Systemic Sclerosis — Continuing Progress in Developing Clinical Measures of Response

DANIEL FURST, DINESH KHANNA, MARCO MATUCCI-CERINIC, PHILIP CLEMENTS, VIRGINIA STEEN, JANET POPE, PETER MERKEL, IVAN FOELDVARI, JAMES SEIBOLD, DAVID PITTROW, RICHARD POLISSON, and VIBEKE STRAND

ABSTRACT. Few randomized controlled trials (RCT) have shown a demonstrable treatment effect in systemic sclerosis (SSc), making it difficult to evaluate outcome measures in this disease indication. Results from recent RCT, including those evaluating cyclophosphamide for SSc interstitial lung disease and endothelin receptor antagonists for pulmonary hypertension, have allowed analysis of certain organ-specific endpoints using the OMERACT filter. An earlier metaanalysis established that skin score, measures of Raynaud's, pulmonary function tests, blood pressure, pain, Health Assessment Questionnaire, and Medical Outcomes Survey Short-Form 36 are validated outcome measures in SSc. At OMERACT 8, data regarding validation of high-resolution computed tomography of the lungs, 6-minute walk test, and patient reported outcomes in SSc were presented. A Delphi exercise to develop consensus regarding a combined set of noninvasive measures for pulmonary arterial hypertension (PAH) is under way. Given the protean nature of this illness and its multiorgan system involvement, a composite responder index may be preferable. Another Delphi exercise is designed to develop consensus regarding a combined SSc response index to be validated in future RCT. (J Rheumatol 2007;34:1194–200)

Key Indexing Terms:
OMERACT SYSTEMIC SCLEROSIS COMBINED RESPONSE
RENAL LUNG CARDIOVASCULAR

Systemic sclerosis (SSc) is a heterogeneous disorder affecting multiple organs that often leads to severe pathologic fibrosis, organ failure, and death. The pathogenesis of SSc involves a complex and poorly understood interaction of inflammatory and immunological reactivity, an obliterative vasculopathy, and an unregulated fibroblastic response. This unusual pathobiology, which waxes and wanes, makes it challenging to develop good outcome measures for randomized controlled trials (RCT) in SSc.

Outcome measurements from clinical trials have been outlined previously, including skin thickness score, measures of Raynaud's, forced vital capacity (FVC), diffusing capacity (DLCO), Health Assessment Questionnaire disability index (HAQ-DI), and Medical Outcomes Survey Short-Form 36 (SF-36), which have demonstrated statistically significant changes by treatment groups (Table 1)[1,2].

Progress has been made in further validating organ-specific outcomes in SSc particularly based on the Scleroderma Lung Study (SLS; N = 158) and a similar although smaller UK study, the Fibrosing Alveolitis in Scleroderma Trial (FAST; N = 45; Table 2). Certain outcome measures, such as mortality or time to organ involvement, are not feasible at this time, unless the clinical trial involves very high risk patients with rapidly progressive diffuse scleroderma.

This report discusses progress since OMERACT 7, and describes proposed methodology to assess the validation of organ-specific system outcome measures. The format of this document will be to review available outcome measures and their background, apply the OMERACT filter to the extent possible, and summarize.

From the University of California at Los Angeles, Los Angeles, California; University of Cincinnati, Cincinnati, Ohio, USA; University of Florence, Florence, Italy; Georgetown University, Washington, DC, USA; University of Western Ontario, London, ON, Canada; Boston University School of Medicine, Boston, Massachusetts, USA; Klinikum Eiber, Hamburg, Germany; University of Michigan, Ann Arbor, Michigan, USA; Technical University of Dresden, Dresden, Germany; Massachusetts General Hospital, Boston, MA; and Stanford University, Palo Alto, CA, USA.

Dr. Khanna is supported by the Arthritis and Scleroderma Foundations (Physician Scientist Development Award), the Scleroderma Foundation (New Investigator Award), and a National Institutes of Health BIRCHW Award (grant HD051953). Dr. Merkel is supported by a Mid-Career Clinical Investigator Award (NIAMS K24 AR2224-01A1).

D.E. Furst, MD, Professor of Rheumatology, University of California at Los Angeles; D. Khanna, MD, MS, Assistant Professor of Medicine, University of Cincinnati and VAMC, Cincinnati, Ohio; M. Matucci-Cerinic, MD, PhD, Professor of Rheumatology and Medicine, University of Florence; P.J. Clements, MD, MPH, Professor of Rheumatology, University of California at Los Angeles; V. Steen, MD, Professor of Medicine, Georgetown University; J.E. Pope, MD, MPH, FRCP, Professor of Medicine, University of Western Ontario; P.A. Merkel, MD, MPH, Associate Professor of Medicine, Boston University School of Medicine; I. Foeldvari, MD, Senior Consultant in Pediatric Rheumatology, Klinikum Eiber; J. Seibold, MD, Professor of Medicine, University of Michigan; D. Pittrow, MD, Associate Professor of Clinical Pharmacology, Technical University of Dresden; R.P. Polisson, MD, MHS, Associate Clinical Professor of Medicine, Massachusetts General Hospital/Harvard Medical School; V. Strand, MD, Adjunct Clinical Professor of Rheumatology and Immunology, Stanford University.

Address reprints requests to Prof. D.E. Furst, David Geffen School of Medicine at UCLA, 1000 Veteran Ave., Room 32-59, Los Angeles, CA 90025. E-mail: defurst@mednet.ucla.edu

Table 1. Fully validated measures of outcome in SSc[2].

System	Measure
Skin	Modified Rodnan Skin Score
Cardiopulmonary	Forced vital capacity/DLCO
	Right heart catheterization
	Congestive heart disesae by clinical examination
Vascular	Raynaud's condition score*
	Patient Raynaud's phenomenon activity (VAS)
	Physician Raynaud's phenomenon activity (VAS)
	Raynaud's phenomenon frequency
	Raynaud's phenomenon duration
	Patient digital ulcer activity (VAS)
	Physician digital ulcer count
Renal	Blood pressure
Patient pain/function	Pain VAS
	HAQ Disability Index
	SF-36

* The Raynaud's condition score (RCS) is a summary rating of patients' self-report of their disease activity. Using a 0–10 Likert scale patients document in a daily diary their assessment of the combination of severity, frequency, duration, and impact of their Raynaud's for that day. The reporting RCS is the mean of 2 weeks of daily ratings.

Cutaneous outcomes in SSc

Although the modified Rodnan Skin Score is a fully validated outcome measure in SSc, it has shown modest responsiveness in 3 RCT, and better measures of response in skin manifestations are still needed. Other methods of assessing skin disease, such as use of a mechanical durometer to measure skin hardness, ultrasound to measure skin thickness, and other devices to measure the viscoelastic properties of skin, are presently in development.

Raynaud's outcomes in SSc

Raynaud's phenomenon is a vasculospastic clinical syndrome resulting in characteristic blanching, cyanosis, and erythema of the digits distal to the metacarpals. In severe cases, it causes painful, nonhealing ulcers, and occasionally results in amputation. It is a very common clinical problem in SSc. Based on data from earlier SSc RCT, several outcome measures have been validated (See Table 1; discussed previously and not reviewed here).

Table 2. Randomized controlled trials in SSc.

	Duration	N	Study	Published	Comments
Bosentan vs placebo in SSc-ILD	48 weeks	164	Seibold[39]	No	1,2,3,4
Bovine type I collagen vs placebo in SSc	60 weeks	168	Postlethwaite[47]	No	1,2,3,4
Cyclophosphamide vs placebo in SSc alveolitis	52 weeks	162	Tashkin[37]	No	1,2,3,4,5 Dis dur ≥ 7 yrs
D-penicillamine vs placebo in diffuse SSc	104 weeks	134	Clements[34]	Yes	1,2,3,4 Dis dur < 2 yrs
Methotrexate vs placebo in diffuse SSc	48 weeks	27	van den Hoogen[40]	Yes	1,2,3,4 Dis dur < 3 yrs
Methotrexate vs placebo in diffuse SSc	48 weeks	72	Pope[41]	Yes	1,2,3,4 Dis dur < 3 yrs
Relaxin vs placebo in diffuse SSc	24 weeks	231	Seibold[31]	No	1,2,3,4 Dis dur < 4 yrs
Alpha interferon for diffuse SSc	48 weeks	35	Black[42]	Yes	1,2,3,4
Anti-TGF beta for diffuse SSc	24 weeks	45	Denton[43]	No	1,2,3,4 phase 1b
Iloprost for Raynaud's phenomenon	6 weeks	103	Black[44]	Yes	1,2,3,4
RAPIDS 1 and 2 Bosentan vs placebo for Raynaud's phenomenon	16–24 weeks	310	Korn/Seibold[46]	No	1,2,3,4,5

1: double-blind; 2: multi-center; 3: placebo controlled; 4: diffuse SSc; 5: limited SSc.

Table 3. Status of measures in addition to Table 1 that have been examined since OMERACT 7.

Measure	Truth: Face Validity	Truth: Content Validity	Reliability	Discrimination: Discriminant Validity	Construct Validity	Feasibility	Ready for Use in Clinical Trials
Creatinine clearance	Val	Val	Val	Not tested	Val	Val	Yes
Chest HRCT for ILD	Val	Val	Val	Val	PV	Val	Yes
6MWT for PAH	Val	Val	Val	Val	Val	Val	Yes
6MWT for ILD	Val	No	Val	No	PV	Val	No
HAQ-DI	Val	Val	Val	Val	Val	Val	-YV
SF-36	Val	Val	Val	Val	Val	Val	Yes
Durometer for skin disease	Val	Val	Val	PV	PV	Val	Yes

HRCT: high resolution CT; 6MWT: 6 minute walk test; HAQ-DI: Health Assessment Questionnaire-Disability Index; ILD: interstitial lung disease; SF-36: Short-form 36; Val: validated; PV: partially validated.

Gastrointestinal outcomes in SSc

The gastrointestinal (GI) manifestations of SSc are probably related to a vasculopathy of the vasa nervorum, with eventual progressive fibrosis of the esophagus, stomach, small intestine, and colon. GI outcome measures are presently under evaluation — a patient-oriented outcome has recently been developed by the members of the Scleroderma Clinical Trial Consortium[2a].

Renal outcomes in SSc

Background. "Scleroderma renal crisis" (SRC) is a vasculopathic process that leads to rapidly progressive renal failure. However, it is difficult to choose measurable biomarkers that predict or indicate SRC. To date, only changes in blood pressure have been explicitly validated as an outcome measure (Table 1). Even urine sediment is not validated for SRC or SSc trials.

Although many use serum creatinine as the most feasible measure of renal function, creatinine levels may be reduced if muscular mass decreases for any reason (including SSc), so evaluating renal function by serum creatinine concentration alone may lead to underestimates of renal involvement. Direct measurement of creatinine clearance (CrCl) resolves this problem but presents 2 major drawbacks: (1) the need to perform an accurate 24-hour urine collection, and (2) individual variability in tubular secretion of creatinine. Clinical equations involving variables that are easily collected in clinical trials have been evaluated with a goal of increased accuracy in estimating CrCl. The Modification of Diet in Renal Disease (MDRD) equation is thought to be more accurate than the more commonly used Cockcroft-Gault equation[35]. The MDRD equation was established from data obtained in patients with some degree of renal failure [mean glomerular filtration rate (GFR) of the MDRD population 39.8 ml/min], and this may explain its tendency to underestimate GFR in patients with no abnormality in GFR. It is recommended that this formula be used in patients with abnormal renal function, although it may underestimate true GFR in patients with normal renal function, and may not compensate completely for decreases in muscle mass.

OMERACT Filter

Truth. Thirty-five patients with SSc underwent ^{99M}Tc-DTPA radioisotopic measurement of GFR by V injection. Results were correlated with measured 24-hour ClCr and Cockcroft-Gault and MDRD equations. In these 35 patients, CrCl using 24-hour urine collections was closely correlated with GFR measured by radioisotopic methods (^{99m}Tc-DTPA; r = 0.8207) considered the gold standard in this study. The 24-hour urine CrCl correlated more closely with the DTPA GFR than the Cockcroft-Gault and MDRD equations — respectively, r = 0.6536 and r = 0.6089.

Discrimination. CrCl has not yet been demonstrated to have discriminant validity in SSc.

Short-term reproducibility of GFR estimate by CrCl is good, with a median change of 5.35% (total range 0–20%). CrCl corrected to the standard body surface area of 1.73 m^2 can be used to evaluate renal function in SSc patients irrespective of sex, age, and clinical subset[35,36].

Feasibility. The method is readily available and not expensive, suffering only from inaccuracy of urine collections.

Summary of renal outcomes in SSc

Despite their simplicity, serum creatinine measurement, 24-hour collections of urine for CrCl calculations, and formulaic estimates of GFR are not validated measures of renal function in SSc. Further research regarding the validity of other measures of renal function, including ^{99m}Tc-DTPA radioisotopic assessment of GFR, is needed.

Cardiopulmonary outcomes in SSc

General background. Pulmonary function tests (PFT), including carbon monoxide diffusing capacity (DLCO), high-resolution computed tomography (HRCT) scan of the lungs, 6-minute walking test (6MWT), dyspnea scores, and New York Heart Association (NYHA) Functional Classification, have all been proposed as outcome measures in SSc trials. Unfortunately, the pathophysiologic processes in SSc often involve both heart and lung and the close functional interrelationship between cardiac and pulmonary organ systems makes differentiating discrete clinical effects on myocardium, coronary circulation, pulmonary parenchyma, or pulmonary vasculature problematic. It was thus agreed at OMERACT 6 that cardiac and pulmonary organ systems should be evaluated together. However, it is recognized that some outcome measures may be appropriate for certain organ involvement [e.g., 6MWT for PAH; FVC for interstitial lung disease (ILD)].

Pulmonary function tests: FVC and DLCO

Background. While PFT have long been used in SSc, only FVC has been validated as an outcome measure in RCT. Two recent RCT in patients with SSc lung disease used FVC and DLCO as outcome measurements. While the FVC changed in these trials, DLCO did not. The lack of sensitivity to change of the DLCO may have been due to (1) true lack of responsiveness (low signal to noise ratio), (2) the relatively short duration of the trials (only 1 year), (3) lack of progression in placebo-treated SSc patients, and/or, (4) the modest effectiveness of the medications tested (oral and intravenous cyclophosphamide)[3].

High-resolution CT scans of the lungs

Background. The histology of SSc is thought to include nonspecific interstitial pulmonary (NSIP) disease (with relatively more inflammation and less fibrosis), usual interstitial pulmonary (UIP) disease characterized by fibrosis and scarring or some combination of these 2[4-7,37,38]. Latsi's review noted either NSIP, UIP, or both[4,5,37] correlation have been found

between histopathology and HRCT in idiopathic pulmonary fibrosis[5,6,38].

In other diseases (e.g., ankylosing spondylitis) correlations were found between lung HRCT scores and various components of PFT in lung transplant patients, and it has been used in numerous other diseases (e.g., idiopathic pulmonary fibrosis, Wegener's granulomatosis, and polymyositis)[9,10-19].

OMERACT Filter

Truth. HRCT appears to be useful in SSc idiopathic lung disease (SSc-ILD). Devenyi and Czirjak[20] demonstrated that 86% of 21 patients with diffuse and 57% of 70 with limited cutaneous SSc had abnormal HRCT. Weaknesses of these and similar studies include the lack of uniform technology and a semiquantitative scoring system. Diot, *et al* used a semiquantitative scoring system and examined the relationship between HRCT and PFT in pulmonary fibrosis in 52 SSc patients[12,21]. The specificity for PFT abnormalities was 0.83 and sensitivity was 0.6 with a positive predictive value of 0.82. Other data, however, have shown inconsistent results[15,22,23].

One small SSc series (15 patients) showed an excellent segmental correlation of ground-glass appearance on HRCT and alveolitis by bronchoalveolar lavage examining the right middle lobe and lingula, but did not show this correlation for lower lobes[7]. In contrast, there was a good correlation between fibrosis of the lower lobes on HRCT and alveolitis, but not for the middle lobe. Overall, better semiquantitative methods to measure alveolitis by HRCT are needed, and are being examined.

HRCT fibrosis scores correlated moderately with histopathologic fibrosis scores ($r = 0.53$, $p = 0.001$), and the HRCT ground-glass scores correlated somewhat with histopathologic inflammatory scores for each lobe ($r = 0.27$, $p = 0.03$).

In idiopathic pulmonary fibrosis (IPF), a randomized trial of interferon-γ (IFN-γ) in 315 patients demonstrated that HRCT correlated with the DLCO[24]. In multivariate analysis an increased fibrosis score predicted death ($p < 0.0001$) and increasing DLCO and IFN-γ predicted less death ($p \leq 0.04$).

Discrimination. The ability of the HRCT to detect change in SSc is controversial, with one 14-patient open study showing responsiveness[30] and one 21-patient open study not doing so. The reproducibility of HRCT scoring for IPF was examined by Kazerooni, *et al* in 25 patients with IPF[38]. Interobserver variability (assessed by kappa statistic) was good for ground-glass ($\kappa = 0.59$–0.81) and for fibrosis ($\kappa = 0.72$–0.83).

The 1-year, 158-patient multicenter Scleroderma Lung Study trial, comparing oral cyclophosphamide to placebo in patients with active SSc alveolitis, demonstrated good inter-reader agreement for determination of the absence or presence of pure ground-glass ($\kappa = 0.76$) and fibrosis ($\kappa = 0.74$), but only fair agreement for honeycombed cysts ($\kappa = 0.29$). The extent of fibrosis was significantly and negatively correlated with FVC (Pearson, $r = -0.22$; $p = 0.01$) and DLCO ($r = -0.42$; $p = 0.0001$)[39,40]. Whether cyclophosphamide will protect the lung from fibrosis on HRCT awaits further analysis of the HRCT performed at baseline and at 12 months in the SLS and FAST trials (unpublished data).

Feasibility. Thoracic HRCT are quite feasible, but they are expensive, and require consensus readings, limiting their sensibility in RCT. While somewhat cumbersome, they are being used[39,40] and thus have demonstrated feasibility.

Summary of HRCT in SSc

The truth and feasibility of pulmonary HRCT in SSc have been demonstrated. Discrimination has been shown in other diseases, but not yet in SSc.

6-Minute Walk Test (6MWT) in SSc PAH

Background. The 6MWT measures the ability to walk a certain distance in a 6-minute period and has been used extensively in cardiology and pulmonary RCT. Biologically and *en face* it should be a good test of cardiopulmonary and musculoskeletal function. The 6-MWT is presently the most widely used primary endpoint for studies investigating PAH (idiopathic and associated forms, including SSc and systemic lupus erythematosus), and is the only measure of exercise capacity accepted by the US Food and Drug Administration[24]. It has not been validated in patients with less severe PAH such as New York Heart Association/WHO classes I and II.

Miyamoto, *et al* showed that in patients with PAH, compared with other noninvasive studies (e.g., echocardiogram and brain natriuretic peptide), only the 6MWT correlated with survival ($p < 0.01$)[25]. In a systematic review of the use of the 6MWT in chronic heart failure by Olsson, *et al*[26], significant increases in 6MWT were observed in only 9 of 47 RCT, leading the authors to recommend against its use. In an observational study of arthroplasty in OA patients by Kennedy, *et al*[27], a subset of 21 patients was chosen for test-retest reliability: intraclass correlation coefficient was 0.94 (95% CI 0.88–0.98).

OMERACT Filter

Truth. In RCT in PAH, including about 20% patients with SSc, the 6MWT improved in the same direction as fatigue and dyspnea[22], with cardiac index[28], or correlated with changes in mean pulmonary artery pressure, cardiac index, pulmonary vascular resistance, and clinical worsening[29].

Discrimination. The 6MWT was able to separate active from placebo treatment in the SSc epoprostenol trial of PAH [36 m improvement (epoprostenol) vs 15 m decline (placebo)][22]. For sildenafil, a subset of 11%–16% of SSc patients was examined from 3 trials in PAH patients comprising fewer than 100 SSc patients overall, and showed discriminant ability (75 m on sildenafil vs 0 m on placebo; $p < 0.05$)[22].

Feasibility. With some training, the 6MWT has been successfully incorporated into RCT of SSc PAH, and can be used in other trials with little difficulty.

6MWT in SSc interstitial lung disease

Truth. In an observational cohort of 50 SSc patients, the Borg index (Borg dyspnea score: a validated score of an individual's perceived effort on exertion on a 0–10 scale) correlated with the 6MWT ($r = 0.59$, $p < 0.0001$), lung VAS ($r = 0.5$, $p < 0.0001$), and University of California San Diego dyspnea index ($r = 0.53$, $p < 0.001$). While no correlation with FVC was found, there was a low to moderate correlation with DLCO ($r = 0.4$, $p < 0.005$). It should be noted, however, that this was a group of patients specifically defined as high risk for PAH and probably did not represent typical SSc patients with ILD or mixed ILD/PAH (Steen, unpublished data, 2006). In a study of bosentan for SSc-ILD, the 6MWT was highly reproducible with a correlation of 0.95 on test-retest after 2 hours to 2 weeks in over 160 patients (Seibold, unpublished data, 2006).

Correlations in a study of bosentan versus placebo in SSc-ILD with pulmonary function tests (FVC, $r = 0.19$ $p < 0.02$; DLCO, $r = 0.06$, p = nonsignificant) and Borg dyspnea score were low ($r = 0.28$, $p < 0.0004$; Seibold, unpublished data, 2006).

Discrimination. The test-retest reliability of the 6MWT was examined in a recent 1-year multicenter double-blind RCT comparing bosentan versus placebo in 163 SSc-ILD patients without complicating PAH (J. Seibold, unpublished data). In a repeat set of 6MWT, performed between 2 hours and 4 weeks apart, the mean 6MWT was 396.6 m versus 400 m (Pearson correlation coefficient 0.95, $p < 0.0001$). The study was not positive so the responsiveness of the 6MWT in SSc-ILD could not be addressed.

Feasibility. As in PAH, the 6MWT can be easily incorporated into RCT in ILD.

Summary

The 6MWT has demonstrated truth and feasibility in SSc-PAH and SSc-ILD and is sensitive to change in SSc-PAH, but it has not yet been shown to be sensitive to change in SSc-ILD.

Noninvasive measures to define pulmonary arterial hypertension

General background. While right heart catheterization is a validated measure for evaluating pulmonary hypertension, less invasive measures of PAH are clearly desirable. After initial discussion among qualified experts, a 3-stage Delphi exercise is under way to develop consensus regarding the best combination of noninvasive measures most likely to correlate with right heart catheterization for measurement of PAH. During the first round, 41 rheumatologists, 23 pulmonologists, and 16 cardiologists identified 17 domains and more than 150 instruments in SSc-PAH.

Patient reported measures in SSc

Physical function and health-related quality of life

Background. Since OMERACT 7, the Health Assessment Questionnaire Disability Index (HAQ-DI) and SF-36 have been shown to meet the OMERACT filter in SSc, based on analyses from 3 RCT in limited and diffuse SSc: a 24-week trial of relaxin in early diffuse SSc, a 15-month trial administering oral type I collagen in diffuse SSc, and the 1-year Scleroderma Lung Study trial in limited and diffuse SSc[8,28-31].

OMERACT Filter

Truth. In the D-penicillamine study, HAQ-DI scores ≥ 1.0 at baseline predicted increased mortality at 4 years compared with patients with diffuse SSc with baseline scores < 1.0, indicating poorer physical function is associated with earlier mortality[36]. The D-penicillamine trial also showed that improvement of HAQ-DI over 2 years was associated with improvement in the skin score. In the SLS trial, physical component summary (PCS) scores of the SF-36 had at least moderate (defined as ≥ 0.40) Pearson correlation coefficients with measures of breathlessness and person's global disease activity — the Mahler Baseline Dyspnea Index (BDI), VAS, 0–100 mm for breathing, and patient global assessment (PGA; 0–100 mm) of disease activity ($r = 0.51$ for all, $p < 0.01$), for HAQ-DI a moderate correlation coefficient with BDI and PGA ($r = 0.46$, $p < 0.01$), and a lower correlation coefficient ($r = 0.33$, $p < 0.01$) with VAS breathing. SF-36 mental component summary (MCS) scores showed low but significant correlations with BDI, VAS breathing, and PGA ($r = 0.26$ to 0.31, $p < 0.05$). Coefficients of correlation were not significant between the SF-36 PCS, MCS scores, and HAQ-DI with %FVC and %DLCO predicted ($r \leq 0.16$).

The internal consistency/reliability for SF-36 and HAQ-DI were estimated in the relaxin study[31]. Cronbach's alpha for SF-36 domains and PCS/MCS scores ranged from 0.76 to 0.93 and for HAQ-DI from 0.69 to 0.91 — good to excellent for both instruments. Although the test-retest reliability for SF-36 and HAQ-DI has been acceptable in other rheumatologic diseases, no study has formally evaluated reliability in SSc.

Discrimination. Responsiveness was examined using effect size (ES) in the relaxin study[31]. SF-36 PCS score showed a larger magnitude of responsiveness (ES = 0.36) for worsening and improvement (ES = 0.59) in overall disease (by PGA) compared with HAQ-DI (ES = 0.33 for worsening; 0.30 for improvement). For disease-specific measures (such as skin score and %FVC predicted), the HAQ-DI demonstrated a larger magnitude of responsiveness for worsening, but not improvement, than the SF-36 PCS. In the SLS, both SF-36 PCS and MCS scores discriminated between the severity of breathlessness measured by BDI and VAS breathing, and between more versus less impairment in %DLCO and %FVC predicted, respectively. The HAQ-DI complemented the SF-36 by discriminating between patients with limited and diffuse disease ($p < 0.001$).

Minimally important difference (MID). Using investigator's assessment as an anchor (measured on a 7-item Likert scale

from much worse to much better), the MID for the HAQ-DI ranged from 0.10 to 0.14 units out of 3 in the D-penicillamine study[34]. In the same study, MID estimates for modified Rodnan Skin Score ranged from 3.2 to 5.3 units out of 51[34]. Since there is an inherent uncertainty in the MID estimates, ranges are preferred over a point estimate. Correlations with patient-reported changes in global disease activity and/or pain will be necessary to confirm the MID estimates for HAQ-DI.

Feasibility. Both instruments have been extensively utilized in RCT in other rheumatic diseases and are feasible for use in RCT in SSc.

Summary

The SF-36 and HAQ-DI have fulfilled the OMERACT filters of truth, discrimination, and feasibility and are ready for use in RCT in patients with diffuse SSc, although effect sizes are not high. Both the SF-36 and HAQ-DI are partially validated in patients with limited SSc based on analyses performed in the SLS trial. Future trials should incorporate preference-based health-related quality of life measures such as standard gamble, time tradeoff, EQ-SD, SF-6D (derived from the SF-36), Health Utilities Index-3, and the Quality of Well-Being Scale. These measures are useful for pharmacoeconomic evaluations of therapies that are increasingly important in determining healthcare utilization.

Combined SSc response index

A Delphi exercise to reach a consensus for a combined index of response in SSc is under way for use in 1-year multicenter controlled clinical SSc trials. Thirteen domains were identified, including skin, pulmonary, cardiac, combined cardiopulmonary, gastrointestinal, renal, musculoskeletal, Raynaud's and digital ulcers, quality of life and function, global health, biomarkers, and others. Fifty scleroderma experts suggested 212 items; the items were decreased to 177 after eliminating the redundant items and are being refined further.

The SSc Special Interest Group recommended the following research agenda.

Renal

(1) Corroborate 24-hour urine creatinine clearance versus the first morning urine, to be employed as a valid measure of renal function. (2) Test whether CrCl is an independent measure of renal outcome or survival in SSc (e.g., different than blood pressure changes).

High resolution computed tomography (HRCT)

(1) As a semiquantitative scoring system of fibrosis and ground-glass has already been partially validated in one RCT, it will be necessary to corroborate responsiveness of HRCT in another RCT.

6MWT research agenda

(1) Validate 6MWT in an SSc population with mixed ILD and PAH.

PAH Delphi research agenda

(1) Complete the Delphi exercise.
(2) Be cognizant of the need to include patient perspective (6 of 14 domains are patient-reported), assuring that patient input remains an important part of ongoing SSc efforts.
(3) Validate results of the Delphi exercise to develop a combination of noninvasive measures to supplement or replace right heart catheterization for assessment of PAH in RCT.

Combined index of SSc response–research agenda

(1) Complete the present Delphi exercise.
(2) Validate results of the Delphi exercise (in RCT or longterm observational studies).
(3) Explore organizing the domains within 3 physiological groupings (fibrotic, immune, vasculopathic) rather than body systems to see if this improves sensitivity to change.

REFERENCES

1. Pope JE, Bellamy N. Outcome measurement in scleroderma clinical trials. Semin Arthritis Rheum 1993;23:22-33.
2. Merkel PA, Clements PJ, Reveille JD, Suarez-Almazor ME, Valentini G, Furst DE. Current status of outcome measure development for clinical trials in systemic sclerosis. Report from OMERACT 6. J Rheumatol 2003;30:1630-47.

2a. Furst DE, Khanna D, Mattucci-Cerinic M, Silman AJ, Merkel PA, Foeldvari I; OMERACT 7 Special Interest Group. Scleroderma — developing measures of response. J Rheumatol 2005;32:2477-80.

3. Demedts M, Behr J, Buhl R, et al. High-dose acetylcysteine in idiopathic pulmonary fibrosis. N Engl J Med 2005;353:2229-42.
4. Fujita J, Yoshinouchi T, Ohtsuki Y, et al. Non-specific interstitial pneumonia as pulmonary involvement of systemic sclerosis. Ann Rheum Dis 2001;60:281-3.
5. Flaherty KR, Toews GB, Travis WD, et al. Clinical significance of histological classification of idiopathic interstitial pneumonia. Eur Respir J 2002;19:275-83.
6. Bonnefoy O, Ferretti G, Calaque O, et al. Serial chest CT findings in interstitial lung disease associated with polymyositis-dermatomyositis. Eur J Radiol 2004;49:235-44.
7. Clements PJ, Goldin JG, Kleerup EC, et al. Regional differences in bronchoalveolar lavage and thoracic high-resolution computed tomography results in dyspneic patients with systemic sclerosis. Arthritis Rheum 2004;50:1909-17.
8. Strand V, Crawford B. Improvements in health related quality of life in patients with systemic lupus erythematosus following sustained reductions in anti-dsDNA antibodies. Expert Rev Pharmacoeconomics Outcomes Res 2005;5:317-26.
9. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635-41.
10. Koldingsnes W, Jacobsen EA, Sildnes T, Hjalmarsen A, Nossent HC. Pulmonary function and high-resolution CT findings five years after disease onset in patients with Wegener's granulomatosis. Scand J Rheumatol 2005;34:220-8.
11. Komocsi A, Reuter M, Heller M, Murakozi H, Gross WL, Schnabel A. Active disease and residual damage in treated Wegener's granulomatosis: an observational study using pulmonary high-resolution computed tomography. Eur Radiol 2003;13:36-42.
12. Diot P, Palmer LB, Smaldone A, DeCelie-Germana J, Grimson R, Smaldone GC. RhDNase I aerosol deposition and related factors in cystic fibrosis. Am J Respir Crit Care Med 1997;156:1662-8.
13. Ziora D, Jastrzebski D, Lubina M, Wojdala A, Kozielski J. High-resolution computed tomography in hypersensitivity

pneumonitis — correlation with pulmonary function. Ann Agric Environ Med 2005;12:31-4.
14. Pignone A, Matucci-Cerinic M, Lombardi A, et al. High resolution computed tomography in systemic sclerosis. Real diagnostic utilities in the assessment of pulmonary involvement and comparison with other modalities of lung investigation. Clin Rheumatol 1992;11:465-72.
15. Morelli S, Barbieri C, Sgreccia A, et al. Relationship between cutaneous and pulmonary involvement in systemic sclerosis. J Rheumatol 1997;24:81-5.
16. Zaporozhan J, Ley S, Gast KK, et al. Functional analysis in single-lung transplant recipients: a comparative study of high-resolution CT, 3He-MRI, and pulmonary function tests. Chest 2004;125:173-81.
17. Altin R, Ozdolap S, Savranlar A, et al. Comparison of early and late pleuropulmonary findings of ankylosing spondylitis by high-resolution computed tomography and effects on patients' daily life. Clin Rheumatol 2005;24:22-8. Epub 2004 Jul 20.
18. El Maghraoui A, Chaouir S, Abid A, et al. Lung findings on thoracic high-resolution computed tomography in patients with ankylosing spondylitis. Correlations with disease duration, clinical findings and pulmonary function testing. Clin Rheumatol 2004;23:123-8.
19. Magkanas E, Voloudaki A, Bouros D, et al. Pulmonary sarcoidosis. Correlation of expiratory high-resolution CT findings with inspiratory patterns and pulmonary function tests. Acta Radiol 2001;42:494-501.
20. Devenyi K, Czirjak L. High resolution computed tomography for the evaluation of lung involvement in 101 patients with scleroderma. Clin Rheumatol 1995;14:633-40.
21. Griffiths B, Miles S, Moss H, Robertson R, Veale D, Emery P. Systemic sclerosis and interstitial lung disease: a pilot study using pulse intravenous methylprednisolone and cyclophosphamide to assess the effect on high resolution computed tomography scan and lung function. J Rheumatol 2002;29:2371-8.
22. Badesch DB, Tapson VF, McGoon MD, et al. Continuous intravenous epoprostenol for pulmonary hypertension due to the scleroderma spectrum of disease. A randomized, controlled trial. Ann Intern Med 2000;132:425-34.
23. Kowal-Bielecka O, Kowal K, Rojewska J, et al. Cyclophosphamide reduces neutrophilic alveolitis in patients with scleroderma lung disease: a retrospective analysis of serial bronchoalveolar lavage investigations. Ann Rheum Dis 2005;64:1343-6.
24. Lynch DA, David Godwin J, Safrin S, et al. High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. Am J Respir Crit Care Med 2005;172:488-93.
25. Miyamoto S, Nagaya N, Satoh T, et al. Clinical correlates and prognostic significance of six-minute walk test in patients with primary pulmonary hypertension. Comparison with cardiopulmonary exercise testing. Am J Respir Crit Care Med 2000;161 (2 Pt 1):487-92.
26. Olsson LG, Swedberg K, Ducharme A, et al. Atrial fibrillation and risk of clinical events in chronic heart failure with and without left ventricular systolic dysfunction: results from the Candesartan in Heart failure-Assessment of Reduction in Mortality and morbidity (CHARM) program. J Am Coll Cardiol 2006;47:1997-2004. Epub 2006 Apr 27.
27. Kennedy DM, Stratford PW, Hanna SE, Wessel J, Gollish JD. Modeling early recovery of physical function following hip and knee arthroplasty. BMC Musculoskelet Disord 2006;7:100.
28. McLaughlin VV, Hill N, Tapson VF, et al. Sitaxsentan improves 6MW in patients with pulmonary arterial hypertension related to connective tissue diseases [abstract]. Arthritis Rheum 2004;50 Suppl:S692.
29. Thumboo J, Fong KY, Ng TP, et al. Validation of the MOS SF-36 for quality of life assessment of patients with systemic lupus erythematosus in Singapore. J Rheumatol 1999;26:97-102.
30. Khanna D, Furst DE, Clements PJ, et al. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. J Rheumatol 2005;32:832-40.
31. Khanna D, Clements PJ, Furst DE, et al. Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with systemic sclerosis and active alveolitis: results from the Scleroderma Lung Study. Arthritis Rheum 2005;52:592-600.
32. Khanna D, Furst DE, Hays RD, et al. Minimally important difference in diffuse systemic sclerosis: Results from the D-Penicillamine Study. Ann Rheum Dis 2006;65:1325-9. Epub 2006 Mar 15.
33. Kingdon EJ, Knight CJ, Dustan K, et al. Calculated glomerular filtration rate is a useful screening tool to identify scleroderma patients with renal impairment. Rheumatology Oxford 2003; 42:26-33.
34. Clements PJ, Wong WK, Hurwitz EL, et al. The Disability Index of the Health Assessment Questionnaire is a predictor and correlate of outcome in the High-dose Versus Low-dose Penicillamine in Systemic Sclerosis Trial. Arthritis Rheum 2001;44:653-61.
35. Latsi PI, Wells AU. Evaulation and management of alveolitis and interstitial lung disease in scleroderma. Curr Opin Rheumatol 2003;15:748-55.
36. Kazerooni EA, Martinez FJ, Flint A, et al. Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: Correlation with pathologic scoring. AJR Am J Roentgenol 1997;169:977-83.
37. Tashkin DP, Clements P, Furst DE, et al, for SLS Investigators. The Scleroderma Lung Study (SLS) shows the beneficial effects of cyclophosphamide (CYC) over placebo (PL) in systemic sclerosis (SSc) patients with active alveolitis [abstract]. New Engl J Med 2006;52:S624.
38. Goldin J, Lynch D, Strollo R, et al, and the SLS Research Group. CT detection of alveolitis: is there inter-reader agreement? Am J Respir Crit Care Med 2007; (in press).
39. Seibold JR, Black CM, Denton CP, et al. Bosentan versus placebo in interstitial lung disease secondary to systemic sclerosis (SSc): The Build 2 Study. Proc Am Thorac Soc 2006;3:A243.
40. 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;35:364-72.
41. Pope JE, Bellamy N, Seibold JR, et al. A randomized, controlled trial of methotrexate versus placebo in early diffuse scleroderma. Arthritis Rheum 2001;44:1351-8.
42. Black CM, Silman AJ, Herrick AI, 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.
43. Denton CP, Merkel PA, Furst DE, et al. Cat-192 Study Group; Scleroderma Clinical Trials Consortium. Recombinant human anti-transforming growth factor beta 1 antibody therapy in systemic sclerosis: a multicenter, randomized, placebo-controlled phase I/II trial of CAT-192. Arthritis Rheum 2007;56:323-33.
44. Black CM, Halkier-Sorensen L, Belch JJ, et al. Oral iloprost in Raynaud's phenomenon secondary to systemic sclerosis: a multicentre, placebo-controlled, dose-comparison study. Br J Rheumatol 1998;37:952-60.
45. Korn JH, Mayes M, Matucci Cerinic M, et al. Digital ulcers in systemic sclerosis: prevention by treatment with bosentan, an oral endothelin receptor antagonist. Arthritis Rheum 2004;50:3985-93.
46. Seibold JR, Denton CP, Furst DE, et al. Bosentan prevents occurrence but does not speed healing of digital ulcers in patients with systemic sclerosis (SSc) (abstract). Arthritis Rheum 2005;52 Suppl:4057.
47. Postlethwait AE, Wong W, Ingles J, et al. Maximal T cell reactivity to type I collagen (CI) is present during the first 3 years of diffuse systemic sclerosis (SSc) (abstract). Arthritis Rheum 2005;52 Suppl:1209.