Challenges in systemic sclerosis trial design

Christopher P. Denton
Professor of Experimental Rheumatology
Royal Free Hospital and University College London, UK

Correspondence to:

c.denton@ucl.ac.uk
Tel: 020 7794 0432
Abstract

Systemic sclerosis (scleroderma; SSc) is an autoimmune rheumatic disease with high clinical burden and unmet need due to connective tissue fibrosis and vascular damage. It has the highest case specific mortality of any rheumatic disease, with approximately half of patients diagnosed eventually dying as a direct result of SSc. There are no approved disease modifying treatments. This is partly related to the difficulty of conducting clinical trials for regulatory approval.

Traditionally skin thickness has been assessed using the modified Rodnan skin score (MRSS) that has been shown to correlate with survival and risk of complications in SSc. However recent trials have highlighted the limitations of MRSS which often improves over time, even on placebo.

A new composite measure integrating changes in multiple domains of lung function, skin, patients and physician global and HAQ disability index has been developed, the CRISS (Composite Response Index for Systemic Sclerosis). This measure looks promising and has provisional acceptance by American College of Rheumatology (designated ACR CRISS) but is unlikely to be strongly persuasive to Health Authorities in isolation unless there are also clinically meaningful changes in relevant domains that reflect how patients feel, function or survive.
Introduction

Systemic sclerosis is a multisystem disease with high burden and unmet need due to the limited efficacy of most current treatments that are directed more at symptom control or management of specific complications such as scleroderma renal crisis, pulmonary arterial hypertension of digital ulcers. The universal problems of skin fibrosis and musculoskeletal limitation and the common problem of interstitial lung disease have much fewer effective treatments available. In many cohorts, more than half of patients diagnosed with systemic sclerosis will ultimately die as a direct result of their illness as a result of serious cardio-respiratory, renal or gastrointestinal tract complications.

Clinical trial design for systemic sclerosis

Trial design has been challenging because of the multifaceted nature of the disease and the limitations in endpoints that can be reliably used in trials to confirm efficacy and the limitations in understanding the pathogenesis of SSc together with the complex underlying pathobiology and clinical diversity of the disease.

There have been advances in trial design that have emerged following several initiatives from experts to refine and improve study templates. An important advance was the publication of an article providing an overview of 22 points to consider in designing clinical trials in SSc [1]. This covered general approaches to study design and disease assessment but also highlighted the importance of specific trial templates for certain complications of SSc such as digital ulcers or Raynaud’s phenomenon. In some cases, it is possible to adapt study designs that have been developed for other organ-based diseases that occur as complications of SSc. One example is pulmonary arterial hypertension where basket trial designs have included a range of different causes of PAH and developed appropriate endpoints. It is notable that in the early period of PAH therapy trials where short term gains in exercise capacity were the main endpoint, especially the 6-minute walk test, that SSc cases often underperformed due to the disease impact on the endpoint that was independent of PAH, however it is reassuring that as more complex and clinically robust endpoints such as the time to clinical worsening emerge then SSc cases generally perform comparably to those with idiopathic PAH.

Digital ulcers have been assessed in clinical trials that have led to licensing of bosentan as a treatment for this specific complication of SSc, but endpoints remain a challenge because new ulcer rates in the timeframe of studies is much lower than in the first pivotal trials. This reflects better
background therapy and changes in the clinical trial population but is likely to explain why several recent trials of promising treatments have been negative. New endpoints and trial designs are being actively developed.

Other organ-based complications such as dyspepsia and indigestion or renal dysfunction are amenable to clinical trial design that is used in other types of disease although few studies have been undertaken despite the clear importance of these specific complications and the availability of potential therapies.

An important consideration for SSc trials is the stage of the clinical study. Phase I trials may evaluate new drugs or experimental approaches and likely focus on tolerability, pharmacokinetics (PK) and pharmacodynamic assessment of target engagement and mechanism of action. Such studies must always have a control arm receiving placebo so that there can be proper interpretation of the potential effect of therapy. This is critical because disease heterogeneity confounds the interpretation of uncontrolled studies.

The schematic in Figure 1 highlights some of the potential clinical trial designs that are appropriate for different phases of drug development. Usually at least one of the Phase II or Phase III trials will need to be dose ranging and so the overall study will need to be larger.

**Assessment of skin fibrosis**

Skin fibrosis is a hallmark of systemic sclerosis and defines the two major disease subsets. Diffuse cutaneous SSc (dcSSc) causes skin sclerosis proximal to the knees or elbows and is generally considered the most severe forms of SSc. Limited cutaneous SSc (lcSSc) has skin thickening confined to the face, neck and extremities. It is important to consider that any of the major complications of SSc can occur in either of the two main subsets although the frequency differs between these subsets. Generally, lung fibrosis and cardiac or renal involvement is more frequent in dcSSc. The link to organ-based complications underlies the different overall survival between subsets and is especially apparent in the first 5 years from disease onset defined by the first non-Raynaud’s SSc manifestation. It is acknowledged that this is a rather artificial as a timepoint for disease onset because vascular symptoms including Raynaud’s phenomenon often present as the first disease feature and the duration before other manifestations varies and is longer for lcSSc. However, the operational definition of first non-RP manifestations is practical for trial design and recruitment and allows some comparability between limited or diffuse subsets.
There is clear evidence of association between MRSS and outcome and this has underpinned its use as a trial endpoint for SSc. Thus, bad or progressive skin is associated with worse mortality, function and increased risk of major organ-based complications. However, the association is more complex and confounded by the natural history of MRSS that leads to significant improvement in at least half of all dcSSc cases over the first 3 to 5 years from disease onset [2]. This spontaneous improvement which may occur even more in cases receiving standard immunosuppressive therapy has been a major confounder in clinical trials because when spontaneous improvement occurs in the placebo arm it is hard to demonstrate treatment benefit. This has led to innovative methods to enrich trials for more progressive or informative cases as discussed below.

**Skin as an outcome measure in clinical trials – performance of MRSS**

The modified Rodnan skin score has emerged as a robust measure of skin involvement in SSc based upon studies that confirm association between MRSS and outcome including mortality and the risk of major internal organ involvement [3]. It has also been shown to be measurable in clinical trials and to have an acceptable intra-observer variability. The interobserver variability is more challenging but can be reduced with standardised training. There has been recent progress in developing SCTC and WSF endorsed standard methods for assessment of MRSS and for training in clinical trials [4]. However, one of the major challenges of MRSS is that there is generally improvement in the score at a group level over time and this confounds demonstration of treatment effect in clinical trials. The basis of this spontaneous improvement has been debated but it seems likely to reflect the natural history of skin fibrosis in that scars form and remodel over time and so it is clear that there is capacity for skin sclerosis to improve. It is likely that the extent of this improvement varies between individuals and contributes to the inherent diversity of skin score trajectory in SSc. Moreover, the intensity and nature of the profibrotic drivers of SSc skin disease are also likely to be relevant and this may differ for intrinsic subsets of SSc and reflect the diversity of profibrotic divers in SSc subsets and stages. This is likely to underpin the different trajectories of MRSS improvement that have been defined in observational cohorts. Thus, it has been shown that cases of early dcSSc can be categorised into milder cases that improve over time and cases with much more severe skin involvement. Some of which improve and others that do not. The latter represent a particularly important group clinically because not only do they have severe skin involvement but also a much higher mortality which suggests that failure to improve in the skin is reflective of failure to improve in more important internal organ manifestations of SSc.
One approach to MRSS for use in clinical trials has been to try to define better cases with progressive or regressive skin. In analysis of the EUSTAR cohort early disease and low baseline MRSS were identified as predictive of worsening. In the prospective ESOS dataset this was confirmed by ARA was also identified as predicting cases likely to be progressing at later stages or with higher baseline skin scores.

Group level changes in MRSS have been disappointing in recent trials because it has been hard to differentiate active treatment and placebo at a group level. This suggests that stratification of cases at baseline or se of a responder analysis might be more robust and this is being explored in current clinical trials.

**Development of skin biomarkers**

The challenge of clinical assessment of MRSS and growing understanding of likely pathogenic drivers and markers in skin in SSc has fuelled the development of candidate biomarkers of skin involvement. These include serum markets that have been correlated with MRSS in many cross-sectional studies but proven less robust in longitudinal analyses. Also, gene expression in skin biopsies has been studies [5]. The first 4 gene biomarker skin score was attractive as a measure of skin involvement that was altered by targeted therapies but in fact this did not perform well. Single gene markers have been informative, especially COMP and TSP. These have been incorporated into a 2 gene biomarker skin score that has been useful in open label studies and in one prospective clinical trial that suggested benefit for MRSS but where the biomarker response was earlier and much more robust that clinical change in MRSS.

Other recent studies have explored serum markers in SSc and associated changes with improvement in MRSS but the challenge of serum markers in a multifaceted disease is clear and new studies that define integrated markers or that examine subsets of patients without other major organ-based complications may be more informative.

**Bio-sampling to explore treatment mechanisms and target engagement**

One of the strengths of undertaking early stage clinical trials in SSc is the opportunity for bio sampling of skin or blood to better understand aspects of target engagement, mechanism of action or to determine efficacy using molecular pharmacodynamic markers. This has been achieved in several recent studies although interpretation has been limited by incomplete understanding of
pathogenesis and a lack of independent validation of candidate markers. One of the first robust examples of this approach was in the open label study of fresolumimab that showed improvement in MRSS but more persuasively there were changes in skin histology, notably the myofibroblasts score by immunostaining for aSMA, and gene expression markers of TGFbeta activation including TSP1 and COMP. In addition, the 2 gene skin score biomarker was significantly reduced agree treatment with anti TGFbeta monoclonal antibody. In a trial of the novel biological agent HICS there were changes in serum markers that reflected the likely mechanism of action by stimulating the hypothalamo-pituitary axis and upregulating αMSH which has antifibrotic activity in preclinical models [6]. There was a link between markers of fibrosis and clinical improvement in MRSS over 26 weeks in this clinical trial [7]. A particularly robust treatment effect was observed in molecular analysis of the phase II TCZ trial in dcSSc [8]. Here the 2 gene skin score biomarker achieved significance at 26 weeks. At the same time point there was almost complete reversal of the profibrotic fibroblast phenotype observed in tissue culture [9]. The availability of control fibroblast from patients receiving placebo made these data especially insightful and confirmed that IL6 blockade in vivo largely attenuated the TGFbeta activated phenotype of explant dcSSc fibroblasts [9]. Once trial that was conducted precisely to confirm target engagement and explore candidate biomarkers was a trial of an LPA1 antagonist that showed a trend of benefit for skin at a very early time point but highly significant impact on an LPA1 defined gene expression signature in skin biopsies [10].

**Lung assessment in clinical trials**

Lung fibrosis is an important complication of SSc and affects patients with both limited and diffuse cutaneous SSc. The timing and frequency of development of significant lung fibrosis is now well established and the impact of change in PFT and outcome has been defined. This it has been shown that short term changes in PFT can predict long term outcome including mortality. Studies have shown that change in FVC can best determined in clinical trials and this was best shown in two placebo-controlled studies of cyclophosphamide [11,12,13]. Recent trials have again shone the robustness of FVC as measure with significant differences observed for two placebo-controlled trials of tocilizumab [8]. It appears that changes can be shown in 6 months and the much more significant statistical differences compared with MRSS suggests that FVC might be one of the best single markers of disease progression treatment response in SSc.

The recently published trial of nintedanib in systemic sclerosis associated lung fibrosis was the largest placebo-controlled trial yet performed in the disease and met its primary end point by slowing the worsening of lung function test decline in a cohort of SSc-ILD [14]. Interestingly the trial
was stratified for the use of background immunosuppression and data showed that mycophenolate mofetil was also associated with a beneficial effect on lung function. For those patients on background MMF there was additional benefit seen from nintedanib compared with placebo suggesting that immunosuppression and antifibrotic therapy may work in complementary ways in this patient group. It is notable that although the primary endpoint was met there was no apparent benefit for skin score or any other lung parameter including patient reported outcomes, and the placebo subtracted difference modest at a group level being 1.2% predicted FVC decline preserved by nintedanib.

Other markers have been investigated including composite measures of FVC and DLco that could help define changes for clinical practice but have not yet been validated in clinical trials [15]. It is important to consider that not all SSC patients develop lung fibrosis but that both lcSSc and dcSSc can be informative. In addition, serum markers and ANA subsets have proven useful in identifying cases at most risk. Also, extent of disease seems to predict likelihood of future progression and this forms the basis of a simple staging system that is used in clinical practice and has been validated in multiple cohorts [16]. Other serum markers are also emerging. Together these tools offer great potential for future stratification in clinical trials design or for management.

**Composite endpoints**

Recent trial and other disease areas such as RA have that have highlighted the limitation of single endpoints in assessing treatment effect fuelled the development of composite endpoints. These have the advantage of measuring outcome over a range of domains and potentially having greater sensitivity and clinical validity as indices of overall treatment response.

The best developed endpoint is the CRISS that was developed by experts using recognised methods to identify important domains and feasible assessment tools [17]. Once the domains had been defined, they were tested in a prospective observational cohort and in available trial datasets. The limitation is that the trials were not generally regarded as positive and so the ability to define MCID and robustly develop the tool was limited. However, some recent trials have explored CRISS as a post hoc endpoint and in prospective designed analyses and the endpoint perform well in some studies although less well in others. Further refinement and development are likely to be needed but at present the CRISS has been given provisional approval by ACR. It is notable that in recent trials it has often outperformed MRSS over 48 months. Determination of MCID and development of new ways of assessing CRISS are ongoing.
HSCT trials redefine the “gold standard” for treatment response in clinical trials

The published major trials of HSCT reported have all suggested better long-term outcome and impressive short-term impact on skin or lung and functional scores in SSc [18,19]. This provides a new gold standard that has been hard to achieve with less intensive treatment strategies. This is the goal for future studies, and it is likely that combination treatments may be needed to achieve such good outcomes. However there are limitations to the available data, not least that the active comparator of high dose IV cyclophosphamide is well recognised to have high morbidity and toxicity and so the priority of HSCT over other approaches needs to be seen also in the context of good outcomes observed in recent cohorts using less intensive treatments.

Conclusions

In conclusion, there is more activity in clinical trials in SSc than ever and this reflects the exciting potential therapies that are emerging together with the high unmet medical need. Recent trials have been encouraging, especially for lung fibrosis endpoints [20], and it is important to distinguish between promising agents that differentiate active for placebo but just fail to reach primary endpoint significance and those trials that are unequivocally negative. This is important to advance the field and ensure that promising treatment strategies are not abandoned prematurely. It is a very realistic view that positive trials for skin and lung will soon be achieved and this will really advance practice and provide further impetus for good quality clinical research in SSc. It also seems likely that targeted therapies that tackle one pathway or mediator are likely to show response in particular stages and subsets and this needs to be considered when designing trials that will be interpreted at a group level. It also suggests that stratification of cases for trial recruitment and in clinical practice will ultimately become the usual approach.
Tables and Figures:

Table 1 Summary of endpoint data from recent placebo-controlled trials in systemic sclerosis

Figure 1 Schematic summarising overall clinical trial templates for systemic sclerosis.

This schematic illustrates some of the possible approaches for trial design in SSc. Definitive studies will need to be placebo controlled although background therapy maybe permitted. Endpoints can include skin, lung function or composite clinical endpoints such as CRISS. Molecular or biomarker endpoints may be considered earlier at 3-6 months linked to blood or skin biopsy studies although these will not assess clinical utility.
Reference list (max 20)


<table>
<thead>
<tr>
<th>Trial</th>
<th>Total (n)</th>
<th>Duration (months)</th>
<th>Drug</th>
<th>MRSS</th>
<th>Δ MRSS</th>
<th>FVC</th>
<th>CRISS</th>
<th>HAQ-DI</th>
<th>Global-patient</th>
<th>Global-physician</th>
</tr>
</thead>
<tbody>
<tr>
<td>faSScinate</td>
<td>86</td>
<td>48</td>
<td>tocilizumab</td>
<td>0.06*</td>
<td>-3.5</td>
<td>0.03</td>
<td>0.002</td>
<td>0.53</td>
<td>0.51</td>
<td>0.03</td>
</tr>
<tr>
<td>focuSSced</td>
<td>212</td>
<td>48</td>
<td>tocilizumab</td>
<td>0.1*</td>
<td>-1.73</td>
<td>0.002</td>
<td>0.02</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>ASSET</td>
<td>88</td>
<td>52</td>
<td>abatacept</td>
<td>0.28</td>
<td>-1.75</td>
<td>0.11</td>
<td>0.03</td>
<td>0.005</td>
<td>0.73</td>
<td>0.03</td>
</tr>
<tr>
<td>RISE-SSc</td>
<td>121</td>
<td>52</td>
<td>riociguat</td>
<td>0.08</td>
<td>-2.34</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>JBT-101-SSc</td>
<td>38</td>
<td>4</td>
<td>lenabasum</td>
<td>0.085</td>
<td>-2.6</td>
<td>ns</td>
<td>0.04</td>
<td>0.03</td>
<td>0.1</td>
<td>0.02</td>
</tr>
<tr>
<td>FASST</td>
<td>145</td>
<td>48</td>
<td>lanifibranor</td>
<td>NS</td>
<td>+0.9</td>
<td>NS</td>
<td>NA</td>
<td>NA</td>
<td>0.08</td>
<td>NA</td>
</tr>
<tr>
<td>SENSCIS</td>
<td>580</td>
<td>52</td>
<td>Nintedanib</td>
<td>0.58</td>
<td>-0.2</td>
<td>0.035</td>
<td>NS</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*there was significant reduction in meaningful worsening of MRSS

NS not statistically significant

NA data not available at time of writing
Schematic summarising overall clinical trial templates for systemic sclerosis. This schematic illustrates some of the possible approaches for trial design in SSc. Definitive studies will need to be placebo controlled although background therapy maybe permitted. Endpoints can include skin, lung function or composite clinical endpoints such as CRISS. Molecular or biomarker endpoints may be considered earlier at 3-6 months linked to blood or skin biopsy studies although these will not assess clinical utility.