, Funaro A, Kazlauskas A, Avvedimento EV, Gabrielli A. Stimulatory autoantibodies to the PDGF receptor in systemic sclerosis. The New England journal of medicine. 2006;354(25):2667-76. Epub 2006/06/23. doi: 10.1056/NEJMoa052955. PubMed PMID: 16790699.

9. Varga J. Antifibrotic therapy in scleroderma: Extracellular or intracellular targeting of activated fibroblasts? Current rheumatology reports. 2004;6(2):164-70. doi: 10.1007/s11926-004-0062-8.

10. Clements P, Lachenbruch P, Siebold J, White B, Weiner S, Martin R, Weinstein A, Weisman M, Mayes M, Collier D, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. The Journal of rheumatology. 1995;22(7):1281-5. Epub 1995/07/01. PubMed PMID: 7562759.

11. Clements PJ, Hurwitz EL, Wong WK, Seibold JR, Mayes M, White B, Wigley F, Weisman M, Barr W, Moreland L, Medsger TA, Jr., Steen VD, Martin RW, Collier D, Weinstein A, Lally E, Varga J, Weiner SR, Andrews B, Abeles M, Furst DE. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. Arthritis and rheumatism. 2000;43(11):2445-54. Epub 2000/11/18. doi: 10.1002/1529-0131(200011)43:11<2445::aid-anr11>3.0.co;2-q. PubMed PMID: 11083267.

12. Steen VD, Medsger TA, Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. Arthritis and rheumatism. 1997;40(11):1984-91. Epub 1997/11/19. doi: 10.1002/1529-0131(199711)40:11<1984::AID-ART10>3.0.CO;2-R. PubMed PMID: 9365087.

13. Khanna D, Clements PJ, Furst DE, Korn JH, Ellman M, Rothfield N, Wigley FM, Moreland LW, Silver R, Kim YH, Steen VD, Firestein GS, Kavanaugh AF, Weisman M, Mayes MD, Collier D, Csuka ME, Simms R, Merkel PA, Medsger TA, Jr., Sanders ME, Maranian P, Seibold JR. Recombinant human relaxin in the treatment of systemic sclerosis with diffuse cutaneous involvement: a randomized, double-blind, placebo-controlled trial. Arthritis and rheumatism. 2009;60(4):1102-11. Epub 2009/04/01. doi: 10.1002/art.24380. PubMed PMID: 19333948; PMCID: PMC3711466.

14. Khanna D, Berrocal VJ, Giannini EH, Seibold JR, Merkel PA, Mayes MD, Baron M, Clements PJ, Steen V, Assassi S, Schiopu E, Phillips K, Simms RW, Allanore Y, Denton CP, Distler O, Johnson SR, Matucci-Cerinic M, Pope JE, Proudman SM, Siegel J, Wong WK, Wells AU, Furst DE. The American College of Rheumatology Provisional Composite Response Index for Clinical Trials in Early Diffuse Cutaneous Systemic Sclerosis. Arthritis & rheumatology (Hoboken, NJ). 2016;68(2):299-311. Epub 2016/01/26. doi: 10.1002/art.39501. PubMed PMID: 26808827; PMCID: PMC4826472.

15. Hankinson JL OJ, Fedan KB. Spirometric reference values from a sample of the general U.S.

population. Am J Respir Crit Care Med. 1999;159(1):179–87.

16. Little RJ, Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. Biometrics. 1996;52(1):98-111. Epub 1996/03/01. PubMed PMID: 8934587.

17. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons; 2004.

18. Molenberghs G, Verbeke G, SpringerLink. Models for discrete longitudinal data. New York ; London: Springer; 2005. xxii, 683 p. p.

16 Listing of Tables, Listings and Figures

See separate document.