

---

7-1-2012

## Systemic sclerosis disease modification clinical trials design: quo vadis?

Fabian A. Mendoza, MD  
*Thomas Jefferson University*

Lynette L Keyes-Elstein  
*NIH/National Institute of Allergy and Infectious Diseases*

Sergio A. Jimenez  
*Thomas Jefferson University*

Follow this and additional works at: <https://jdc.jefferson.edu/sclerodermaf>

 Part of the [Medical Sciences Commons](#)

[Let us know how access to this document benefits you](#)

---

### Recommended Citation

Mendoza, MD, Fabian A.; Keyes-Elstein, Lynette L; and Jimenez, Sergio A., "Systemic sclerosis disease modification clinical trials design: quo vadis?" (2012). *Scleroderma Center Faculty Papers*. Paper 7.

<https://jdc.jefferson.edu/sclerodermaf/7>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Scleroderma Center Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

## Systemic Sclerosis Disease Modification Clinical Trials Design: Quo Vadis?

Fabian A. Mendoza, MD<sup>1</sup>, Lynette L. Keyes-Elstein, DrPH<sup>2</sup> and Sergio A. Jimenez, MD<sup>1</sup>

<sup>1</sup>Jefferson Institute of Molecular Medicine and Scleroderma Center, Thomas Jefferson University

<sup>2</sup>Statistical and Clinical Coordinating Center for Autoimmune Disease Clinical Trials (SACCC-ADCT), NIH/NIAID

**Keywords:** systemic sclerosis, scleroderma, clinical trial design.

### Corresponding Author

Sergio A. Jimenez, M.D.

Jefferson Institute of Molecular Medicine

Thomas Jefferson University

233 S. 10<sup>th</sup> Street, Suite 509 BLSB

Philadelphia, PA 19107. USA.

T: 215-503-5042 / F: 215-923-4649

Sergio.Jimenez@jefferson.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/acr.21667

© 2012 American College of Rheumatology

Received: Jun 08, 2011; Revised: Nov 16, 2011; Accepted: Mar 09, 2012

**Abstract**

The purpose of this manuscript is to discuss relevant aspects of clinical trials for Systemic Sclerosis (SSc) and to identify important considerations for the design of SSc disease modification clinical trials. Placebo randomized controlled trials with appropriate identification of SSc patients with diffuse progressive SSc skin involvement of recent onset, along with a rescue strategy for patients with worsening lung and skin involvement are suggested. If change in skin thickening is a major outcome of the study, the selection of patients with recent onset of disease and a predetermined degree of skin involvement are crucial requirements. The trial duration should be of at least 12 months. Sample size calculations should consider differences that exceed the Minimal Important Difference. Other relevant trial designs and potential threats to study validity are also discussed. Previous SSc-disease modifying trials have been beset by high dropout rates. Analyses on the subset of subjects completing the trial or applying the last-observation-carried-forward approach can potentially lead to biased estimates and false conclusions. Strategies for retention of subjects should be included at the design stage and analyses to account for missing data should be performed.

### Significance and Innovation

- There is an urgent unmet need for the development of effective disease modifying therapies for Systemic Sclerosis (SSc).
- Most disease modification SSc clinical trials have failed to identify effective disease modifying therapeutic interventions.
- During the design phase of the study, predictors of disease worsening should be identified based on knowledge of the natural history of the organ system under study. Predictors should be considered during both the design and analysis phases of the study.
- Important considerations relevant to trial design are discussed and evaluated with the goal of enhancing clinical research to identify effective disease modifying therapies for SSc.

## Introduction

Since the first description of scleroderma treatment by Carlos Curzio in 1753 which included warm milk and vapor baths, bleeding from the foot, and oral administration of small doses of quicksilver (1), until the most recent clinical trials of anti-TGF- $\beta$  monoclonal antibodies and novel tyrosine kinase inhibitors (2-10), numerous medications have been employed for treating this condition. Despite intense efforts to develop effective therapies, systemic sclerosis (SSc) remains as the autoimmune rheumatic disease with the highest case-specific mortality. Furthermore, there is a general perception that no disease modifying therapeutic modalities are available and that any therapeutic approaches currently employed are clinically ineffective. Although numerous therapeutic agents have been suggested to be effective in uncontrolled clinical trials or in descriptive unblinded studies rigorous evaluation failed to confirm these suggestions.

The relative rarity of SSc, its clinical heterogeneity and the lack of quantitative and unbiased assessment tools to evaluate the effectiveness of therapeutic interventions have imposed serious limitations to the design and performance of informative clinical trials in this disease. Based on different pathophysiological concepts, diverse trial design approaches have been used to develop phase 1 and phase 2 trials for potential SSc disease modifying drugs (11). Despite these efforts, effective treatments for SSc remain elusive with no clear and widely accepted standard for treating its broad spectrum of clinical manifestations. The paucity of data showing effective disease modifying therapeutic interventions for SSc, our best wishes to help patients with this disease and the clamorous necessity for effective therapies, coupled with the expected increase of new trials in the near future render the limitations in the performance of controlled clinical trials an unavoidable obstacle that must be surmounted. Consequently, we will

focus this review towards establishing a framework and describing an approach to the design and development of disease modifying SSc clinical trials that may be useful for investigators designing and engaging in such trials.

It is important to emphasize that cutaneous SSc involvement constitutes a specific subset of the broad disease spectrum and that paradigms and clinical response expectations for other manifestations of the disease, such as pulmonary artery hypertension or SSc-related interstitial lung disease (ILD) should not be directly extrapolated from the observations related to skin involvement.

We are aware that satisfying all clinical trial requirements in terms of statistical correctness, ethical perspective and feasibility will be difficult for SSc disease modification trials. Thus, recognizing that it will not be possible to design a universally acceptable fixed clinical trial design, it is our purpose to discuss several crucial aspects of trial design that will need to be considered and addressed in the design of future SSc disease modification clinical trials.

### **Randomized Controlled Trials vs Single-Arm Open Label trials**

Single-arm open label trials are used to obtain preliminary data regarding the safety of a drug and may help to develop and refine the parameters for the outcome variables to be evaluated. In any single arm trial, observed improvements may not be attributable to study treatment. For example, if the outcome of interest is not an objective measure, both physicians and patients may believe they observe improvement that is not genuine. Measures based on rater interpretation, such as the modified Rodnan Skin Score (mRSS) or visual analog scales (VAS), are vulnerable to this type of observation bias. Another possible explanation for observed improvements not related to the study treatment could be the natural course of the disease. For

example, skin will soften spontaneously in some SSc patients. On the other hand, disease course could also obfuscate a positive treatment effect; patients may show no improvement during the trial, but disease would have progressed in the absence of the study treatment. In SSc, the heterogeneous clinical presentation and the variability in the natural course of the disease render single arm studies of limited value to explore effectiveness of any potential therapeutic agent. To illustrate how the variable course of the disease can affect treatment outcome, it should be remembered that Carlos Curzio's treatment of the scleroderma patient he described resulted in softening of her skin after 11 months of therapy.

The evaluation of safety data for a single arm trial can also lead to ambiguous findings. For a heterogeneous disease, such as SSc, that gives rise to a complex array of clinical manifestations impacting multiple organ systems, distinguishing adverse reactions related to the disease from those attributable to the therapeutic intervention can be a significant challenge. These considerations became quite apparent in a recent single-arm open study of imatinib mesylate in which it was extremely difficult to adjudicate with certainty numerous clinical manifestations to either the disease course or as adverse reactions to the therapeutic intervention (Robert Spiera, personal communication). However, it should be emphasized that there is an important role of single drug open label trials for the acquisition of preliminary information about safety as well as to obtain valuable mechanistic insights of the proposed intervention and to assist in the development and honing of outcome measures. Furthermore, these trials are substantially less difficult and less expensive to conduct since they have an easier recruitment and design.

A well-designed blinded randomized controlled clinical trial (RCT) can overcome the weaknesses outlined above for the single-arm study. In the discussion below, we review issues

associated with the design and analysis of RCTs as they apply to trials in SSc. We attempt to maintain our recommendations consistent with published guidance documents from the Food and Drug Administration (FDA) and the International Conference on Harmonization (ICH), which provide principles and methodology to ensure that the assessments of efficacy and safety are unbiased and scientifically valid.

### **Randomized Placebo Control Trials**

A key decision when designing a RCT is the choice of control group. Placebo-controlled trials are preferred when the objective is to assess efficacy and safety of the study treatment after accounting for the influences of disease course and physician or patient expectations. Placebo-controlled trials are considered ethical when no effective treatment is known or when the use of placebo does not expose the subjects to excessive risk of harm provided that subjects are informed of the risks. Even when use of placebo is deemed ethical, physicians and their patients may be reluctant to participate in a placebo-controlled study, which presents a problem for recruiting sufficient numbers of subjects for the trial. When use of placebo presents ethical or practical concerns, the study treatment can be compared to a prior existing therapy or standard of care. When the control is an active therapy, the objective of the trial is to demonstrate either non-inferiority or superiority of the study treatment compared to the active control, the choice likely depending on the perceived effectiveness of the active control. Regarding SSc, the most relevant question in this matter is whether there is an accepted standard of disease modifying treatment for these patients.

Although SSc is characterized by broad multi-systemic organ involvement, the involvement of skin in SSc is almost universal and constitutes a specific subset of the broad disease spectrum. At the present time, there is no universally accepted disease modifying

therapy for the cutaneous manifestations of the disease. Only a few studies have demonstrated a statistically significant impact on the mRSS. In the study described by Tashkin et al. (12), subjects with SSc-associated lung disease treated with 1 year of oral cyclophosphamide had significant improvement in mRSS at 12 months compared to subjects receiving placebo. However, owing to its adverse event profile, cyclophosphamide is not a preferred treatment option for skin disease in the absence of interstitial lung disease (ILD).

From the review of 8 RCT for diffuse skin disease in SSc, excluding trials for treatment of Raynaud's phenomena and SSc-associated ILD (2-9), only one study showed a statistically significant difference between treated and control groups. This was a small trial (n=29) employing methotrexate (MTX) as a disease modifying drug, which showed a statistically significant improvement over placebo (p=0.03) at 24 weeks on a composite response variable (defined by improvement in mRSS, DLCO, or VAS for general well being). The impact on mRSS alone was not significant (p=0.06) (5). The lack of significance of the effect on mRSS for this MTX trial and the failure of the other RCTs to find statistically significant treatment effects does not necessarily mean that all the treatments are ineffective. A failed study may have included too few subjects to detect a meaningful clinical difference, or may have been too short in duration for the treatment to impact the pathologic effects of the disease on the skin, or biased in some way against the treatment. For example, in a second RCT designed to evaluate MTX (n=71 randomized), analysis of 47 subjects who completed the study showed no statistical difference between MTX and placebo (8). A reanalysis of the data using a Bayesian approach and applying multiple imputations methods for missing data to reduce bias associated with loss of subjects suggested that subjects on MTX were more likely to have a favorable mRSS response than subjects on placebo. While this demonstrates that evaluation of a simple hypothesis

sometimes does not provide the entire picture and that a thoughtful analysis can glean information important to evaluating the full potential of a new therapy, the frequentist approach (i.e. hypothesis testing) is currently still the standard by which FDA will approve a drug for a labeling indication and by which clinical practice is changed. The example of MTX for treatment of SSc skin manifestations illustrates the importance of overcoming design challenges in order to obtain unambiguous results.

Another interesting example of a study that failed to find significant difference between treatment groups is the high dose versus low dose D-penicillamine trial (2). In this trial, a low dose of D-penicillamine was deemed by the investigators to be an appropriate control. Although this study has been widely quoted as supporting the concept that D-penicillamine is not an effective disease modifying therapy for SSc, this conclusion rests on the validity of the assumption that the low dose of D-penicillamine is not at a therapeutic or biologically active level. Recent research, however, demonstrating an extremely important role of extracellular matrix stiffness in the fibrotic process calls the validity of this assumption into question. Remarkable results from several recent studies (13-16) have shown that increased matrix stiffness resulting from lysyl oxydase 2 (LOX-2)-mediated cross-linking of collagen molecules and from interactions with specific integrins can result in a potent activation of profibrotic pathways and in increased production of collagen and other extracellular matrix proteins with worsening of the fibrotic process. These studies clearly indicate that disruption or blockage of collagen cross-linking would be expected to exert a potent antifibrotic effect. D-penicillamine is one of the most potent inhibitors of collagen cross-linking. These effects can last for prolonged periods of time depending on the turnover rates of the affected collagen molecules. Therefore, even very low doses of the drug may have quite potent anti-fibrotic effects mediated by the

modification of extracellular matrix stiffness. Based on this new information, a more appropriate conclusion is that the high dose versus low dose D-penicillamine study was not able to differentiate between the two interventions or determine whether either intervention was effective or ineffective. As such, rejection of D-penicillamine as a potential disease modifying agent for SSc may have been premature.

Given the absence of a universally accepted disease modifying treatment for skin manifestations, ethical concerns associated with designing a placebo controlled study are diminished. Furthermore, placebo controlled trials are essential to establishing the absolute benefit of a new treatment over the natural course of disease. Placebo-controlled trials, however, can be difficult to recruit especially with a disease like SSc where disease progression has serious and irreversible consequences. Physicians are eager to offer SSc patients a disease modifying therapy, even one that is not proven or universally accepted, if there is a reasonable expectation of benefit. In addition, owing to anxiety about the future progression of their disease, patients are often unwilling to accept randomization to placebo, and if they do consent they may be quick to leave the study in the absence of perceived benefit. Physician and patient reluctance to participate combined with the rarity of SSc can compromise the ability to complete an adequately powered blinded parallel-group placebo-controlled trial within a reasonable timeframe.

To address this issue, use of historical controls as a comparator rather than placebo controls has been employed in some studies to describe the natural course of the disease. Generally, this approach will not produce comparable treatment groups, and hence bias is not adequately controlled, and subsequent findings are inconclusive. While use of historical controls cannot be recommended in most situations, valid alternative placebo-controlled designs that

minimize exposure to placebo and/or provide a potentially active treatment to all subjects are worth consideration at the protocol development stage.

Since no single design will be satisfactory in all situations, several alternative placebo-controlled designs options are discussed below. Some are standard and well-known, others are progressive designs developed for trials in diseases other than SSc, but all should be considered for their potential to offer a way forward for clinical research on treatments for SSc. Among several alternative designs outlined in the 2001 FDA guidance on choice of controls, “add-on”, “early escape”, and “randomized withdrawal” designs are some standard options that might be considered for SSc trials (17). Each has advantages and disadvantages.

#### Add-on Design

For the add-on design, all subjects are randomized to either study treatment or placebo in combination with a background therapy on which all subjects are maintained. The background therapy would typically be a “standard of care”. If no proven or uniformly accepted standard of care therapy is available, a control therapy with reasonable expectation of benefit (e.g. MTX for cutaneous disease) might be appropriate given certain limitations (noted below). The advantage of this design is that all patients will receive some form of treatment, which might make the trial more acceptable to physicians and patients and easier to recruit, but there are several considerations. First, the effect of study drug alone cannot be assessed. If the combination of background-plus-study treatment proves to be better than background-plus-placebo, no definitive conclusions can be drawn about how the study treatment would fare as a monotherapy. Second, if the background therapy is effective, then it might be reasonable to assume that the difference between the background-plus-study treatment combination and background-plus-placebo combination would be smaller than the difference between study drug and placebo in the absence

of background therapy. Under this assumption, the add-on study would need to be larger than the parallel-group placebo-controlled trial to achieve the same level of power. Third, if the effectiveness of the control is not proven (e.g. MTX for cutaneous disease), and a clinically significant difference between groups is not detected, the study will be inconclusive i.e. both drugs might or might not be effective. Finally, this design is likely to be most useful in cases where the background and study therapies have different mechanisms of action, but the protocol development team should also be cognizant of potential adverse interactions between the therapies.

#### Early Escape Design

In the early escape design, randomized treatment (placebo or study treatment) is discontinued if the disease worsens or doesn't improve by a pre-specified point in time. The advantage of this design is that subject exposure to an ineffective treatment is minimized. Responder status or time-to-failure endpoints are the most appropriate for this design. Since the consequences of SSc progression can be severe and irreversible, close monitoring for disease progression and implementation of plans for rescue therapy are necessary in every study on ethical grounds. However, it is important to recognize the limitations that institute of rescue therapy could pose on analysis of quantitative endpoints (see discussion below on Patient selection challenges and risk of progression).

#### Randomized withdrawal

In the randomized withdrawal trial, all subjects receive study drug during an open label phase of the study. Subjects who respond during the open-label phase are randomized to either stay on study treatment or withdraw. For example, SSc patients could be treated with open label study treatment for a predetermined period, for example of 9 months. After this period subjects

whose mRSS have remained stable or improved could be randomized to either stay on study treatment or withdraw. After another period of observation which would be of 6-9 months, if the subjects who remained on study treatment had significantly better improvement in mRSS than those who were withdrawn, the study treatment would be considered effective. One important consideration in this design is that if the effect of the study treatment is expected to continue on after withdrawal, then the observation period will have to be long enough to witness deterioration of the response in order to observe differences between study arms. Another consideration is that study treatment should be tapered slowly to minimize the risk of disease exacerbation associated with rapid withdrawal. A limitation of this design is that only the subset of the population that had a favorable response is included in the randomized portion of the study, so estimated treatment effects cannot be generalized to the population as a whole.

Over the last decade some innovative designs have been developed to improve efficiency, enhance recruitment, and address the ethical issue of equipoise when prior studies suggest a reasonable expectation of benefit. Some of these designs include the following: Three-stage design (18,19). Response Conditional cross-over (20), Placebo-phase design (21), and various Response-adaptive designs (22). These designs can allow smaller numbers of subjects required for the study, as well as, a reduction in the number of subjects in the placebo arm and have been successfully employed in studies of rare diseases. However, the relatively slow development of skin induration changes in SSc necessary to cause a meaningful modification of the mRSS as an outcome measure, the potential carry over effect of previous interventions, and the prolonged duration of disease modifying interventions for SSc may limit the potential application of these designs for SSc clinical trials.

### **Randomized Active Controlled Trials**

In situations where an effective therapy is available, a placebo-controlled trial may no longer be ethical. In SSc, some of the target organ manifestations such as renal crisis and pulmonary hypertension have highly effective therapy options. Marginal efficacy of cyclophosphamide has also been demonstrated in SSc-associated ILD. If available, an active control should be considered for randomized clinical trials when failure to treat could result in life threatening consequences or significant morbidity. For example, the currently ongoing autologous stem cell transplantation SCOT trial (and its mirror European study, ASTIS) use cyclophosphamide as an active control. Results of these trials are not known at present. It should be pointed out that in these two trials, the aggressive treatment scheme, the inclusion of patients with SSc-related alveolitis and the requirement of “life threatening disease” in their inclusion criteria clearly justify the use of a potent although potentially harmful active control. However, the well documented lack of durable effects of cyclophosphamide on SSc lung involvement and the substantial frequency of severe side effects suggest great caution in the use of this drug as a comparator active control. Obviously, standard therapy for other manifestations of the disease should be allowed in both groups.

### **Patient selection challenges and risk of progression**

To mitigate effects of the variable clinical course of SSc, and owing to the paradigm that early interventions are crucial for an effective SSc therapy most of the recent trials have been focused on subpopulations of patients with recent onset disease. Recent onset disease has been defined in different ways. A cut-off point of 18 to 24 months from the appearance of the first non-Raynaud’s SSc-involvement has been utilized in some studies and seems reasonable. However, to provide greater homogeneity regarding the skin involvement it may be appropriate

to define the onset of the disease as the appearance of clinically detectable skin induration to avoid problems that may be introduced, for example, by the presence of gastrointestinal symptoms such as gastroesophageal reflux. Furthermore, if skin thickening is a major outcome of the study, patients should have a substantial degree of skin involvement at the time of recruitment. For example, a requirement for subjects to have diffuse cutaneous SSc with mRSS >16 at enrollment will enable the study to demonstrate the effectiveness of the therapeutic intervention employing the currently available outcome tools.

Although the selection of recent onset and moderate to severe skin involvement is useful from the clinical and mechanistic perspective there is strong evidence to suggest that there is a substantial risk of rapid and even fatal disease progression in patients with rapidly progressive skin involvement, particularly related to worsening renal and pulmonary involvement (2-9). Given the substantial risk of potentially serious, irreversible or life-threatening consequences associated with severe disease progression, aggressive monitoring is required to detect rapid progression of skin disease, and to recognize early signs of renal disease or alveolitis. Criteria for use of salvage therapies for patients with worsening lung and skin involvement should be pre-specified in the protocol and are strongly recommended. However, choosing the drugs for salvage is difficult and represents another point of controversy. The recently published recommendations from EULAR/EUSTAR for management of SSc may assist investigators in selection of appropriate salvage therapies. One difficulty is that with the exception of the use of ACE inhibitors for renal crisis there are no available data conclusively demonstrating that the use of any drug can decrease SSc mortality. Pulmonary involvement has emerged as the most frequent cause of mortality in SSc, accounting currently for about 60% of all SSc-related deaths. Since cyclophosphamide has been shown to stabilize recent onset SSc-related alveolitis and ILD

(12), it is suggested as the rescue therapy for patients who develop clinically significant alveolitis and ILD. For patients with rapidly progressing skin involvement indicative of a dangerous progression of disease, there is no clear choice of rescue therapy. Given the modest effects observed for MTX in two RCTs, and positive evidence from uncontrolled trials suggesting possible effectiveness of mycophenolate mofetil (23-28), one of these two agents may be appropriate choices.

Although failure to provide appropriate salvage therapy in the event of severe disease progression would be unethical, implementation of this design feature poses significant challenges with respect to analysis of the data and interpretation of the study results. Whereas randomization helps to assure treatment groups are comparable at the start of the trial, salvage therapy, which will be implemented erratically and differentially across treatment arms, undermines comparability and weakens the ability of the trial to adequately assess the impact of study treatment on patient outcomes. If the “need for salvage therapy” is a component of the primary endpoint (e.g. response failure or time to failure), the impact of bias could be minimal providing criteria for implementing salvage therapies are pre-specified, clearly and objectively defined, and implemented consistently. On the other hand, numeric endpoints, for example mRSS, may be highly sensitive to bias if the time point for the primary assessment is after some subjects have received salvage therapy. Under this scenario, the potential for bias should be recognized during the design phase with strong consideration of (1) how often salvage therapy is likely to be needed in the population under study, (2) what baseline characteristics would be useful predictors of subjects who might go on to salvage, and, importantly, (3) how will the chosen salvage therapies impact the primary endpoint of interest. Consider the example where the primary endpoint is mRSS at a given time point, and mycophenolate mofetil is chosen as the

salvage therapy for rapidly progressing skin disease. Assuming that both study treatment and mycophenolate mofetil are effective in lowering the skin score, if salvage with mycophenolate mofetil is used more often in the placebo arm than in the study treatment arm, then the observed difference between placebo and study treatment arms will be smaller than would have been observed in the absence of the salvage therapy. That is, the estimated effect will be biased against the study treatment and statistical significance will be more difficult to achieve. Typical data analysis approaches including exclusion of subjects who have received salvage therapy, ignoring that some subjects have received salvage, and carrying forward the last observation prior to salvage will all yield biased results. If baseline predictors of need for salvage are available, an alternative might be to consider primary endpoint data for subjects receiving salvage as missing, and apply pre-specified approaches for handling missing data (see discussion below). Importantly, since no statistical technique can overcome the impact of bias with absolute certainty, the study development team should discuss the likely severity of the bias due to salvage and risk posed to valid inference. If the risk is too high, alternative endpoints or designs should be considered.

As noted above, criteria for implementing salvage therapies should be clearly and objectively defined, but these decisions are also challenging. For cutaneous disease, it is suggested that a change exceeding the Minimal Important Difference (MID) as it was defined based on several prior SSc clinical trials and described by Khanna et al. (29) should be utilized. Khanna et al. estimated that the MID for the mRSS improvement to be 3.2 to 5.3 (e.g. ~ 15-25% for an mRSS of 21), thus we suggest that a worsening of > 6 on the mRSS (e.g. ~30% worsening in skin involvement for a mRSS of 21) should trigger the institution of the salvage therapy arm. Similarly, for evaluation of pulmonary worsening and based on data from previous clinical trials

a 15% reduction in TLC or FVC accompanied by the demonstration of new onset of alveolitis or fibrotic lesions in a high resolution computerized axial tomography scan should indicate the institution of salvage therapy with cyclophosphamide. Based on the above mentioned considerations, a RCT design (placebo vs study drug) with salvage therapy for patients with worsening SSc pulmonary and skin involvement is suggested as shown in **Figure 1**.

Subjects meeting a criterion for salvage therapy should remain in the study and continue with all planned study procedures and assessments. Furthermore, subjects and study personnel should generally remain blinded to the study treatment assignment after salvage therapy is instituted. Per ICH E9 guidelines, breaking the blind for an individual subject should only be considered when management of patient care would vary depending on the treatment assignment. In contrast to these guidelines, however, an FDA guidance on safety reporting requirements that went into effect on March 2011 outlines different reporting requirements for study treatment and comparator arms, which results in unblinding the treatment assignment for any subject experiencing an unexpected serious adverse event. According to the new guidelines serious unexpected suspected adverse reactions (SUSARs), which by definition can only occur in the study treatment arm, are reported in an expedited manner, but serious unexpected adverse experiences occurring in the comparator arm are not to be reported on the expedited timeline. Since expedited reports for SUSARs are also sent to Institutional Review Boards, site personnel will be unblinded. In a study with few unexpected serious adverse events, the impact of unblinding the treatment assignment on a few subjects may be minimal. However, the scientific validity of the study could be jeopardized if the fraction of unblinded treatment assignments becomes too large. Study teams should monitor this and consider consulting with the FDA on alternate reporting strategies should a concern arise.

### **Duration of the trial**

Owing to the slow turnover of collagen in the tissues, the trial requires a long observation period to detect clinically evident changes. For example, a recent study employing the tyrosine kinase inhibitor imatinib mesylate showed that improvement in mRSS was not detectable at 3 months and that a statistically significant improvement was observed at 6 months of treatment which continued until the termination of the study at 12 months (10). The same pattern was observed in a recent retrospective trial with mycophenolate mofetil showing statistical differences compared with a historical cohort only after 6 months (28). Furthermore, our own analysis of a cohort of recent onset rapidly progressive SSc patients treated with mycophenolate mofetil showed worsening mRSS skin scores during the first 6 months of treatment which was followed by statistically significant improvement in mRSS in the subsequent 6 months of the trial (unpublished observations). Thus, based on the experience from these studies as well as from other studies a 12 month intervention phase would be required as the minimal duration for a meaningful trial. Follow up at the end of the intervention is also advisable either as a part of the trial itself or as a continued extension.

### **Endpoints**

A comprehensive review of potential endpoints for the study of SSc is beyond the scope of this discussion, but would include a wide variety of time-to-event outcomes, responder definitions, quantitative measures associated with individual manifestations of the disease, quality of life measures, patient reported outcomes, and biomarkers. One endpoint used almost universally in SSc clinical trials is the mRSS, the only validated clinical measure of changes in the extent and severity of skin involvement in SSc (30-32). The mRSS is an improvement over the original description of the skin score (30), in terms of practical application and reduction of

interobserver variability and standard deviation. In a study of intra- and inter-rater variation in the assessment of the mRSS in a diverse population of SSc patients with both diffuse and limited disease of variable severity, the coefficient of inter-rater variation (i.e. a measure of agreement between raters) was estimated at 25%. In contrast, the intra-rater coefficient of variation (i.e. within rater agreement) was 12%. These two findings taken together suggest that repeated mRSS assessments on an individual subject should be made by the same rater. If the mRSS is the primary endpoint for a longitudinal study, the logistics of retaining the same rater throughout the study should be considered by the protocol development team.

A significant challenge has been to identify endpoints that describe disease progression across multiple organ systems. An example is “event-free survival” (EFS), an endpoint of key interest in both the SCOT and ASTIS trials, where events are defined as death or significant organ failure. A second example, also from the SCOT study, is a new global composite rank score (GCRS) that reflects each subject’s “order” relative to every other subject in the study based on the following hierarchy of component outcome variables: death, failure of EFS, change in FVC, change in SHAQ, and change in mRSS. Although it has intuitive appeal, the GCRS will have to be evaluated to determine whether scores represent an accurate assessment of disease progression or improvement.

The development of new outcome measures is a priority in SSc clinical research. With more precise, sensitive, and quantitative outcome measures, clinical trials with fewer individuals enrolled and of shorter duration will be achievable.

## **Threats to Study Validity**

### *Drop-outs and Missing Data*

High dropout rates have been observed in past trials on disease modifying therapy in SSc. For example, 50% of subjects were lost in the D-penicillamine trial and 20% in the bovine collagen trial (2,6). Missing data due to subject withdrawal can result in biased estimates of the effect of the study treatment on both efficacy and safety endpoints and undermine the validity of study conclusions. A comprehensive review of methods for prevention and handling of missing data in clinical trials has been recently published by the National Research Council (33). Only a few highlights from these recommendations are repeated in the discussion below.

Importantly, statistical approaches to account for missing data in the analysis will always rely on unverifiable assumptions about the missing data mechanism. Hence, strategies designed to minimize missing data on key outcomes during the design and implementation phases of the study have more power to ensure study integrity than statistical techniques. The protocol design team should consider reasons why subjects might withdraw prematurely and consider potentially mitigating study design features. For example, SSc subjects may withdraw prematurely due to intolerance to the study treatment, lack of perceived treatment effect, or inconvenience attributed to long follow-up times or too many visits. To address withdrawal due to tolerability issues, the study design might include flexible dosing or include a run-in period to ensure tolerability prior to randomization. Options to minimize dropouts due to lack of perceived treatment effect include implementation of salvage therapies or considering a design that minimizes exposure to ineffective treatment of offers active treatment to most or all subjects. The problem of long follow-up times is challenging in SSc trials given the long time required to observe changes in fibrosis, but care should be taken to keep the number study visits to a minimum and avoid unnecessary data collection at each visit. While the study is actively following subjects, site personnel should be trained to understand the importance of working to keep subjects in the

study and to collect and report data. Subjects who are withdrawn from treatment should still be encouraged to continue to return for study visits and data collection. Sites should be adequately compensated for these efforts as an incentive to maintain vigilance in these efforts. Finally, data managers should monitor and query sites for missing data on a routine basis.

Despite best efforts, some degree of missing data will be present in nearly every trial. A pre-specified analysis plan to evaluate the study objectives should include information on all randomized subjects. This is consistent with the principles of an intent-to-treat (ITT) analysis and requires that missing data for randomized subjects that withdrew prior to assessment of the endpoint of interest be accounted for the analysis. This approach is in contrast to many published reports on SSc trials where analysis of data from “completers” is presented as the primary analysis. Analysis of completers is only unbiased in the rare situation where missing subject data is “missing completely at random” (MCAR), which means missingness is attributable to reasons completely unrelated to any study variable. For example, data may be missing due to equipment failure or withdrawal of the subject due to an out-of-town transfer away from the study clinical facilities. Furthermore, analysis of “completers” only also wastes valuable information that is available on withdrawn subjects.

Many statistical analysis tools are currently available that reduce bias associated with missing data and make use of available information on withdrawn subjects. These include single imputation methods, maximum likelihood, multiple imputation, Bayesian methods, and methods based on generalized estimating equations. The choice of method depends on assumptions regarding the missing data mechanism and underlying the methodology. Single imputation methods such as last-observation-carried-forward (LOCF) are commonly applied in SSc trials but rely on the often unrealistic assumption that the observation at the time of withdrawal would

not change if measured later. In the case where subjects outcomes are expected to worsen with time the LOCF approach can underestimate the treatment effect if a disproportionate number of subjects on placebo arm withdraw. Reasonable assumptions about missing data should be discussed with the study team prior to choosing the statistical approach. In addition, the team should consider baseline assessments and historical information that might be predictive of future withdrawal, which could be incorporated into the analytic plan. In particular, the skin thickness progression rate, which has been shown to be predict mortality and early internal organ involvement should be assessed at baseline for studies in diffuse SSc (34). Finally, sensitivity analyses are also important to assess the degree to which the estimated treatment effects rely on the assumptions.

#### *Inadequate power*

Another consequence of the premature withdrawal of subjects from a trial is that the subjects remaining are likely to be a subset of subjects with more stable disease or with greater improvement in disease who have not experienced intolerable toxicities. This subset no longer represents the target population. In addition, the outcome assessments in the placebo and study treatment arms are likely to be more similar than would have been expected if subjects had not withdrawn. As a consequence, statistical power for the test of the treatment effect is reduced.

A secondary consequence is in the design of future studies. In building a rationale for sample size estimates for a new trial, statisticians turn to published trials to extract estimates of responder rates and/or mean outcomes for the comparator arm along with estimates of the standard deviation. Extreme outcome values are likely to be under-represented among “completers”, which will bias estimated means and tend to reduce standard deviations. If the standard deviation estimate is too small, the resultant sample size estimate will also be too small.

Care must be taken when using estimates from a trial with a high dropout rate in order to ensure that the assumptions used in the design of a new trial are appropriate.

In addition to ensuring estimates derived from prior studies are appropriate, several other factors can impact the validity of sample size/power estimates and can, if ignored, undermine the success of the trial. First, the statistical test on which the sample size/power estimate is based must be consistent with the objectives and design of the study as well as the analysis planned for the primary endpoint upon completion of the trial. Sample size and power needs for key secondary endpoints should also be considered. If the analysis methods underlying the sample size/power calculations are inconsistent with the methods planned for analysis, the study may be under-powered to detect the treatment effect of interest or over-powered and larger than necessary.

Second, the goal in selecting the sample size is to ensure the study is adequately powered to detect a “clinically meaningful effect”. A “clinically meaningful effect” should not be confused with a “statistically significant effect”; a treatment group difference of ANY magnitude no matter how small can be statistically significant if the study is large enough. The objective is to design an efficient study that has a reasonable probability (i.e. power) of detecting a clinically meaningful difference between treatment groups should one in fact exist. Since there is not always consensus on the magnitude of the “clinically meaningful effect”; the protocol development team should consider this question during the design stage consulting literature and expert opinion as needed. This question has been considered by investigators in the field for studies where mRSS is of interest. A recent Delphi exercise showed that a reduction of at least 35% from the baseline is necessary to consider the effect clinically relevant and statistically

significant (35). This estimate is more generous than the MID calculation discussed above, but it may represent a more desirable goal in patients with severe skin disease.

Third, the adequately powered study must be feasible. Too often, site investigators will over-estimate how many subjects can be randomized within a reasonable time frame, and a large study will ultimately fail due to insufficient recruitment. Another ill-advised practice is choosing a feasible sample size first then selecting a “clinically significant effect” large enough to produce the desired sample size. The danger in this practice is that if the treatment effect for a new drug is smaller than that used for planning but still worthy of clinical consideration, the study may have insufficient power to detect the effect. Although compromise is often needed owing to limited resources, a better approach is to consider alternate design and endpoint options before making a final decision on how best to meet the study objectives.

### **Translational Research Considerations**

Given the clamorous necessity for the development of new disease modifying drugs for SSc, every effort should be made to address the potential mechanism of action of the proposed intervention. Skin biopsies can provide invaluable information with a low risk for the patient. Since many of the intracellular pathways involved in the expression of the fibrotic phenotype are not well defined in the context of the natural course of the disease and depending on the intervention proposed, it is suggested that both placebo and treatment groups should undergo skin biopsies. Analysis of these skin biopsies will allow the development and validation of more sensitive outcomes and the testing of potential therapeutic interventions in a shorter time and with a smaller number of subjects. The identification of serum biomarkers of disease activity should also become a high priority. Cytokine and autoantibody profiles from SSc patient samples

are valuable sources for biomarker candidate selection and subsequent testing in SSc and healthy populations.

Although the mRSS has been considered a validated outcome for SSc disease modifying trials, it is not quantitative and is prone to subjective variability. Therefore, substantial efforts should be devoted to the development and validation of more sensitive and quantitative tools such as, for example, quantification of ECM protein expression on skin biopsies before and after treatment (36) or novel approaches based on newer molecular technologies such as global gene expression analyses (37-39), proteomics (40), or kinomics (41). Proteomics, microarrays and kinomics are extremely powerful tools which have provided valuable mechanistic information, particularly in oncologic disorders, however, are still underused in SSc. Their inclusion in any SSc mechanistic studies will provide invaluable data to direct and guide future research in the field.

### **Acknowledgements**

The expert assistance of Melissa Bateman in the preparation of the manuscript is gratefully acknowledged. Supported in part of NIH grant AR 19616 to Sergio A. Jimenez.

## References

1. Rodnan GP, Benedek TG. An account of the study of Progressive Systemic Sclerosis. *Ann Int Med* 1962; 57: 305-318
2. Clements PJ, Furst DE, Wong WK, Mayes M, White B, Wigley F, et al. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis: analysis of a two-year, double-blind, randomized, controlled clinical trial. *Arthritis Rheum* 1999; 42:1194-1203.
3. Seibold JR, Korn JH, Simms R, Clemente PJ, Moreland LW, Mayes MD, et al., Recombinant human relaxin in the treatment of scleroderma. A randomized, double-blind, placebo-controlled trial, *Ann Intern Med* 2000; 132 :871–879
4. Black CM, Silman AJ, Herrick AI, Denton CP, Wilson H, Newman J, et al. Interferon-alpha does not improve outcome at one year in patients with diffuse cutaneous scleroderma: results of a randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 1999; 42:299-305.
5. van den Hoogen FH, Boerbooms AM, Swaak AJ, Rasker JJ, van Lier HJ, van de Putte LB. Comparison of methotrexate with placebo in the treatment of systemic sclerosis: a 24 week randomized double-blind trial, followed by a 24 week observational trial. *Br J Rheumatol* 1996; 35:364-372.
6. Postlethwaite AE, Wong WK, Clements P, Chatterjee S, Fessler BJ, Kang AH, et al. A multicenter, randomized, double-blind, placebo-controlled trial of oral type I collagen treatment in patients with diffuse cutaneous systemic sclerosis: I. oral type I collagen does not improve skin in all patients, but may improve skin in late-phase disease. *Arthritis Rheum* 2008; 58:1810-1822.
7. Clements PJ, Seibold JR, Furst DE, Mayes M, White B, Wigley F, et al. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis trial: lessons learned. *Semin Arthritis Rheum* 2004; 33:249-263.
8. Pope JE, Bellamy N, Seibold JR, Baron M, Ellman M, Carette S, et al. A Randomized control trial of Methotrexate versus placebo in early diffuse Scleroderma. *Arthritis Rheum* 2001; 44:1351-1358.
9. Denton CP, Merkel PA, Furst DE, Khanna D, Emery D, Hsu VM, et al. Recombinant human anti-transforming growth factor beta1 antibody therapy in systemic sclerosis: a multicenter, randomized, placebo-controlled phase I/II trial of CAT-192. *Arthritis Rheum* 2007; 56:323-333.
10. Spiera RF, Gordon JK, Mersten JN, Magro Cm, Metha M, Wildman HF, et al. Imatinib mesylate (Gleevec) in the treatment of cutaneous systemic sclerosis: results of a 1-year phase IIa, single-arm, open-label clinical trial. *Ann Rheum Dis* 2011. doi:10.1136/ard.2010.143974.

11. White B, Bauer EA, Goldsmith LA, Hochberg MC, Katz LM, Korn JH, et al. Guidelines for Clinical Trials in Systemic Sclerosis (Scleroderma): I. Disease-Modifying Interventions *Arthritis Rheum* 1995; 38:351-360.
12. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease, *N Engl J Med* 2006; 354 : 2655–66.
13. Wells RG. The role of matrix stiffness in hepatic stellate cell activation and liver fibrosis. *J Clin Gastroenterol.* 2005;39:S158-61.
14. Georges PC, Hui JJ, Gombos Z, McCormick ME, Wang AY, Uemura M, Mick R, Janney PA, Furthe EE, Wells RG. Increased stiffness of the rat liver precedes matrix deposition: implications for fibrosis. *Am J Physiol Gastrointest Liver Physiol.* 2007;293:G1147-54.
15. Lopez B, Gonzalez A, Hermida N, Valencia F, deTeresa E, Diez J. Role of lysyl oxidase in myocardial fibrosis: from basic science to clinical aspects. *Am J Physiol Heart Circ Physiol.* 2010;299:H1-9.
16. Liu F, Mih JD, Shea BS, Kho AT, Sharif AS, Tager AM, Tschumperlin DJ. Feedback amplification of fibrosis through matrix stiffening and COX-2 suppression. *J Cell Biol.* 2010;23:693-706.
17. Guidance for Industry. E10 Choice of Control Group and Related Issues in Clinical Trials. Rockville, MD: U.S. Food and Drug Administration; 2001.
18. Honkanen VEA, Siegel AF, Szalai JP, Berger V, Feldman BM and Siegel JN. A three-stage clinical trial design for rare disorders. *Statist. Med.* 2001;20:3009-3021.
19. Hull K. “Study Designs for Rare Diseases”, (2010) The 2<sup>nd</sup> Conference on Clinical Research for Rare Diseases (Online presentation).
20. Deng C, Hann K, Bril B, Dalakas MC, Donofrio P, van Doorn PA, Hartung HP, Merkies ISJ. Challenges of clinical trial design when there is lack of clinical equipoise: use of a response-conditional crossover design. *J Neurol.* 2011;Online:Aug 2011.
21. Feldman B, Wang E, Willan A, Szalai J.P. The randomized placebo-phase design for clinical trials. *J Clin Epidemiol* 2001;54:550–557.
22. Chow SC, Chang M. Adaptive design methods in clinical trials – a review. *Orphanet J of Rare Dis* 2008;3:11.
23. Nihtyanova SI, Brough GM, Black CM, Denton CP. Mycophenolate mofetil in diffuse cutaneous systemic sclerosis—a retrospective analysis. *Rheumatology (Oxford).* 2007;46:442-5.
24. Gerbiño AJ, Goss CH, Molitor JA. Effect of mycophenolate mofetil on pulmonary function in scleroderma-associated interstitial lung disease. *Chest.*2008;133:455-60.

25. Saketkoo LA, Espinoza LR. Experience of mycophenolate mofetil in 10 patients with autoimmune-related interstitial lung disease demonstrates promising effects. *Am J Med Sci*. 2009;337:329-35.
26. Derk CT, Grace E, Shenin M, Nalk M, Schulz S, Xiong W, et al. A prospective open-label study of mycophenolate mofetil for the treatment of diffuse systemic sclerosis. *Rheumatology (Oxford)*. 2009;48:1595-9.
27. Koutroumpas A, Ziogas A, Alexiou I, Barouta G, Sakkas Li, et al. Mycophenolate mofetil in systemic sclerosis-associated interstitial lung disease. *Clin Rheumatol*. 2010:1167-8.
28. Le EN, Wigley FM, Shah AA, Boin F, Hummers LK, et al. Long-term experience of mycophenolate mofetil for treatment of diffuse cutaneous systemic sclerosis. *Ann Rheum Dis*. 2011;70:1104-7.
29. Khanna D, Furst DE, Hays RD, Park GS, Wong WK, Seibold JR. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. *Ann Rheum Dis*. 2006; 65: 1325–1329.
30. Rodnan GP, Lipinski E, Luksick J. Skin thickness and collagen content in progressive systemic sclerosis and localized scleroderma. *Arthritis Rheum* 1979; 22:130-140.
31. Kahaleh MB, Sultany GL, Smith EA, Huffstutter JE, Loadholt CB, LeRoy EC. A modified scleroderma skin scoring method. *Clin Exp Rheumatol* 1986; 4:367-9.
32. Furst DE, Clements PJ, Steen VD, Medsger TA Jr, Masi AT, D'Angelo WA, et al. The modified Rodnan skin score is an accurate reflection of skin biopsy thickness in systemic sclerosis. *J Rheumatol*. 1998;25:84-8.
33. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC (2010). The National Academies Press.
34. Domsic, R.T., Rodrigues-Reyna, T., Lucas, M. Fertig, N., and Medsger, T.A. (2011) Skin thickness progression rate: a predictor of mortality and early internal organ involvement in diffuse scleroderma. *Ann Rheum Dis*. 2011;70:104-9.
35. Gazi H, Pope JE, Clements P, Medsger TA, Martin RW, Merkel PA, et al. Outcome Measurements in Scleroderma: Results from a Delphi Exercise. *J Rheum* 2007; 34:501-509.
36. Busquets J, Del Galdo F, Kissin EY, Jimenez SA. Assessment of tissue fibrosis in skin biopsies from patients with systemic sclerosis employing confocal laser scanning microscopy: an objective outcome measure for clinical trials? *Rheumatology*. 2010;49:1069-1075.

37. Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi JT, Pergamenschikou A, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. *Proc Natl Acad Sci U S A*. 2003;100:12319-24.
38. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, et al. Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS One*. 2008:e2696.
39. Sargent JL, Milano A, Bhattacharyya S, Varga J, Connolly MK, Chang HY, et al. A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. *J Invest Dermatol*. 2010;130:694-705.
40. Del Galdo F, Shaw MA, Jimenez SA. Proteomic analysis identification of a pattern of shared alterations in the secretome of dermal fibroblast from systemic sclerosis and nephrogenic systemic fibrosis. *Am J Pathol*. 2010;177:1638-46.
41. Aden N, Nuttall A, Shiwen X, de Winter P, Leask A, Black CM, et al. Epithelial cells promote fibroblast activation via IL-1alpha in systemic sclerosis. *J Invest Dermatol*. 2010;130:2191-200.

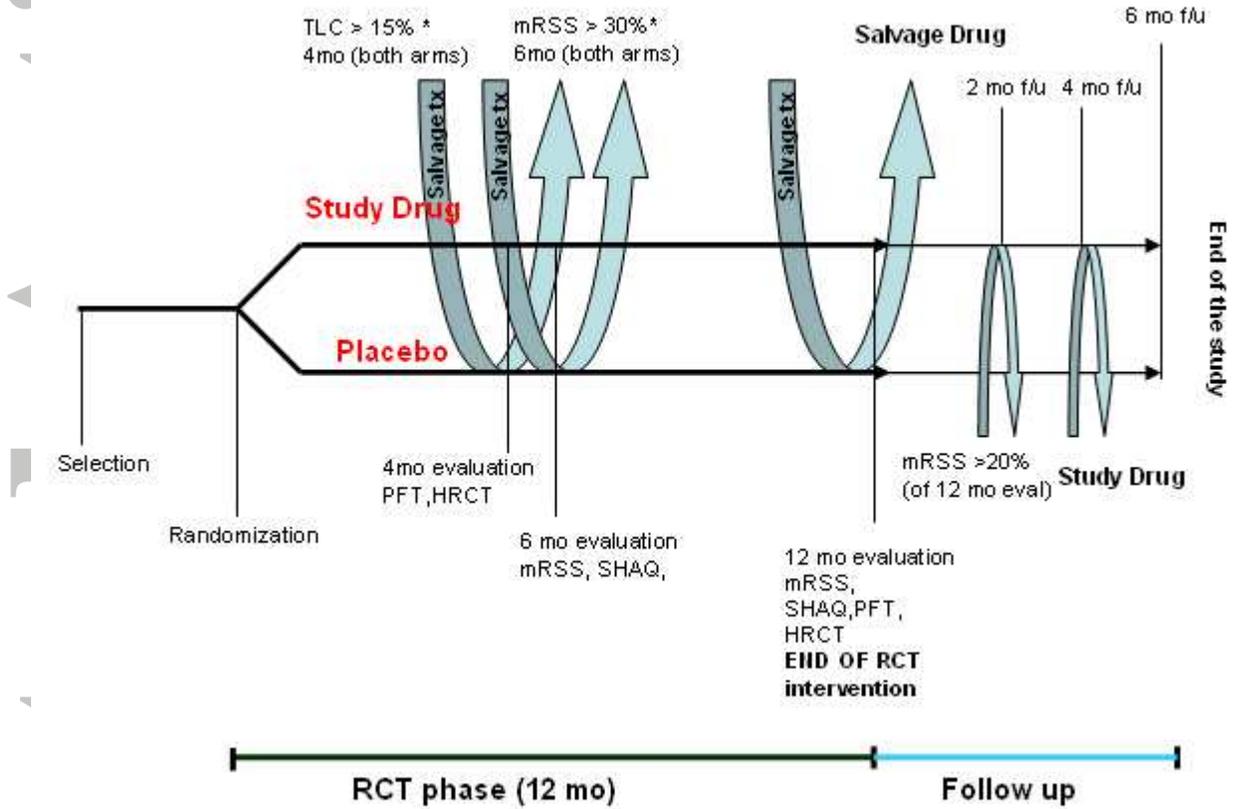
Accepted

**Figure 1. Schematic example of a SSc disease modification RCT design.**

Note that the trial includes salvage therapy arms for worsening skin involvement and deterioration of lung function caused by new onset alveolitis or progression of lung fibrosis. The study also includes a post-study follow up phase.

Accepted Article

Figure 1



Accep