

Current Status of Outcome Measure Development for Clinical Trials in Systemic Sclerosis. Report from OMERACT 6

PETER A. MERKEL, PHILIP J. CLEMENTS, JOHN D. REVEILLE, MARIAE. SUAREZ-ALMAZOR, GABRIELE VALENTINI, and DANIELE. FURST

ABSTRACT. Systemic sclerosis (SSc), also known as scleroderma, is a complex, multisystem disease that results in severe morbidity, disability, and life-threatening complications. The wide range of biological systems clinically involved, and the large spectrum of disease activity and damage that occurs in this disease have limited clinical investigation of SSc, especially therapeutic trials. The OMERACT 6 Workshop on Systemic Sclerosis was conceived as a starting point for the process of assessing the current state of the science in outcome assessment in SSc and then setting a realistic research agenda and priorities for future work in this area. This article reviews the current status of assessment tools for research in SSc within the main fundamental domains of illness in SSc, including problems related to skin, pulmonary, cardiac, peripheral vascular (Raynaud's phenomenon and digital ulceration), renal, and musculoskeletal systems as well as health-related quality of life and physical functioning. The current level of validation of outcome measures in SSc is assessed according to the guidelines set forth by prior OMERACT meetings. (*J Rheumatol* 2003;30:1630–47)

Key Indexing Terms:

SCLERODERMA

OUTCOME MEASURES

VALIDATION

Systemic sclerosis (SSc), also known as scleroderma, is a complex disease involving increased vascular damage, immunological activity, inflammation, connective tissue matrix deposition, and fibrosis in multiple organ systems and various types of tissues. This diversity of pathologic processes and systemic involvement leads to a wide variety of clinical presentations among patients with SSc. Patients may have limited disease burden or may evolve to have extensive disease that may be organ and life-threatening and physically and emotionally devastating. Further, the various biological systems clinically involved in SSc may progress at different rates of disease. Identifying the large spectrum of disease activity and damage in SSc is a challenge to clinical investigators attempting to both study the natural

history of the disease and measure meaningful clinical outcomes in therapeutic trials.

The past several years have seen a markedly increased interest by pharmaceutical companies in conducting research and development into effective agents for SSc. These increased drug development efforts have been accompanied by improved study designs and productive international collaboration among major SSc treatment centers. These advances in SSc research need to be matched by improvements in the outcome measures available to evaluate new therapies. Prior attempts at outcome measure development have often been based on expert opinion. The need for data-driven validation of outcome selection is evident and mandatory as investigators, industry representatives, government regulators, and patients work together towards more effective treatment of SSc.

The OMERACT 6 Workshop on Systemic Sclerosis was conceived as a starting point for the international scleroderma research community to begin the process of assessing the current state of the science in outcome assessment in SSc and then setting a realistic research agenda and priorities for future work in this area. The focus of this OMERACT meeting was the examination of outcome or response measures for trials involving patients with early diffuse scleroderma (SSc). Members of the working committee prepared presentations in various domains of illness in SSc. These preliminary documents and presentations were shared and reworked by the group. The final presentations were reviewed by other OMERACT 6 partici-

From the Arthritis Center, Boston University School of Medicine, Boston, MA; the Department of Medicine, UCLA School of Medicine, Los Angeles, CA; Division of Rheumatology and Clinical Immunogenetics, University of Texas Health Science Center; Baylor College School of Medicine, Houston, TX, USA; Seconda Università degli Studi di Napoli, Napoli, Italy; and Department of Rheumatology, UCLAMedical School.

Dr. Merkel is supported by a Clinical Associate Physician Award from the National Institute of Arthritis, Musculoskeletal, and Skin Diseases.

P.A. Merkel, MD, MPH, the Arthritis Center, Sections of Rheumatology and Clinical Epidemiology, Boston University School of Medicine; P.J. Clements, MD, MPH, Department of Rheumatology, UCLAMedical School; J.D. Reveille, MD, Division of Rheumatology and Clinical Immunogenetics, University of Texas Health Science Center; M.E. Suarez-Almazor, MD, Baylor College School of Medicine; D.E. Furst, MD, Department of Rheumatology, UCLAMedical School.

Address reprint requests to Dr. P.A. Merkel, Arthritis Center, E5 Boston University School of Medicine, 715 Albany Street, Boston, MA 02118. E-mail: pmerkel@bu.edu

pants and important expert feedback was received from people outside scleroderma research.

This article reviews the current status of a number of SSc outcome assessments as based on published data. The domains of illness addressed include problems related to skin, pulmonary, cardiac, peripheral vascular (Raynaud's phenomenon and digital ulcerations), renal, and musculoskeletal systems as well as health-related quality of life and physical functioning. Although other areas of disease certainly exist, these represent the most common organ systems involved by SSc. The various domains of illness are discussed separately since trials are often aimed at only one organ system. There is also a need to develop measures of overall disease activity, and such efforts are under way as well.

Each section in this report outlines the current level of validation of outcome measures in SSc according to the guidelines set forth by prior OMERACT meetings¹. The goal is that each outcome measure in SSc meets the standards of the "OMERACT Filter"^{2,3} of truth (face, content, construct, and criterion validity), discrimination (reliability and sensitivity to change), and feasibility. Since construct validity is often difficult to assess, convergent and divergent validity are often evaluated instead. Although outcome measure development, including for SSc, often begins with expert opinion of logical and practical measures, OMERACT strives to be a data-driven process.

The measures in this article are each summarized and judged based mostly on the evidence of their validity and utility derived from research data. The quality of these data is also evaluated. Five components of validation are rated based on the currently available data, as follows: A measure is considered "validated" if proper data are available that confirm the particular component of validity. A measure is "partially validated" if there are some but not complete data available regarding a component of validity. A measure is deemed "not tested" if no data are available to evaluate the measure. A measure is considered "not valid" if proper data are available to test the component of validity, and the results show the measure not to meet appropriate criteria. Additionally, feasibility is subjectively rated as "excellent," "good," or "poor" based on the measure's ease of use, cost effectiveness, availability in different centers, and overall practicality.

Each section of this article ends with a discussion of suggested areas or directions for future research, based on priorities established by participants of OMERACT 6.

Outcome Measures for Skin Involvement in Systemic Sclerosis

Thickening of the dermis and subcutaneous tissues is one of the earliest clinical events in the skin in SSc. Initially the skin thickens secondary to edema and infiltration of excess collagen. In the years that follow, the skin softens or thins in many patients, particularly those who develop widespread

skin thickening. Several groups have shown that assessment of skin thickness by clinical palpation has a good correlation with the weights of uniform diameter skin core biopsies (convergent validity)^{4,5}. These and other data suggest that assessing skin thickness by clinical palpation by a trained observer is an accurate, noninvasive method for assessing skin thickness. Table 1 lists the main outcome measures investigated for skin disease in SSc and the current status of validation according to the OMERACT guidelines.

Skin scoring. Three methods for assessing skin involvement by clinical palpation have not only been proposed, but also have been at least partially validated: modified Rodnan skin (thickness) score (MRSS), Kahaleh skin (thickness) score and the UCLA skin (tethering) score⁶⁻¹⁰. All 3 methods for assessing thickness/tethering make clinical sense (face validity). Since all 3 assess multiple anatomic areas, they assess skin involvement in a global or total organ sense (content validity). Interobserver test-retest reliability has been quantified and found acceptable for all 3 methods, while intraobserver reliability has been quantified and found acceptable for the MRSS and the UCLA skin score^{6,8,10,11}. For example, the inter- and intraobserver coefficients of variation were shown to be 25% and 12% for the MRSS in one study¹⁰, and 8% and 6% for the UCLA skin score in its initial report⁷. All 3 methods have demonstrated sensitivity to change in longitudinal cohort or parallel randomized controlled studies (discriminant validity)^{7,12-18}. For example, the MRSS declined 5.4 units over 2 years from a baseline value of 20.4 in the penicillamine trial¹⁵; the UCLA skin score declined 3.8 units over 3 years from a baseline value of 13.1 units in the chlorambucil trial⁷; and the Kahaleh skin score declined 5.9 units over 10 months from a baseline value of 21.3 in the photopheresis trial¹³ (all changes $p < 0.05$). While all 3 skin scoring methods distinguish diffuse from limited SSc, only the MRSS and the UCLA skin score have been shown to be useful in predicting which patients are at risk of developing scleroderma renal crisis and early mortality (divergent validity)^{7,8,12,17,19}. The absolute value of, and the changes in, MRSS have been shown to correlate with the absolute value of, and changes in, other features of the disease including oral aperture, functional disability, hands spread, finger-to-palm distance, joint tenderness, and survival (convergent validity)^{17,20,21}. The other 2 skin score techniques have not been evaluated for convergent validity. Skin scoring requires training, and the interobserver variability is quite high. Validity has been established only when the same investigator assesses a subject throughout the trial.

Skin biopsies. The weights of the skin cores obtained by uniform diameter skin biopsies correlated well with skin thickness scoring in the 2 studies reported: correlation coefficient of 0.81 for Rodnan's original technique, which measured 5 degrees of thickness by clinical palpation and 0.55 for a more recent Rodnan version, which measured 4 degrees of thickness by clinical palpation^{4,5}. The mean

Table 1. Validation of outcome measures for skin disease in systemic sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
Modified Rodnan skin score ^{9-12, 15-17,20, 21}	V	V	V	V	V	Good	Yes
Kahaleh skin score ^{6, 13}	V	V	V	V	NT	Good	No
UCLAskin score ^{7, 8, 11, 18}	V	V	V	V	PV	Good	No
Punch core biopsy ^{4, 5}	V	PV	NT	NT	PV	Good	No
Ultrasound of skin thickness ²²	V	PV	PV	NT	PV	NT	No

V: validated; PV: partially validated; NT: not tested.

weights of the cores were lower in limited than in diffuse SSc (divergent validity). Since the technique has been applied only to forearms, it has not yet been validated in multiple skin areas (content validity). It has not been tested for convergent validity (except for the correlation with skin thickness scoring by clinical palpation), and test-retest studies have not been performed. Skin biopsy has face validity, is feasible, and can be learned and performed easily.

There has also been some work on quantifying collagen and other connective tissue matrix components in skin biopsies for use as a clinical trial outcome measure. However, these techniques are not standardized, nor have validation studies been done to date.

Ultrasound. Ultrasound (using 20 MHz frequency) has been studied in one report that demonstrated a difference between limited and diffuse SSc (divergent validity)²². Although 3 body locations were studied, other frequently affected body areas were not. Interobserver reliability was high and there was good correlation between local MRSS and ultrasound readings (accuracy and convergent validity). Ultrasound was not compared to other related outcome response measures or skin techniques other than the MRSS (convergent validity), but it does make clinical sense as a measure of skin thickness. Since the study was cross-sectional, no testing over time was reported. The technique is sophisticated and results came from one center only, therefore, feasibility remains a question.

Other skin assessment techniques. New techniques for assessment of skin disease in SSc are under investigation for use in clinical trials including measures of skin elasticity²³, durometer measurements of hardness²⁴, spectrophotometry²⁴, and serum markers of connective tissue metabolism. However, there are not enough data on these new methodologies to comment on their validity.

Conclusions and future directions of research. Of the 3 methods using clinical palpation to assess skin involvement, only the MRSS is ready for use in clinical trials as a fully validated outcome or response measure. At this time ultra-

sound, skin biopsy, or other investigational techniques cannot be considered validated outcome or response measures for clinical trials. The OMERACT 6 workshop determined that the future direction of research should include (1) methods to make skin scoring more precise and reliable (e.g., training, 2 blinded measurers); (2) a comparison of the Kahaleh, UCLA, and MRSS methods with respect to all aspects of validity; and (3) further development and validation of skin biopsies, ultrasound, and new techniques to measure skin involvement in SSc.

Outcome Measures for Lung Involvement in Systemic Sclerosis

Lung disease, which in SSc includes interstitial lung disease (ILD) and pulmonary artery hypertension (PHT) together, is now the leading cause of death in SSc^{25,26}. Infiltration of the interstitium of the interalveolar septae by inflammatory cells and excessive collagen are the processes that lead to ILD typical of SSc. These processes lead to: (1) stiffening of the lung tissues and resulting loss of vital capacity and (2) distortion of the lung architecture with resulting mismatch of functional alveoli and arterioles and decreased diffusing capacity for carbon monoxide (DLCO). As fibrosis progresses, traction bronchiectasis, honeycombing, and fibrotic strands become evident and fewer intact alveoli remain perfused. The more the lung is damaged, the more the patient notes impaired function, reduced quality of life, decreased ability to exercise, and shortened lifespan²⁶.

The pulmonary vasculature can also be adversely affected in SSc. Disease of the pulmonary vasculature may be the result of either or both of 2 processes: (1) a progressive, obliterative-obstructive vasculopathy (typical of SSc), which is characterized by intimal proliferation and adventitial scarring/fibrosis, and (2) infiltration of the interstitium by inflammatory cells and fibrosis, which causes destruction of alveoli and pulmonary vessels. The result is that the pulmonary vascular supply is diminished enough that the pulmonary vascular bed cannot expand to meet the increased cardiac output that exercise demands. The right side of the heart must exert ever greater force to push blood

through the lungs. This may give rise to PHT, which initially occurs primarily during exercise but eventually occurs even at rest²⁶.

Table 2 lists the main outcome measures investigated for lung disease in SSc and the current status of validation according to the OMERACT guidelines.

Vital capacity. Vital capacity is often used as a surrogate for distensibility of the lung, although chest wall weakness and poor patient effort may confound the interpretation. As the lung loses distensibility, the vital capacity declines. The validity of vital capacity as an outcome measure in SSc has been assessed in multiple studies^{25,27-47}.

Diffusing capacity for carbon monoxide. Factors that affect the alveoli as well as the pulmonary vasculature may influence (i.e., diminish) the DLCO, largely by increasing the mismatch between arteriolar blood supply and the alveolar airspaces. The validity of DLCO has been assessed and confirmed in SSc^{26-38,41,42,48-57}. Its reproducibility is not as good as that of the vital capacity and its variability is wider, especially if performed in laboratories that lack strict quality control of procedures⁵⁶. Because both interstitial and pulmonary vascular involvement may affect DLCO, its utility to assess interstitial or pulmonary vascular involvement separately may be confounded.

High-resolution chest computed tomography (HRCT). The hope for HRCT is that it can provide a diagnosis of inflammation noninvasively (“ground-glass opacification”) and that the degree and extent of inflammation and scarring can be quantified. HRCT has been assessed and its validity has been confirmed primarily in idiopathic ILD, and to a lesser extent in SSc^{27,29,36,37,50,52,58-62}. The technique is widely available and fairly easily learned. However, centers that are less experienced, less standardized, and use older equipment may demonstrate increased variability. Therefore, standardization of centers is absolutely essential. Although fibrosis

on HRCT has fair to good correlation with fibrosis as seen in pathologic lung specimens ($r = 0.53$), the correlation of “ground-glass opacifications” on HRCT with “inflammation” on lung specimens (the gold standard) was much lower (0.27)⁵⁸.

Bronchoalveolar lavage (BAL). The hope for BAL is that it can assess “inflammation” in the alveoli in a less invasive manner than an open lung biopsy. Its validity has been assessed and confirmed in idiopathic ILD and in SSc^{35,45,46,50,52,63-65}. When BAL has been repeated over time, the differential cell counts did not change consistently⁶⁶. Its reliability has not otherwise been assessed. The correlation of inflammatory findings from BAL (polymorphonuclear and eosinophilic leukocytes in particular) and histopathologic specimens from open lung biopsy was low ($r = 0.29$)⁶³. Although the technique of BAL is widely available, great care is essential for standardization in the performance of the BAL and the interpretation of cell differential counts, especially in multicenter trials.

Plain chest radiograph. Because of its superiority in detecting early interstitial disease (either inflammatory or fibrotic), HRCT has superseded plain chest radiography in the diagnosis and measurement of response in SSc lung disease^{12,28,36,43,52,67}.

Dyspnea indices. The dyspnea indices were initially developed for grading lung and heart disease from other causes, but they have recently been applied to SSc. These questionnaires ask the patient to determine what degree of physical activity leads to dyspnea (i.e., walking on the flat, climbing up a gently sloping hill, climbing 3 flights of stairs). The Borg index assesses dyspnea immediately after patients complete the 6 minute walk distance test. Other indices (i.e., Mahler, the pulmonary visual analog scale of the scleroderma Health Assessment Questionnaire) allow overall assessment of dyspnea with daily activities. The answers

Table 2. Validation of outcome measures for lung disease in systemic sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
FVC ^{25, 27-47}	V	V	V	V	V	Good	Yes
DLCO ^{26,27-38,41,42,48-57}	V	V	PV	V	V	Good	No
HRCT ^{27,29,36,37,50,52,58-62}	V	PV	PV	PV	PV	Good	No
BAL ^{35,45,46,50,52,63-66}	V	NT	NT	NT	PV	Good	No
Chest radiograph ^{36,52,67}	V	NV	PV	NV	PV	Good	No
Dyspnea indices ^{34,38,44,68-72}	V	PV	PV	PV	PV	Good	No
Exercise tests ^{27-30,32,33,37,38,43,44,47,49,68,72-74}	V	PV	PV	PV	PV	Fair	No
6-min walk distance test ^{32,34,69-71,73,74}	V	PV	V	PV	PV	Good	No
Right heart catheterization ^{55,69-71,73,75-78}	V	V	V	V	V	Good	Yes
2-D/M-mode echocardiogram ^{55,75}	V	NT	PV	NT	PV	Good	No

V: validated; PV: partially validated; NT: not tested; NV: not valid.

may be recorded on visual analog or Likert scales. Their validity has been partially assessed in ILD and in PHT^{34,38,44,68-72}. Dyspnea indices need to be further tested and validated in SSc, particularly in those SSc patients with ILD.

Exercise tests. Several exercise techniques have been developed to test patients with cardiac or pulmonary diseases, including SSc ILD and pulmonary vascular disease. The maximal cardiopulmonary stress test employs gradually increasing degrees of exercise (commonly on a bicycle ergometer) until the patient reaches their maximum exercise capacity, at which time measurements of the maximum O₂ consumption (Max VO₂), arterial O₂ saturation, and other variables are determined. Although reproducibility and reliability have been quite good in idiopathic pulmonary fibrosis and healthy controls (at least at individual investigating sites)^{43,44}, these have been less than adequately tested in SSc. Construct validity is excellent in idiopathic ILD but has not been well tested in the ILD of SSc^{27-30,32,33,37,38,49,68,72-74}. They have shown sensitivity to change in PHT (both worsening and improving) and in ILD (primarily worsening, with lesser degrees of improving)^{47,73}. Unless they are rigorously standardized, these techniques may have restricted use in multicenter studies. Whether they can be applied to the SSc population at large is not clear. Given that many SSc patients have musculoskeletal problems, joint contractures, fatigue, and deconditioning, exercise testing needs to be evaluated in SSc directly.

The 6 minute walk distance test. The 6 min walk distance test has been quite useful in the development, US Food and Drug Administration (FDA) approval, and marketing of drugs for treatment of PHT. Although originally developed for testing patients with congestive heart failure and “pulmonary” diseases, it has more recently found favor in the investigation of PHT. Because patients responded with improved distance walked in 6 min (sensitivity to change) in response to therapy, the test has been used as the primary outcome measure justifying FDA approval for epoprostenol and bosentan for SSc-related PHT⁶⁹⁻⁷¹. After patients have had 3 training walks, reproducibility-reliability measurements are good to excellent⁷⁴. It has good construct validity for PHT^{34,69-71,73,74}, but construct validity has not been well studied in the ILD of SSc. It has good face validity, is easily learned and used, and is readily available at investigational sites⁷⁴.

Right heart catheterization. Right heart catheterization (RHC) is the gold standard for measurement of pulmonary artery pressure, pulmonary vascular resistance, pulmonary capillary wedge pressure, right atrial pressure, and cardiac index^{31,45-47,49,51-55,69-71,73,75-78}. Its greatest benefits are in documenting PHT and in evaluating the response of PHT to therapy^{45-47,49,53,54,69-71,73,77,78}. RHC measurements have been assessed and their validity in pulmonary hypertension (both

primary and SSc-related) confirmed^{31,45-47,49,51,54,55,69-71,73,75,78}. Even though RHC is readily available, its use requires rigorous standardization among clinical centers and is an invasive technique.

Echocardiogram. Because RHC is invasive and not suited for repeated or frequent outpatient assessment of pulmonary artery pressures, the use of less invasive, more easily performed techniques for measuring such as echocardiography are being explored. In contrast to RHC, however, echocardiography cannot distinguish between elevated pulmonary artery pressures resulting from primary pulmonary vascular disease or from elevation secondary to left heart dysfunction⁵⁵. Because of technical shortcomings, a direct reading of pulmonary artery pressures may not be possible with echocardiography in upwards of 20–30% of patients⁷⁵. Although its accuracy has been tested against RHC (the gold standard)⁷⁵, its test-retest reliability has not been validated in SSc. The validity of the echocardiogram otherwise has been partially confirmed in measuring pulmonary artery pressures in SSc⁷⁵.

Conclusions and future directions of research. Of the several potential lung and pulmonary vascular response measures discussed, only forced vital capacity and RHC have been fully validated as response measures for clinical trials in early diffuse scleroderma. Current trials in lung disease in SSc will help further validate some other measures, including HRCT, echocardiography, and dyspnea indices. The OMERACT 6 workshop identified the following for the future direction of research: (1) consider combining cardiac and pulmonary components of SSc for outcome assessment; (2) determine if the 6 minute walk test, maximal and submaximal exercise tests, and/or other physical/functional measures should be used as surrogates for cardiopulmonary involvement in SSc; (3) investigate the utility and validity of newer tests of cardiopulmonary function in SSc such as positron emission tomography, magnetic resonance imaging, and magnetic resonance angiography; (4) complete the validation of DLCO, dyspnea index(es), BAL, maximum exercise testing, encouraged 6 minute walk test, and HRCT in SSc.

Outcome Measures for Heart Involvement in Systemic Sclerosis

Heart disease in systemic sclerosis (SSc) can be either primary or secondary, the latter being caused by lung or kidney involvement. Primary heart involvement in SSc occurs in the first 3 to 4 years from disease onset in patients with the diffuse cutaneous subset⁷⁹. It is symptomatic in a few patients in whom it can present either as pericardial disease (acute fibrinous pericarditis or moderate to large pericardial effusion) occurring in about 7% of SSc patients (small pericardial effusion being both asymptomatic and devoid of any clinical significance)^{80,81}, or as myocardial disease (small intramyocardial coronary artery involvement

and myocardial fibrosis), which can manifest as symptomatic arrhythmias (7%), congestive heart failure (3%), or sudden death (5%)⁷⁹.

Symptoms of heart disease in SSc including fatigue, dyspnea, palpitations, chest pain, dizziness, and syncope are nonspecific. In contrast, clinical signs of heart involvement, such as fixed splitting of the first and second heart sounds, pericardial rubs, cardiac enlargement, jugular venous distension, and peripheral edema, even if infrequent, are more specific. Tachycardia, however, may also be secondary to autonomic neuropathy and is less specific.

Symptomatic heart disease including acute pericarditis, moderate to large pericardial effusions, and congestive heart failure have construct and criterion validity and are sensitive to change, but occur in few patients and are, therefore, less useful for clinical trials¹. Table 3 lists the main outcome measures investigated for heart disease in SSc and the current status of validation according to the OMERACT guidelines.

Anatomic cardiac alterations. Cardiac blocks^{82,83}, fixed defects at perfusional scintigraphy⁸⁴, and videodensitometric alterations⁸⁵ strictly reflect damage and cannot be used as outcome measures in clinical trials¹. When coronary artery disease is excluded, reversible defects detected by perfusional scintigraphy could be suitable outcome measures since they have been shown to be sensitive to change⁸⁶, but the technique is not feasible in all centers.

Cardiac arrhythmias. Arrhythmias detected by Holter monitoring⁸⁷ must be investigated for their sensitivity to change.

Ejection fraction. Ejection fraction has recently been shown to change after prednisolone therapy⁸⁸. However, the meaning of an increase in an otherwise normal ejection fraction must be understood. The lack of an increase in ejection fraction during exercise^{84,89}, as well as left and right ventricular filling abnormalities⁸⁹⁻⁹², await further validation.

Conclusions and future directions of research. Outcome measures of cardiac disease in SSc that have been properly validated are currently limited to symptomatic congestive heart failure and pericardial disease, and these measures may not reflect the full spectrum of SSc-related heart disease. The severity of cardiac complications in SSc and their close links to pulmonary manifestations of SSc necessitate further development of better outcome measures. It may be most useful to start considering cardiopulmonary impairment as a single domain of illness in SSc and evaluate outcome measures to reflect this understanding.

Outcome Measures for Raynaud's Phenomenon and Digital Ulcers in Systemic Sclerosis

Raynaud's phenomenon (RP), vasospasm of digital arteries with associated cyanosis, blanching and reperfusion, and potential for ischemic digital ulcers and chronic permanent arteriopathy occur in more than 90% of patients with SSc. This section will specifically relate to the secondary form of RP associated with SSc but many of these ideas may be adaptable to studies of primary RP.

The clinical seriousness of RP in scleroderma should not be underestimated. Patients suffer not only annoying and

Table 3. Validation of outcome measures for heart disease in systemic sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
CHF clinical Exam ⁷⁹	V	V	V	V	V	Excellent	Yes
Blocks (ECG) ^{82,83}	V	PV	V	Not Valid	V	Excellent	No
Q waves (ECG) ^{82,83}	V	PV	V	Not Valid	V	Excellent	No
Pericardial disease (clinical exam, ECG, echocardiogram) ^{80,81}	V	PV	V	V	V	Good	Yes
Supraventricular tachycardia (Holter ECG) ⁸⁷	V	PV	V	NT	V	Good	No
Ejection fraction (echocardiogram) ⁸⁸	PV	PV	PV	V	V	Good	No
Reversible defects (scintigraphy) ⁸⁶	PV	PV	PV	V	V	Poor	No
Ejection fraction during exercise (scintigraphy; echocardiogram) ^{84,89}	PV	PV	NT	NT	V	Poor	No
Fixed defects (scintigraphy) ^{84,86}	V	PV	PV	Not Valid	V	Poor	No
Ventricular filling (scintigraphy; echocardiogram) ⁸⁹⁻⁹²	V	PV	PV	PV	PV	Good	No
Quantification of fibrosis (videodensitometry) ⁸⁵	V	PV	NT	Not Valid	V	Poor	No

CHF: congestive heart failure. V: validated; PV: partially validated; NT: not tested.

painful “attacks” of RP, but also longterm disability from resulting digital ulcerations, autoamputations, infections, and the psychological burden of the disease.

Research in RP has focused on (1) diagnostic testing to both establish a diagnosis of RP and to differentiate primary from secondary forms; and (2) measurements of biological phenomena associated with RP (blood flow, oxygen delivery, tissue integrity). Much of development of outcome measures for clinical trials of RP has either been empiric or derived from these 2 areas of investigation. It is important to determine whether specific studies were geared toward understanding the biology of RP or developing and validating outcome measures.

Developing outcome measures and clinical trial design in general for RP presents some interesting challenges: (1) RP is episodic; (2) pain, tingling, and numbness are each inherently subjective and thus require patient-based data collection; and (3) psychological effects and environmental factors must be considered. Similar problems are encountered when studying digital ulcers.

A wide variety of outcome measures have been investigated for use in clinical trials of RP and/or digital ulcers in SSc. These measures vary from simple patient global self-assessments to laser Doppler measurements. Some proposed outcome tools require special machinery and training and

may not be feasible for multicenter trials. Many measures have not been standardized or validated in a comprehensive fashion. Table 4 lists the main outcome measures investigated for RP and digital ulcers in SSc and the current status of validation according to the OMERACT guidelines.

The simplest and most appealing method of assessment of RP is merely counting the frequency and duration of attacks. However, what appears to be an easy task is actually complicated since attacks are intermittent, do not occur with a frequency that allows direct observation by an investigator, and are subjective. Similarly, the severity of attacks, even if measurable objectively, varies from attack to attack and requires extended observation to obtain accurate representation of the patient’s status.

Patient-completed assessments. To overcome these problems in quantifying the frequency, duration, and severity of RP attacks, investigators began to design methods of assessment that are patient-derived and could take into consideration extended time periods. The most widely used method is to provide study subjects with diaries in which they record the number and duration of RP attacks each day⁹³⁻¹⁰⁴. More recently, diaries have also included a severity scale (visual analog scale or Likert) that was either completed for each attack or recorded once daily as a summary rating. The best described and most tested version of this latter method is the

Table 4. Validation of outcome measures for Raynaud’s phenomenon (RP) and digital ulcerations (DU) in system sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
Raynaud’s phenomenon							
Raynaud’s condition score ^{98,99,104}	V	V	V	V	V	Good	Yes
Patient RP activity (VAS) ¹⁰⁴	V	V	V	V	V	Excellent	Yes
Physician RP activity (VAS) ¹⁰⁴	V	V	V	V	V	Excellent	Yes
Pain VAS HAQ ^{20,104}	V	V	V	V	V	Excellent	Yes
HAQ disability ^{20,104}	V	V	V	V	V	Excellent	Yes
RP attack frequency ¹⁰⁴	V	V	V	V	V	Good	Yes
RP attack duration ¹⁰⁴	V	V	V	V	V	Good	Yes
Nailfold capillary microscopy ^{102,105-108}	PV	PV	NV	PV	PV	Poor	No
Laser Doppler ^{100,108,116,117}	PV	PV	Not Valid	NT	PV	Poor	No
Infrared thermography ^{94,95,100,101,107-109,113-115}	V	PV	NT	PV	PV	Poor	No
Digital blood pressure ¹⁰⁸⁻¹¹²	PV	PV	Not Valid	PV	PV	Poor	No
Plethysmography cold challenge ^{103,108,109,113,118,119}	PV	PV	NT	PV	PV	Poor	No
Digital ulcerations							
DU count ^{93,94,96,98,99,103}	V	PV	PV	PV	V	Excellent	Yes
Patient DU activity (VAS) ¹⁰⁴	V	V	V	V	V	Excellent	Yes
Physician DU activity (VAS) ¹⁰⁴	V	V	V	V	V	Excellent	Yes
Pain VAS HAQ ¹⁰⁴	V	V	V	V	V	Excellent	Yes
HAQ disability ^{20,104}	V	V	V	V	V	Excellent	Yes
Color photos	PV	NT	NT	NT	NT	Good	No
DU dimension	PV	NT	NT	NT	NT	Good	No

V: validated; PV: partially validated; NV: not validated; NT: not tested.

Raynaud's Condition Score (RCS)^{98,99,104}. The RCS is a 0–10 integer scale patients mark after considering the frequency, duration and severity of their day's RP activity. Thus the RCS diary system collects the raw frequency and duration data and has the patient self-integrate these factors along with severity to produce a series of scalable measures. The mean daily RCS is calculated over a set time period, usually 7 or 14 days prior to a study visit.

Simple 10-centimeter visual analog scales (VAS) are another type of clinical outcome measure frequently used for trials in RP to assess RP activity^{96-99,104}. Both patient and physician scales for RP activity and severity have been studied and a patient-completed scale is incorporated into the Scleroderma Health Assessment Questionnaire (SHAQ)^{20,104}.

RP patient diaries with the RCS as well as patient and physician-completed VAS have recently been used in several large and important clinical trials of investigational drugs for RP^{98,99}. These measures have performed well in these trials. A recent comprehensive analysis was performed to specifically examine the degree and scope of validation of these measures for RP¹⁰⁴. This study used actual clinical trial data and showed the full validity of the RCS and various VAS, and concluded that the RCS was superior to the individual metrics for use in clinical trials. The RCS was shown to have face, content, criterion, discriminant, and construct validity. This study also further validated the SHAQ VAS for RP as well as the Health Assessment Questionnaire (HAQ) disability and pain scales for application to RP trials. Patients with RP suffer significant disability and pain as a result of the chronic damage from vascular disease, including digital ulcers. These investigators concluded that a small number of outcome measures are sufficient for RP trials and proposed a core set for adoption.

Digital blood flow assessments. Several methods of assessing digital blood flow have been applied to clinical research for RP. Nailfold capillary microscopy is a useful diagnostic tool in RP, but its utility as an outcome measure remains to be proven^{102,105-108}. There are several techniques available, including direct observation and digitalized video. Measures such as counts of dilated loops, capillary dimensions, and blood flow velocities are potential quantifiable outcomes, but none have been applied comprehensively enough to allow adoption into clinical trials. Digital arterial blood pressure measurements using hand and/or total body cooling have been used to measure flow dynamics in patients with RP in clinical trial settings, but standardization is inadequate and must be improved before this technique is accepted as valid¹⁰⁸⁻¹¹². Infrared thermography has been used to measure temperature changes in the skin of the hand as a surrogate of blood flow, often in combination with a cold water challenge^{94,95,100,101,107-109,113-115}. Thermography is complex, expensive, requires quite strict environmental conditions for testing, and is not well standardized. Several

methods of laser Doppler have been studied to measure digital blood flow in RP^{100,108,116,117}. However, many problems exist with laser Doppler regarding variability, reproducibility, complexity, and feasibility. Similarly Doppler ultrasound has been used in RP research but mostly for diagnostic purposes. Its use as a clinical trial tool is not established. Although plethysmographic methods of digital blood flow quantification have been proposed to measure the severity of RP, usually in combination with a cold water challenge, few data are available to judge its validity^{103,108,109,113,118,119}. Whether these various techniques to measure blood flow move from interesting tools to study pathophysiology to outcome measures for clinical trials depends on whether data are published or available to show their validity and on the development of technology that is less expensive and operator-dependent.

Digital ulcer assessments. Digital ulceration (DU) secondary to RP in SSc is another major source of morbidity for which development of proper clinical trial outcome measures is urgently needed. Because of the paucity of clinical trials for DU in RP and the lack of proven effective treatment options, there are fewer data upon which to base conclusions of test validity in DU compared to RP itself.

Measuring DU has mostly been done through simple counting of lesions. If study subjects are seen frequently enough to detect bidirectional change, this method may be useful^{93,94,96,98,99,103}. However, there are other aspects of DU besides quantity that merit measuring, such as severity, pain, disability, size, and duration. Further, differentiating ischemic ulcers from skin breakdown and pressure sores is not always simple for patients and even for experienced clinicians. Thus, both patient and physician assessments may be prone to error unless both groups are taught to properly identify lesions.

The same study that helped validate measures of RP based on trial data also showed the validity of patient assessment of DU activity by VAS, pain by VAS, and HAQ disability measurements for DU¹⁰⁴.

Other measures of DU disease activity for clinical trials include serial digital color photographs and direct measurement of DU sizes with calipers. Data to evaluate these measures' validity are not yet available.

Conclusions and future directions of research. The current status of outcome measures for clinical trials of RP and DU is quite good. Several fully validated and easy to use instruments are available to measure multiple domains of illness and impact for the peripheral vascular disease of SSc. These measures have already been incorporated into clinical trials and their availability has helped facilitate industry-sponsored trials and drug development programs.

Future research in outcome measure development in RP and DU will need to focus on measures and markers of biological vascular activity and further refinement of methods to assess digital ulcers.

Outcome Measures for Renal Disease in Systemic Sclerosis

Scleroderma renal crisis and renal failure, previously the leading causes of death in patients with systemic sclerosis (SSc)¹²⁰, have become survivable since the introduction of angiotensin-converting enzyme inhibitors, circa 1980^{120,121}. Table 5 lists the main outcome measures investigated for renal disease in SSc and the current status of validation according to the OMERACT guidelines.

Scleroderma renal crisis. In most patients, renal failure is due to scleroderma renal crisis, an entity characterized by abrupt onset of accelerated hypertension associated with extremely high plasma renin levels¹²² that occurs in 20% of those with diffuse skin involvement, especially in those with rapidly progressive skin thickening of less than 3 years' duration¹²⁰. About 80% of cases with scleroderma renal crisis develop within the first 5 years after disease onset. Although uncommon, progressive azotemia and microangiopathic hemolytic anemia can occur in the setting of a persistently normal blood pressure¹²³. Abnormalities in renal sediment, including mild to moderate proteinuria and hematuria, are characteristic of scleroderma renal crisis, but nephrotic syndrome and cellular casts rarely occur. The new appearance of microangiopathic hemolytic anemia and thrombocytopenia was present in one study in 90% and 83%, respectively, of patients with normotensive scleroderma renal crisis versus 38% and 21% of those without renal crisis ($p < 0.01$ for both comparisons)¹²³. This outcome has been partially validated in normotensive scleroderma renal crisis, but has not been validated in classical hypertensive renal crisis.

Renal failure. Impending renal failure in SSc has been evaluated by measurements of blood pressure, fundoscopic examination, monitoring the complete blood count for microangiopathic hemolytic anemia or thrombocytopenia^{120,123}, serum creatinine, urinary sediment examination, measurement of timed urine collections for creatinine

clearance, and determining para-amino hippurate (PAH) clearance to measure renal plasma flow. Monitoring plasma renin activity has also been advocated and tested^{122,124}, and isolated elevations of plasma renin activity were not found to be predictive of renal crisis or the development of renal insufficiency¹²⁴. Renal biopsies and angiograms have also been employed^{125,126}, though due to their invasive nature and unproven sensitivity to change, these tests are not considered feasible as measures for clinical trials. Significant rise in both systolic and diastolic blood pressure seen before versus at onset of scleroderma renal crisis has been the most consistent measure found^{120,124,127}, accounting for 6 of 7 patients in one study who developed renal impairment¹²⁷. Proteinuria, classically associated with renal involvement in SSc¹²⁸, was found in 100% of those who developed renal failure in one large study¹²⁰ and in 5 out of 7 patients in another¹²⁷, but had poor specificity¹²⁰ and thus is not valid for use in a clinical trial.

The validity of serum creatinine and 24-hour collections for creatinine clearance in evaluating renal function has not been tested in scleroderma renal crisis, but it has been used as the reference against which other measures have been tested.

Conclusions and future directions of research. Future efforts should be directed at further validating the complete blood count and fundoscopic examination, as well as simpler measures of assessing renal function, such as 2-hour urine collections for creatinine clearance.

Outcome Measures for Gastrointestinal Involvement in Systemic Sclerosis

The gastrointestinal (GI) tract in SSc is a cause of a great deal of morbidity and is the third most common cause of mortality in scleroderma¹²⁹. Esophageal abnormalities occur in up to 90% of patients, stomach involvement such as gastric retention and gastric antral vascular ectasia (GAVE or "watermelon stomach") documented in 50% or more of

Table 5. Validation of outcome measures for renal disease in systemic sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
Blood pressure ^{120,123,124,127}	V	V	V	V	V	Excellent	Yes
Fundoscopy ^{120,121}	PV	PV	PV	NT	PV	Excellent	No
Complete blood count ^{120,123}	PV	Not Valid	Not Valid	NT	Not Valid	Excellent	No
Serum creatinine ^{120,121,123,127}	V	NT	NT	V	NT	Excellent	Yes
Urinalysis	PV	Not Valid	Not Valid	NT	NT	Excellent	No
24-h creatinine clearance ^{120,127}	PV	NT	NT	NT	PV	Good	Yes
Plasma renin activity ^{122,124}	V	PV	Not Valid	NT	V	Poor	No
Renal plasma flow PAH clearance ¹²⁴	PV	PV	PV	NT	PV	Poor	No
Renal biopsy ^{126,128}	PV	PV	NT	NT	V	Good	No
Renal angiography ¹²⁵	V	V	NT	NT	V	Poor	No

PAH: para-amino hippurate; V: validated; PV: partially validated; NT: not tested.

patients, and small bowel, colonic and anorectal involvement, including dysmotility and malabsorption, occurs in 50% to 70% of patients with SSc¹³⁰.

Defining outcomes to measure change in the GI tract can be approached by considering the probable pathogenesis of this visceral manifestation. The disease probably starts as a vascular insult of the vasa nervorum with secondary neurological dysfunction, smooth muscle atrophy, and increasing fibrosis. The consequences of these changes include dysmotility and, in the lower bowel, bacterial overgrowth. Measurements in the above context can include radiographs, manometry, pH monitoring, and myoelectric change and transit times, as well as biopsies and tests of malabsorption^{131,132}. Table 6 lists the main outcome measures investigated for GI disease in SSc and the current status of validation according to the OMERACT guidelines.

Plain radiographs. Although radiographs of the GI tract are certainly important measures in the clinical care of SSc patients, these tests have poor content/convergent validity or have not been tested as outcome measures. For example, using the lower esophageal sphincter pressure as a standard of comparison versus radiographs, the radiographs of the esophagus yield a false positive rate of 36% and a false negative rate of 17%¹³²⁻¹³⁵. Using biopsies as the standard for comparison the false positive rate is 50% in the esoph-

agus¹³²⁻¹³⁵. Using D-xylose as a standard of abnormality in the small bowel, radiographs give a 50% false negative rate^{132,135}. Reproducibility/reliability have not been examined and, importantly, there have been no data on the sensitivity of this measure to change. Consequently, radiographs of the GI tract in SSc should not at present be used as measures of response in clinical trials.

Tests of esophageal motility. Esophageal motility in patients with SSc may be measured using a number of techniques, including manometry, myoelectric change, biopsy, and transit times. Upper esophageal sphincter pressures by manometry are 100% discordant with respect to radiographs, while 27% of patients with symptoms have normal lower esophageal sphincter pressures. The reproducibility of manometry includes a coefficient of variation of roughly 30% and manometry is able to detect the effects of cisapride although not of metoclopramide^{130,131,136-140}. The convergent validity correlation of myoelectric changes and manometric recordings in the esophagus is excellent. However, reproducibility is not well documented and sensitivity to change is low, except in very severely affected patients^{130,138,140}. The relationship of duodenal myoelectric change to the full range of responses in the GI tract is poorly documented. There is good colonic convergent validity (90% true positive and 100% true negative) for colonic myoelectric

Table 6. Validation of outcome measures for gastrointestinal disease in systemic sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
Motility							
Manometry ^{130,131,135-142,144,145}	V	PV	NT	V	NT	Good	No
Esoph. transit time ^{137,138,140,144}	V	V	PV	V	NT	Good	No
Gastric transit time ^{130,144}	PV	V	PV	V	NT	Good	No
Esophageal myoelectric changes ^{137,138,140,142}	PV	V	PV	NV	NT	Poor	No
Duodenal myoelectric changes ^{130,143}	PV	NT	NT	NV	NT	Poor	No
Colonic myoelectric change ^{130,136,138,141,142,146}	PV	NV	NT	NT	NT	Poor	No
Biopsies							
Esophageal ^{132-135,140}	V	V	V	Good for infections	Excellent	Good	Yes
Jejunal ^{141,143}	V	NV	NT	NT	Excellent	Poor	No
Endoscopy*							
Esophageal ^{134,140,141,143}	V	PV	NT	V for infection	PV	Good	No
Gastric ¹⁴³	V	NT	NT	NT	NT	Good	No
Duo/jejunal ^{141,143}	V	NT	NT	NT	NT	Good	No
Colonic ¹³³	V	NT	NT	NT	NT	Good	No
Esophagitis							
1. PH monitoring ^{137,140}	V	V	NT	V for outcomes	Excellent	Poor	No
2. Acid perfusion test ^{137,140}	V	V	NT	NT	Excellent	Poor	No
1 & 2 combined ¹⁴⁰	V	V	NT	NT	Excellent	Poor	No
Malabsorption							
Hydrogen breath test ^{136,145}	V	V	NT	NT	Excellent	Good	No
D-xylose ^{132,135,141,143,146}	V	PV	NT	Good	Excellent	Good	No
72h fecal fat ^{141,143}	V	PV	NT	V	Excellent	Poor	No

* For inflammation and physical abnormalities N neural or muscular changes. DUO: duodenal. V: validated; PV: partially validated; NT: not tested.

changes versus manometry^{136,138,141,142}. Sensitivity to change of myoelectric measurements in the duodenum and colon of patients with SSc are poorly documented except in severe cases^{136,142}.

Gastrointestinal mucosal biopsy. Although GI mucosal biopsies are an appealing “gold standard” for clinical outcome measures, biopsies only sample the mucosa and some adventitia and are rarely deep enough to examine vasa nervorum or muscular atrophy. There are limited data supporting their use and some question whether tissue samples can fully document functional abnormalities. Biopsies have good content validity in the esophagus: 90% of biopsies from healthy subjects are negative and 94% of biopsies from symptomatic patients are positive for submucosal and muscular abnormalities and inflammation. Data on the reproducibility of biopsy results in the esophagus have not been published and sensitivity to change has only been documented with candidiasis¹⁴⁰. Jejunal biopsies and endoscopy of the duodenum and proximal ileum have unsatisfactory convergent validity^{141,143}. Thirty-nine percent of patients with SSc without symptoms had esophagitis on endoscopy and 10% of asymptomatic healthy controls demonstrated esophagitis¹⁴⁰. There are no data regarding the reproducibility of jejunal biopsies or endoscopy, and sensitivity to change of these tests is also not well documented.

Tests of esophageal and gastric transit time. When studied in patients with SSc, esophageal and gastric emptying times had ceiling effects, limiting the ability to measure the full range of responses (content validity). There is marked within-patient variability of this measure, as its coefficient of variation is 50% to 100%. Esophageal and gastric transit times measure sensitivity to change as analyzed by interquartile ranges across populations¹⁴⁴.

Tests for symptomatic gastroesophageal reflux disease. Heartburn is a common symptom in SSc, so measures to document this abnormality will be treated separately. Using the lack of peristalsis as the standard and abnormality being defined as > 5% reflux time, pH monitoring had a 26% false-negative rate in patients without SSc and symptoms of esophagitis^{137,140}. The sensitivity to change of pH monitoring can be documented when using antacids, but there are no data on reproducibility^{137,140}. The acid perfusion test has a 12% to 15% false-positive rate, but there are no data with respect to its reproducibility or sensitivity to change¹⁴⁰. However, a combination of pH monitoring or esophageal biopsy plus acid perfusion testing has good content validity, with a sensitivity of 87% to 90% but a false-positive rate of 30% in healthy controls. Data on reproducibility and sensitivity to change of these tests are lacking¹⁴⁰.

Tests for intestinal malabsorption. Several tests for intestinal malabsorption have been evaluated for use with patients with SSc. Bile acids or hydrogen breath tests have reasonable content validity versus bacterial counts from the

jejunum¹⁴⁵. For example, the sensitivity of the bile acid breath test as surrogate for bacterial overgrowth is 70% and its specificity ranges from 87% to 90%¹⁴⁵. The D-xylose test correlates with fecal fat and jejunal flora and was abnormally low in 11 of 16 patients with malabsorption^{141,143,146}. Improvement in the D-xylose test occurs after successful treatment of malabsorption with antibiotics. The lactulose test examines small intestinal permeability, but there are few data with respect to this test’s utility in SSc. The 72-hour fecal fat test on a 100-gram fat diet revealed a 100% abnormality among patients with SSc and radiographic abnormalities, and was sensitive to change after administration of either pancreatic enzymes or antibiotics to patients with malabsorption^{141,143} and improved bowel function occurred. Unfortunately, this is not an easy test to perform.

Conclusions and future directions of research. Measures of involvement of scleroderma in the GI domain suffer from a lack of data regarding their content/convergent validity, reproducibility/reliability, and sensitivity to change. Nevertheless, there are sufficient measures with some data on validity to prompt the research necessary to validate appropriate measures for use in clinical trials of SSc. It would be important to examine the use of combinations of measures so that the potential pathophysiology of GI disease in SSc is examined. The development of response measures in this domain will require a great deal of work in the future, although the GI tract is one of the core domains of illness in SSc.

Outcome Measures for Musculoskeletal Disease in Systemic Sclerosis

While joint contractures, especially in the fingers and the elbows, occur frequently in SSc, appreciable joint inflammation is uncommon in patients with SSc¹⁴⁷⁻¹⁵¹, although erosive arthropathy has been described¹⁴⁸. It is difficult to distinguish whether diminished joint mobility and function in patients with SSc is more related to joint, tendon, or skin involvement, or a combination of the 3. Table 7 lists the main outcome measures investigated for musculoskeletal disease in SSc and the current status of validation according to the OMERACT guidelines.

Arthropathy. Joint involvement in SSc has been measured by a number of methods. Arthralgia, a subjective measure, can be quantitated on a visual analog scale and has been tested in other rheumatic diseases, although not in SSc. Swelling over joints in patients with SSc is difficult to distinguish from skin or tendon thickening and thus is not useful or valid. The tender joint count is a standard measure used in scleroderma clinical trials, but it has not been formally tested in SSc. Finger-to-palm distance is also a frequently employed measure, but one validation study of this measure found an interobserver variation of 24% and a coefficient of variation of 0.52¹⁵², suggesting this not to be a valid outcome measure. In the same study, hand-spread

Table 7. Validation of outcome measures for musculoskeletal disease in system sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
Joint disease							
Arthralgia (VAS)	V	NT	Not valid	NT	NT	Excellent	Yes
Swollen joint count	Not valid	NT	Not valid	NT	Not valid	Good	No
Tender joint count	V	V	NT	NT	V	Excellent	Yes
Finger-to-palm distance	V	NT	Not valid	Not valid	PV	Excellent	No
Hand-spread	PV	Not valid	PV	Not valid	NT	Excellent	No
Grip strength	NT	NT	Not valid	NT	NT	Excellent	No
Flexion contractures	Not valid	NT	NT	Not valid	Not valid	Excellent	No
Proximal muscle weakness assessment	NT	NT	NT	NT*	Excellent	Good	No
No. of tendon friction rubs	Not valid	Not Valid	Not valid	Not valid	Not valid	Good	No
Hand x-rays (joints)	PV	PV	Not valid	Not valid	Not valid	Good	No
Tc-pertechnetate scintigraphy	PV	Not valid	Not valid	Not valid	PV	Poor	No
Thermography (joints)	PV	Not valid	Not valid	NT	PV	Poor	No
Muscle disease							
Serum creatine kinase	V	Not valid	NT	NT*	NT	Excellent	Yes*
Electromyogram	PV	PV	PV	NT*	NT	Poor	No
Muscle biopsy	V	V	NT	NT*	V	Poor	No
MRI	PV	PV	NT*	NT*	NT	Good	No
³¹ P-myosin magnetic resonance spectroscopy	NT	NT	NT	NT*	NT	Poor	No
¹¹¹ In-antimyosin antibody scintigraphy	NT	NT	NT	NT	NT	Poor	No

* Good for patients with SSc-polymyositis overlap syndrome, unknown for scleroderma-related bland myopathy. V: validated; PV: partially validated; NT: not tested.

distance was found to perform better, with an interobserver variation of less than 5% and a coefficient of variation of 0.25¹⁵². Thus, this measure is regarded as at least partially validated, although its sensitivity to change is less clear. Grip strength and quantitation of joint flexion contractures are also utilized, but both measures' sensitivities to change in SSc is fair to poor, and they have not been validated in SSc. Abnormalities in hand radiographs, specifically marginal erosions and acro-osteolysis, have also been examined. In one study erosions correlated with tender joint counts (tender joint count 14.5 ± 8.3 in those with erosions versus 7.5 ± 8.4 in those without; $p < 0.02$)¹⁴⁸. However, in 2 other studies, there were no correlations of the presence of erosions with either clinical synovitis^{149,150} or finger-to-palm distance¹⁴⁷, and thus are not valid outcome measures. Other measures, such as Tc-pertechnetate scintigraphy^{147,151} or thermography¹⁵¹, are not widely available and are not feasible for clinical trials. Tendon friction rubs due to inflammatory and fibrotic involvement of tendon sheath occur most typically over the wrists, elbows, ankles, and knees. Measuring the number of tendon friction rubs has not been shown to be sensitive to change nor to correlate with other measures. The Health Assessment Questionnaire has been validated in SSc and is discussed in the next section of this article.

Myopathy. Muscle disease in SSc can occur due to a variety of causes. Muscle weakness can result from disuse atrophy either due to malnutrition or to contractures of fibrotic skin^{153,154}. A bland myopathy may occur in up to 20% of SSc patients, associated with mild elevations of serum creatine kinase, a typical myopathic electromyographic picture, and muscle biopsy showing prominent interstitial fibrosis and mild inflammation and muscle fiber degeneration that is generally poorly responsive to glucocorticoid therapy¹⁵³⁻¹⁵⁵. An overlap syndrome with frank and clinically significant inflammatory myositis occurs much less commonly (less than 5% in most series). There are no reliably validated measures to quantitate muscular involvement in SSc, particularly the bland myopathy that is most frequently encountered. Serum creatine kinase is the most reliable measure of the activity of inflammatory polymyositis. There is no measure to accurately or consistently quantify muscle weakness. Electromyography and muscle biopsy¹⁵⁵ are reliable in polymyositis associated with SSc, but their sensitivity to change (i.e., treatment) has not been validated in sclerodermatous myopathy. T1 and T2-weighted magnetic resonance imaging has been validated in inflammatory myositis¹⁵⁶, but not in scleroderma. ³¹P magnetic resonance spectroscopy¹⁵⁷ and ¹¹¹In-antimyosin scintigraphy¹⁵⁸ have all been examined in polymyositis, but have not been tested in the bland

myopathy associated with SSc and are not widely available, and thus are not feasible for clinical trials.

Conclusions and future directions of research. Future studies should be directed toward further validation of measures with excellent feasibility, such as tender joint counts and grip strength, and at testing sensitivity to change of other measures, such as flexion contractures and hand-spread. Better methods to quantify proximal muscle strength are needed. Defining the sensitivity to change in the bland myopathy of SSc of measures that have been validated in polymyositis, such as serial creatine kinase measurements and MRI scanning, will be important to advance research on SSc-related myopathy. Ongoing work on outcome measures in idiopathic inflammatory myopathies may provide guidance to similar measures in SSc.

Outcome Measures for Function and Health-Related Quality of Life in Systemic Sclerosis

The use of functional and health-related quality of life (HRQOL) measures for chronic diseases has been widely recommended. These measures can be either generic or disease-specific¹⁵⁹. Generic measures are usually multidimensional and encompass various aspects of health and function, most commonly physical, mental, and social functioning, and symptom distress. They can be used across diseases and can provide a good measure of the burden of illness from a given disease in comparison to other conditions. A disadvantage of generic measures is that if they do not include specific and important aspects of a given condition, they may not be adequately sensitive to change for specific disease states. Disease-specific measures focus on the aspects relevant to the disease under study, and are more discriminative and responsive than generic instruments, but they cannot be used across diseases for comparative purposes. It has been suggested that both disease-specific and generic measures be used simultaneously in clinical trials¹⁵⁹. Table 8 summarizes the current state of validation of measures of the various functional and HRQOL measures in SSc.

The use of functional and HRQOL measures in SSc has been limited. To date, only one functional measure, the Health Assessment Questionnaire (HAQ)¹⁶⁰, has been adequately evaluated in SSc^{16,20,104,161-167}, but it only considers physical functioning. The HAQ was originally developed for arthritis, so it is not truly “specific” for SSc. The Scleroderma HAQ (SHAQ) includes the disability and pain scales of the HAQ plus 5 VAS that patients use to rate scleroderma-specific problems in the preceding week including pulmonary disease, digital ulcers, Raynaud’s phenomenon, GI disease, and skin disease²⁰. The SHAQ has also been partially validated by examining its relationship with other measures, its ability to discriminate among patients with severe and mild disease, its properties as a predictor of survival, and its responsiveness in clinical trials and cohort studies^{16,20,104,165,167}. The evidence for responsiveness is limited, but this is due in part to the lack of effective therapies in SSc, which results in very few patients showing detectable improvement.

Three additional functional scales have been developed: (1) the Systemic Sclerosis Questionnaire (SySQ) (German language) was developed by Ruof, *et al*¹⁶⁸; (2) an 11-item Functional Questionnaire specific for SSc that evaluates upper limb function¹⁶⁹; and (3) the Scleroderma Functional Index, an 11-item scale (French language) that has not been widely used¹⁷⁰. In a comparative study the Scleroderma Functional Index was shown to be less discriminative than the HAQ for patients with varying disease severity¹¹.

Several generic instruments have been developed to evaluate HRQOL [1996 #1224] including the SF-36, the Sickness Impact Profile, and the Quality of Well-Being Scale, among others. Although these measures have been validated and are commonly used in a number of rheumatic diseases, their use in SSc has been limited, and there are only limited preliminary published data using the SF-36 in patients with SSc from the GENISOS study and from Italian investigators¹⁷¹⁻¹⁷³. The inclusion of the SF-36 into current therapeutic trials will undoubtedly help validate this instrument for use with patients with SSc.

Table 8. Validation of outcome measures for function and health related quality of life in systemic sclerosis.

Outcome Measure	Face Validity (Credibility)	Content Validity (Comprehensiveness)	Criterion Validity (Accuracy)	Discriminant Validity (Sensitivity to Change)	Construct Validity (Biological Sense)	Feasibility	Ready for Use in Clinical Trials?
HAQ ^{16,20,104,161-167}	V	PV	V	PV	PV	Excellent	Yes
SHAQ Visual Analog Scales ^{16,20,104,161,165,167}	V	PV	V	PV	PV	Excellent	Yes
SySQ ¹⁶⁸	V	V	PV	NT	PV	Excellent (German)	No
Functional Questionnaire, upper limb ¹⁶⁹	V	V	U	U	NT	Excellent	No
Scleroderma Functional Index, upper limb ¹⁷⁰	V	U	NT	NT	NT	Excellent (French)	No
SF-36 ¹⁷¹⁻¹⁷³	V	NT	PV	U	NT	Excellent	No

HAQ: Health Assessment Questionnaire; SHAQ: Scleroderma Health Assessment Questionnaire; SySQ: Systemic Sclerosis Questionnaire; V: validated; PV: partially validated; NT: not tested; U: unclear.

Conclusions and future directions of research. The use of functional and HRQOL measures in SSc has been limited, and only the HAQ has been adequately evaluated. Although it shows good psychometric properties, it only considers physical functioning. SSc is a multisystem disorder, with a broad spectrum of organ manifestations, some of which are unique, and it seems desirable to use both disease-specific and generic measures to evaluate functioning and HRQOL. Additional research on existing or modified instruments is needed. The development of new, more comprehensive patient-centered tools that assess the impact of the various disease components on function and well being is also recommended.

DISCUSSION

Along with an understanding, or at least a model, for the pathogenesis of the disease, validated measures of response will allow movement towards effective treatment of SSc. For example, some understanding of the pathogenesis of renal disease in SSc along with the use of validated measures such as blood pressure measurements and measurement of mortality eventually led to the use of angiotensin-converting enzyme inhibitors in scleroderma renal crisis. This now common treatment of scleroderma renal crisis would not have been possible without accepted and validated measures of response — blood pressure and mortality.

During the past 20 years, slow but steady progress has been made in developing outcomes tools to measure change in SSc, as evidenced by the work outlined above. However, as also outlined above, there are insufficient well validated tools ready for use for many of the core domains of illness in SSc. Of course, outcomes such as mortality require little validation, and there is work presently being done to examine the ability to use tools in SSc that have been developed for other indications or in normal populations. An example of that is the SF-36, which is being tested for its validity in SSc.

The literature with respect to outcome measures in SSc was reviewed at the OMERACT 6 meeting. Presenters and the audience discussed the results and made suggestions about what areas or domains needed to be improved or validated as outcome measures in SSc. It became immediately clear that certain domains would be particularly difficult to separate. For example, it appeared quite difficult to separate, using clinically available or potentially available tools, the cardiac from the pulmonary systems. On the other hand, the manifestation of disease in certain other domains was so uncommon that it was deemed unwise to tackle them at this juncture. An example of such a domain is the peripheral nervous system.

As one of the purposes of this workshop was to suggest directions and research needs in the area of outcome measures in SSc, significant time was taken eliciting ideas

for future research. In this context, the task of OMERACT is to “point the way.” It is not to mandate any task or designate anyone, specifically, to do the task. Agendas for the future direction of outcomes assessment for different domains are outlined at the end of each section above. Any individual or group of individuals can decide on what to do and how to approach the research.

It is our hope that SSc researchers will join with metrologists, epidemiologists, and statisticians to answer some of the questions and research agendas outlined in this article. Combined with the marked increase in new pharmaceutical agents for patients with SSc, progress in development and validation of outcome measures for use in clinical trials in SSc will lead to more effective treatment of SSc as well as to a better understanding of the pathogenesis of this complex and sometimes devastating disease.

REFERENCES

1. Tugwell P, Bombardier C. A methodologic framework for developing and selecting endpoints in clinical trials. *J Rheumatol* 1982;9:758-62.
2. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;25:198-9.
3. Bellamy N. Clinimetric concepts in outcome assessment: the OMERACT filter. *J Rheumatol* 1999;26:948-50.
4. Rodnan GP, Lipinski E, Luksick J. Skin thickness and collagen content in progressive systemic sclerosis and localized scleroderma. *Arthritis Rheum* 1979;22:130-40.
5. Furst DE, Clements PJ, Steen VD, et al. The modified Rodnan skin score is an accurate reflection of skin biopsy thickness in systemic sclerosis. *J Rheumatol* 1998;25:84-8.
6. 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.
7. Furst DE, Clements PJ, Hillis S, et al. Immunosuppression with chlorambucil, versus placebo, for scleroderma. Results of a three-year, parallel, randomized, double-blind study. *Arthritis Rheum* 1989;32:584-93.
8. Clements PJ, Lachenbruch PA, Ng SC, Simmons M, Sterz M, Furst DE. Skin score. A semiquantitative measure of cutaneous involvement that improves prediction of prognosis in systemic sclerosis. *Arthritis Rheum* 1990;33:1256-63.
9. Clements PJ, Lachenbruch PA, Seibold JR, et al. Skin thickness score in systemic sclerosis: an assessment of interobserver variability in 3 independent studies. *J Rheumatol* 1993;20:1892-6.
10. Clements P, Lachenbruch P, Siebold J, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol* 1995;22:1281-5.
11. Pope JE, Baron M, Bellamy N, et al. Variability of skin scores and clinical measurements in scleroderma. *J Rheumatol* 1995;22:1271-6.
12. Steen VD, Medsger TA Jr, Rodnan GP. D-Penicillamine therapy in progressive systemic sclerosis (scleroderma): a retrospective analysis. *Ann Intern Med* 1982;97:652-9.
13. Rook AH, Freundlich B, Jegasothy BV, et al. Treatment of systemic sclerosis with extracorporeal photochemotherapy. Results of a multicenter trial. *Arch Dermatol* 1992;128:337-46.
14. 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.

15. Clements PJ, Furst DE, Wong WK, 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;42:1194-203.
16. Seibold JR, Korn JH, Simms R, et al. Recombinant human relaxin in the treatment of scleroderma. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 2000;132:871-9.
17. Clements PJ, Hurwitz EL, Wong WK, et al. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. *Arthritis Rheum* 2000; 43:2445-54.
18. 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.
19. Steen VD, Medsger TA Jr, Osial TA Jr, Ziegler GL, Shapiro AP, Rodnan GP. Factors predicting development of renal involvement in progressive systemic sclerosis. *Am J Med* 1984;76:779-86.
20. Steen VD, Medsger TA Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. *Arthritis Rheum* 1997;40:1984-91.
21. Steen VD, Medsger TA Jr. Improvement in skin thickening in systemic sclerosis associated with improved survival. *Arthritis Rheum* 2001;44:2828-35.
22. Scheja A, Akesson A. Comparison of high frequency (20 MHz) ultrasound and palpation for the assessment of skin involvement in systemic sclerosis (scleroderma). *Clin Exp Rheumatol* 1997; 15:283-8.
23. Balbir-Gurman A, Denton CP, Nichols B, et al. Non-invasive measurement of biomechanical skin properties in systemic sclerosis. *Ann Rheum Dis* 2002;61:237-41.
24. Merkel PA, Kaluta A, Brancalion L, et al. New methods of assessing skin disease in scleroderma: use of durometers, photonic devices, and comparison to modified Rodnan skin scoring [abstract]. *Arthritis Rheum* 2001;44 Suppl:S330.
25. Steen VD, Conte C, Owens GR, Medsger TA Jr. Severe restrictive lung disease in systemic sclerosis. *Arthritis Rheum* 1994;37:1283-9.
26. Stupi AM, Steen VD, Owens GR, Barnes EL, Rodnan GP, Medsger TA Jr. Pulmonary hypertension in the CREST syndrome variant of systemic sclerosis. *Arthritis Rheum* 1986;29:515-24.
27. Wells AU, King AD, Rubens MB, Cramer D, du Bois RM, Hansell DM. Lone cryptogenic fibrosing alveolitis: a functional-morphologic correlation based on extent of disease on thin-section computed tomography. *Am J Respir Crit Care Med* 1997; 155:1367-75.
28. Risk C, Epler GR, Gaensler EA. Exercise alveolar-arterial oxygen pressure difference in interstitial lung disease. *Chest* 1984;85:69-74.
29. Xaubet A, Agusti C, Luburich P, et al. Pulmonary function tests and CTscan in the management of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1998;158:431-6.
30. Chinet T, Jaubert F, Dusser D, Danel C, Chretien J, Huchon GJ. Effects of inflammation and fibrosis on pulmonary function in diffuse lung fibrosis. *Thorax* 1990;45:675-8.
31. Cherniack RM, Colby TV, Flint A, et al. Correlation of structure and function in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1995;151:1180-8.
32. Watters LC, King TE, Schwarz MI, Waldron JA, Stanford RE, Cherniack RM. A clinical, radiographic, and physiologic scoring system for the longitudinal assessment of patients with idiopathic pulmonary fibrosis. *Am Rev Respir Dis* 1986;133:97-103.
33. Schwaiblmair M, Behr J, Fruhmann G. Cardiorespiratory responses to incremental exercise in patients with systemic sclerosis. *Chest* 1996;110:1520-5.
34. Chang JA, Curtis JR, Patrick DL, Raghu G. Assessment of health-related quality of life in patients with interstitial lung disease. *Chest* 1999;116:1175-82.
35. Behr J, Vogelmeier C, Beinert T, et al. Bronchoalveolar lavage for evaluation and management of scleroderma disease of the lung. *Am J Respir Crit Care Med* 1996;154:400-6.
36. 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.
37. Wells AU, Hansell DM, Rubens MB, et al. Fibrosing alveolitis in systemic sclerosis: indices of lung function in relation to extent of disease on computed tomography. *Arthritis Rheum* 1997; 40:1229-36.
38. King TE Jr, Tooze JA, Schwarz MI, Brown KR, Cherniack RM. Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med* 2001;164:1171-81.
39. Hanson D, Winterbauer RH, Kirtland SH, Wu R. Changes in pulmonary function test results after 1 year of therapy as predictors of survival in patients with idiopathic pulmonary fibrosis. *Chest* 1995;108:305-10.
40. Erbes R, Schaberg T, Loddenkemper R. Lung function tests in patients with idiopathic pulmonary fibrosis. Are they helpful for predicting outcome? *Chest* 1997;111:51-7.
41. Bouros D, Wells AU, Nicholson AG, et al. Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *Am J Respir Crit Care Med* 2002; 165:1581-6.
42. Raghu G, Cain K, Winterbauer R. Improved forced vital capacity, DL50SB and resting P(A-a) O₂ measurements at one year predict long term survival in idiopathic pulmonary fibrosis [abstract]. *Am Rev Respir Dis* 1991;143:A57.
43. Garrard CS, Emmons C. The reproducibility of the respiratory responses to maximum exercise. *Respiration* 1986;49:94-100.
44. Marciniuk DD, Watts RE, Gallagher CG. Reproducibility of incremental maximal cycle ergometer testing in patients with restrictive lung disease. *Thorax* 1993;48:894-8.
45. Silver RM, Miller KS, Kinsella MB, Smith EA, Schabel SI. Evaluation and management of scleroderma lung disease using bronchoalveolar lavage. *Am J Med* 1990;88:470-6.
46. White B, Moore WC, Wigley FM, Xiao HQ, Wise RA. Cyclophosphamide is associated with pulmonary function and survival benefit in patients with scleroderma and alveolitis. *Ann Intern Med* 2000;132:947-54.
47. Ziesche R, Hofbauer E, Wittmann K, Petkov V, Block LH. A preliminary study of long-term treatment with interferon gamma-1b and low-dose prednisolone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 1999;341:1264-9.
48. Harris-Eze AO, Sridhar G, Clemens RE, Zintel TA, Gallagher CG, Marciniuk DD. Role of hypoxemia and pulmonary mechanics in exercise limitation in interstitial lung disease. *Am J Respir Crit Care Med* 1996;154:994-1001.
49. Hansen JE, Wasserman K. Pathophysiology of activity limitation in patients with interstitial lung disease. *Chest* 1996;109:1566-76.
50. Wells AU, Hansell DM, Haslam PL, et al. Bronchoalveolar lavage cellularity: lone cryptogenic fibrosing alveolitis compared with the fibrosing alveolitis of systemic sclerosis. *Am J Respir Crit Care Med* 1998;157:1474-82.
51. Harrison NK, Myers AR, Corrin B, et al. Structural features of interstitial lung disease in systemic sclerosis. *Am Rev Respir Dis* 1991;144:706-13.
52. Warrick JH, Bhalla M, Schabel SI, Silver RM. High resolution computed tomography in early scleroderma lung disease. *J Rheumatol* 1991;18:1520-8.
53. Peters-Golden M, Wise RA, Hochberg MC, Stevens MB, Wigley FM. Carbon monoxide diffusing capacity as predictor of outcome in

- systemic sclerosis. *Am J Med* 1984;77:1027-34.
54. Wells AU, Cullinan P, Hansell DM, et al. Fibrosing alveolitis associated with systemic sclerosis has a better prognosis than lone cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 1994;149:1583-90.
 55. Ungerer RG, Tashkin DP, Furst D, et al. Prevalence and clinical correlates of pulmonary arterial hypertension in progressive systemic sclerosis. *Am J Med* 1983;75:65-74.
 56. Crapo RO, Jensen RL, Wanger JS. Single-breath carbon monoxide diffusing capacity. *Clin Chest Med* 2001;22:637-49.
 57. Steen VD, Graham G, Conte C, Owens G, Medsger TA Jr. Isolated diffusing capacity reduction in systemic sclerosis. *Arthritis Rheum* 1992;35:765-70.
 58. 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.
 59. Diot E, Boissinot E, Asquier E, et al. Relationship between abnormalities on high-resolution CT and pulmonary function in systemic sclerosis. *Chest* 1998;114:1623-9.
 60. Kim EA, Johkoh T, Lee KS, et al. Interstitial pneumonia in progressive systemic sclerosis: serial high-resolution CT findings with functional correlation. *J Comput Assist Tomogr* 2001; 25:757-63.
 61. Gay SE, Kazerooni EA, Toews GB, et al. Idiopathic pulmonary fibrosis: predicting response to therapy and survival. *Am J Respir Crit Care Med* 1998;157:1063-72.
 62. Wells AU, Hansell DM, Rubens MB, Cullinan P, Black CM, du Bois RM. The predictive value of appearances on thin-section computed tomography in fibrosing alveolitis. *Am Rev Respir Dis* 1993;148:1076-82.
 63. Watters LC, Schwarz MI, Cherniack RM, et al. Idiopathic pulmonary fibrosis. Pretreatment bronchoalveolar lavage cellular constituents and their relationships with lung histopathology and clinical response to therapy. *Am Rev Respir Dis* 1987;135:696-704.
 64. Wallaert B, Hatron PY, Grosbois JM, Tonnel AB, Devulder B, Voisin C. Subclinical pulmonary involvement in collagen-vascular diseases assessed by bronchoalveolar lavage. Relationship between alveolitis and subsequent changes in lung function. *Am Rev Respir Dis* 1986;133:574-80.
 65. Witt C, Borges AC, John M, Fietze I, Baumann G, Krause A. Pulmonary involvement in diffuse cutaneous systemic sclerosis: bronchoalveolar fluid granulocytosis predicts progression of fibrosing alveolitis. *Ann Rheum Dis* 1999;58:635-40.
 66. Turner-Warwick M, Haslam PL. The value of serial bronchoalveolar lavages in assessing the clinical progress of patients with cryptogenic fibrosing alveolitis. *Am Rev Respir Dis* 1987; 135:26-34.
 67. Harrison NK, Glanville AR, Strickland B, et al. Pulmonary involvement in systemic sclerosis: the detection of early changes by thin section CTscan, bronchoalveolar lavage and 99mTc-DTPA clearance. *Respir Med* 1989;83:403-14.
 68. O'Donnell DE, Chau LK, Webb KA. Qualitative aspects of exertional dyspnea in patients with interstitial lung disease. *J Appl Physiol* 1998;84:2000-9.
 69. 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.
 70. Rubin LJ, Badesch DB, Barst RJ, et al. Bosentan therapy for pulmonary arterial hypertension. *N Engl J Med* 2002;346:896-903.
 71. Channick RN, Simonneau G, Sitbon O, et al. Effects of the dual endothelin-receptor antagonist bosentan in patients with pulmonary hypertension: a randomised placebo-controlled study. *Lancet* 2001;358:1119-23.
 72. Fulmer JD, Roberts WC, von Gal ER, Crystal RG. Morphologic-physiologic correlates of the severity of fibrosis and degree of cellularity in idiopathic pulmonary fibrosis. *J Clin Invest* 1979;63:665-76.
 73. Wax D, Garofano R, Barst RJ. Effects of long-term infusion of prostacyclin on exercise performance in patients with primary pulmonary hypertension. *Chest* 1999;116:914-20.
 74. Guyatt GH, Sullivan MJ, Thompson PJ, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *CMAJ* 1985;132:919-23.
 75. Denton CP, Cailles JB, Phillips GD, Wells AU, Black CM, Bois RM. Comparison of Doppler echocardiography and right heart catheterization to assess pulmonary hypertension in systemic sclerosis. *Br J Rheumatol* 1997;36:239-43.
 76. Rich S, D'Alonzo GE, Dantzker DR, Levy PS. Magnitude and implications of spontaneous hemodynamic variability in primary pulmonary hypertension. *Am J Cardiol* 1985;55:159-63.
 77. Simonneau G, Herve P, Petitpretz P, et al. Detection of a reversible component in primary pulmonary hypertension: value of prostacyclin acute infusion [abstract]. *Am Rev Respir Dis* 1986;133:A223.
 78. Raffy O, Azarian R, Brenot F, et al. Clinical significance of the pulmonary vasodilator response during short-term infusion of prostacyclin in primary pulmonary hypertension. *Circulation* 1996;93:484-8.
 79. Owens GR, Follansbee WP. Cardiopulmonary manifestations of systemic sclerosis. *Chest* 1987;91:118-27.
 80. McWhorter JE, LeRoy EC. Pericardial disease in scleroderma (systemic sclerosis). *Am J Med* 1974;57:566-75.
 81. Smith JW, Clements PJ, Levisman J, Furst D, Ross M. Echocardiographic features of progressive systemic sclerosis. Correlation with hemodynamic and postmortem studies. *Am J Med* 1979;66:28-33.
 82. Follansbee WP, Curtiss EI, Rahko PS, et al. The electrocardiogram in systemic sclerosis (scleroderma). Study of 102 consecutive cases with functional correlations and review of the literature. *Am J Med* 1985;79:183-92.
 83. Hesselstrand R, Scheja A, Akesson A. Mortality and causes of death in a Swedish series of systemic sclerosis patients. *Ann Rheum Dis* 1998;57:682-6.
 84. Follansbee WP, Curtiss EI, Medsger TA Jr, et al. Physiologic abnormalities of cardiac function in progressive systemic sclerosis with diffuse scleroderma. *N Engl J Med* 1984;310:142-8.
 85. Ferri C, Di Bello V, Martini A, et al. Heart involvement in systemic sclerosis: an ultrasonic tissue characterisation study. *Ann Rheum Dis* 1998;57:296-302.
 86. Kahan A, Devaux JY, Amor B, et al. Nifedipine and thallium-201 myocardial perfusion in progressive systemic sclerosis. *N Engl J Med* 1986;314:1397-402.
 87. Kostis JB, Seibold JR, Turkevich D, et al. Prognostic importance of cardiac arrhythmias in systemic sclerosis. *Am J Med* 1988; 84:1007-15.
 88. Antoniadou L, Sfikakis PP, Mavrikakis M. Glucocorticoid effects on myocardial performance in patients with systemic sclerosis. *Clin Exp Rheumatol* 2001;19:431-7.
 89. Valentini G, Vitale DF, Giunta A, et al. Diastolic abnormalities in systemic sclerosis: evidence for associated defective cardiac functional reserve. *Ann Rheum Dis* 1996;55:455-60.
 90. Giunta A, Tirri E, Maione S, et al. Right ventricular diastolic abnormalities in systemic sclerosis. Relation to left ventricular involvement and pulmonary hypertension. *Ann Rheum Dis* 2000;59:94-8.
 91. Kazzam E, Caidahl K, Hillgren R, Gustafsson R, Waldenstrom A. Non-invasive evaluation of long-term cardiac effects of captopril in systemic sclerosis. *J Intern Med* 1991;230:203-12.

92. Bilan A, Chibowska M, Makaruk B, et al. Enalapril (10 mg/day) in systemic sclerosis. One year, double blind, randomised study (ESS-1): echocardiographic substudy — three months follow-up. *Adv Exp Med Biol* 1999;455:279-83.
93. Belch JJ, Newman P, Drury JK, et al. Successful treatment of Raynaud's syndrome with prostacyclin. *Thromb Haemost* 1981;45:255-6.
94. McHugh NJ, Csuka M, Watson H, et al. Infusion of iloprost, a prostacyclin analogue, for treatment of Raynaud's phenomenon in systemic sclerosis. *Ann Rheum Dis* 1988;47:43-7.
95. Kyle MV, Belcher G, Hazleman BL. Placebo controlled study showing therapeutic benefit of iloprost in the treatment of Raynaud's phenomenon. *J Rheumatol* 1992;19:1403-6.
96. Wigley FM, Wise RA, Seibold JR, et al. Intravenous iloprost infusion in patients with Raynaud phenomenon secondary to systemic sclerosis. A multicenter, placebo-controlled, double-blind study. *Ann Intern Med* 1994;120:199-206.
97. Vayssairat M. Controlled multicenter double blind trial of an oral analog of prostacyclin in the treatment of primary Raynaud's phenomenon. French Microcirculation Society Multicentre Group for the Study of Vascular Acrosyndromes. *J Rheumatol* 1996;23:1917-20.
98. Wigley FM, Korn JH, Csuka ME, et al. Oral iloprost treatment in patients with Raynaud's phenomenon secondary to systemic sclerosis: a multicenter, placebo-controlled, double-blind study. *Arthritis Rheum* 1998;41:670-7.
99. 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.
100. Dziadzio M, Denton CP, Smith R, et al. Losartan therapy for Raynaud's phenomenon and scleroderma: clinical and biochemical findings in a fifteen-week, randomized, parallel-group, controlled trial. *Arthritis Rheum* 1999;42:2646-55.
101. Herrick AL, Hollis S, Schofield D, et al. A double-blind placebo-controlled trial of antioxidant therapy in limited cutaneous systemic sclerosis. *Clin Exp Rheumatol* 2000;18:349-56.
102. Gardinali M, Pozzi MR, Bernareggi M, et al. Treatment of Raynaud's phenomenon with intravenous prostaglandin E1 alpha-cyclodextrin improves endothelial cell injury in systemic sclerosis. *J Rheumatol* 2001;28:786-94.
103. Sambo P, Amico D, Giacomelli R, et al. Intravenous N-acetylcysteine for treatment of Raynaud's phenomenon secondary to systemic sclerosis: a pilot study. *J Rheumatol* 2001;28:2257-62.
104. Merkel PA, Herlyn K, Martin RW, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. *Arthritis Rheum* 2002;46:2410-20.
105. Maricq HR, LeRoy EC. Patterns of finger capillary abnormalities in connective tissue disease by "wide-field" microscopy. *Arthritis Rheum* 1973;16:619-28.
106. ter Borg EJ, Piersma-Wichers G, Smit AJ, Kallenberg CG, Wouda AA. Serial nailfold capillary microscopy in primary Raynaud's phenomenon and scleroderma. *Semin Arthritis Rheum* 1994;24:40-7.
107. Jayanetti S, Smith CP, Moore T, Jayson MI, Herrick AL. Thermography and nailfold capillaroscopy as noninvasive measures of circulation in children with Raynaud's phenomenon. *J Rheumatol* 1998;25:997-9.
108. Herrick AL, Clark S. Quantifying digital vascular disease in patients with primary Raynaud's phenomenon and systemic sclerosis. *Ann Rheum Dis* 1998;57:70-8.
109. Wesseling H, den Heeten A, Wouda AA. Sublingual and oral isoxsuprine in patients with Raynaud's phenomenon. *Eur J Clin Pharmacol* 1981;20:329-33.
110. Leppert J, Jonasson T, Nilsson H, Ringqvist I. The effect of isradipine, a new calcium-channel antagonist, in patients with primary Raynaud's phenomenon: a single-blind dose-response study. *Cardiovasc Drugs Ther* 1989;3:397-401.
111. Leppert J. Limitation of finger systolic pressure measurement as a diagnostic test for primary Raynaud's phenomenon in a female population. *Clin Physiol* 1989;9:457-65.
112. Weber A, Bounameaux H. Effects of low-dose nifedipine on a cold provocation test in patients with Raynaud's disease. *J Cardiovasc Pharmacol* 1990;15:853-5.
113. Chucker FD, Fowler RC, Hurley CW. Photoplethysmometry and thermography in Raynaud's disorders. A preliminary report. *Angiology* 1973;24:612-8.
114. Ring EF, Bacon PA. Quantitative thermographic assessment of inositol nicotinate therapy in Raynaud's phenomena. *J Int Med Res* 1977;5:217-22.
115. Meloni F, Transi MG, Sciacca V, De Felice C, Bagarone A, Sciacca A. Therapeutic efficacy of ketanserin, a selective antagonist of the serotonin (5-HT₂) receptors, in primary and secondary Raynaud's phenomenon. *Angiology* 1987;38:530-6.
116. Whitmore SE, Wigley FM, Wise RA. Acute effect of topical minoxidil on digital blood flow in patients with Raynaud's phenomenon. *J Rheumatol* 1995;22:50-4.
117. Anderson ME, Hollis S, Moore T, Jayson MI, Herrick AL. Non-invasive assessment of vascular reactivity in forearm skin of patients with primary Raynaud's phenomenon and systemic sclerosis. *Br J Rheumatol* 1996;35:1281-8.
118. Pistorius MA, Planchon B, de Faucal P. Plethysmographic cold test for diagnosis and evaluation of the severity of Raynaud's phenomenon. Validation of the method by factorial analysis of correspondences in 541 patients. *Int Angiol* 1994;13:10-4.
119. Kallenberg CG, Wouda AA, Meems L, Wesseling H. Once daily felodipine in patients with primary Raynaud's phenomenon. *Eur J Clin Pharmacol* 1991;40:313-5.
120. Traub YM, Shapiro AP, Rodnan GP, et al. Hypertension and renal failure (scleroderma renal crisis) in progressive systemic sclerosis. Review of a 25-year experience with 68 cases. *Medicine (Baltimore)* 1983;62:335-52.
121. Steen VD, Costantino JP, Shapiro AP, Medsger TA Jr. Outcome of renal crisis in systemic sclerosis: relation to availability of angiotensin converting enzyme inhibitors. *Ann Intern Med* 1990;113:352-7.
122. Gavras H, Gavras I, Cannon PJ, Brunner HR, Laragh JH. Is elevated plasma renin activity of prognostic importance in progressive systemic sclerosis? *Arch Intern Med* 1977;137:1554-8.
123. Helfrich DJ, Banner B, Steen VD, Medsger TA Jr. Normotensive renal failure in systemic sclerosis. *Arthritis Rheum* 1989;32:1128-34.
124. Clements PJ, Lachenbruch PA, Furst DE, Maxwell M, Danovitch G, Paulus HE. Abnormalities of renal physiology in systemic sclerosis. A prospective study with 10-year followup. *Arthritis Rheum* 1994;37:67-74.
125. Warren BH, Rosch J. Angiography in the diagnosis of renal scleroderma. *Radiol Clin* 1977;46:194-202.
126. McCoy RC, Tisher CC, Pepe PF, Cleveland LA. The kidney in progressive systemic sclerosis: immunohistochemical and antibody elution studies. *Lab Invest* 1976;35:124-31.
127. Mallee C, Meijers KA, Cats A. Renal scleroderma, value of clinical markers. *Clin Exp Rheumatol* 1985;3:287-90.
128. Cannon PJ, Hassar M, Case DB, Casarella WJ, Sommers SC, LeRoy EC. The relationship of hypertension and renal failure in scleroderma (progressive systemic sclerosis) to structural and functional abnormalities of the renal cortical circulation. *Medicine (Baltimore)* 1974;53:1-46.
129. Poirier TJ, Rankin GB. Gastrointestinal manifestations of progressive systemic scleroderma based on a review of 364 cases. *Am J Gastroenterol* 1972;58:30-44.
130. Sjogren RW. Gastrointestinal motility disorders in scleroderma.

- Arthritis Rheum 1994;37:1265-82.
131. Cohen S, Fisher R, Lipshutz W, Turner R, Myers A, Schumacher R. The pathogenesis of esophageal dysfunction in scleroderma and Raynaud's disease. *J Clin Invest* 1972;51:2663-8.
 132. Greydanus MP, Camilleri M. Abnormal postcibal antral and small bowel motility due to neuropathy or myopathy in systemic sclerosis. *Gastroenterology* 1989;96:110-5.
 133. Jiranek GC, Bredfeldt JE. Gut and hepatic manifestations. In: Clements PJ, Furst DJ, editors. *Systemic sclerosis*. Baltimore: Williams & Wilkins; 1996:453-81.
 134. McCallum RW, Ippoliti AF, Cooney C, Sturdevant RA. A controlled trial of metoclopramide in symptomatic gastroesophageal reflux. *N Engl J Med* 1977;296:354-7.
 135. Behar J, Biancani P, Sheahan DG. Evaluation of esophageal tests in the diagnosis of reflux esophagitis. *Gastroenterology* 1976;71:9-15.
 136. Lauterburg BH, Newcomer AD, Hofmann AF. Clinical value of the bile acid breath test. Evaluation of the Mayo Clinic experience. *Mayo Clin Proc* 1978;53:227-33.
 137. Baron M, Arzoumanian A. Radionuclide esophageal transit studies in progressive systemic sclerosis: an analysis of longitudinal data. *J Rheumatol* 1991;18:1837-40.
 138. Scudamore HH, Green PA, Hofman HN 2nd, Rosevear JW, Tauxe WN. Scleroderma (progressive systemic sclerosis) of the small intestine with malabsorption. Evaluation of intestinal absorption and pancreatic function. *Am J Gastroenterol* 1968;49:193-208.
 139. Kahn IJ, Jeffries GH, Slesinger MH. Malabsorption in intestinal scleroderma. Correction by antibiotics. *N Engl J Med* 1966;274:1339-44.
 140. Yarze JC, Varga J, Stampfl D, Castell DO, Jimenez SA. Esophageal function in systemic sclerosis: a prospective evaluation of motility and acid reflux in 36 patients. *Am J Gastroenterol* 1993;88:870-6.
 141. Cohen S. The gastrointestinal manifestations of scleroderma: pathogenesis and management. *Gastroenterology* 1980;79:155-66.
 142. O'Mahoney SA, Menzies IS, Nukajam WS, Ciclitira PJ, Black CM. Small intestine permeability and absorptive capacity in 16 systemic sclerosis patients with index symptoms of malabsorption [abstract]. *Arthritis Rheum* 1992;39 Suppl:C227.
 143. Olmsted WW, Madewell JE. The esophageal and small-bowel manifestations of progressive systemic sclerosis. A pathophysiologic explanation of the roentgenographic signs. *Gastrointest Radiol* 1976;1:33-6.
 144. Kahan A, Chaussade S, Gaudric M, et al. The effect of cisapride on gastro-oesophageal dysfunction in systemic sclerosis: a controlled manometric study. *Br J Clin Pharmacol* 1991;31:683-7.
 145. Clements PJ, Kadell B, Ippoliti A, Ross M. Esophageal motility in progressive systemic sclerosis. Comparison of cine-radiographic and manometric evaluation. *Dig Dis Sci* 1979;24:639-44.
 146. Battle WM, Snape WJ Jr, Wright S, et al. Abnormal colonic motility in progressive systemic sclerosis. *Ann Intern Med* 1981;94:749-52.
 147. Lovell CR, Jayson MI. Joint involvement in systemic sclerosis. *Scand J Rheumatol* 1979;8:154-60.
 148. Blocka KL, Bassett LW, Furst DE, Clements PJ, Paulus HE. The arthropathy of advanced progressive systemic sclerosis. A radiographic survey. *Arthritis Rheum* 1981;24:874-84.
 149. Baron M, Lee P, Keystone EC. The articular manifestations of progressive systemic sclerosis (scleroderma). *Ann Rheum Dis* 1982;41:147-52.
 150. Catoggio LJ, Evison G, Harkness JA, Maddison PJ. The arthropathy of systemic sclerosis (scleroderma); comparison with mixed connective tissue disease. *Clin Exp Rheumatol* 1983;1:101-12.
 151. Misra R, Darton K, Jewkes RF, Black CM, Maini RN. Arthritis in scleroderma. *Br J Rheumatol* 1995;34:831-7.
 152. Furst DE, Clements PJ, Harris R, Ross M, Levy J, Paulus HE. Measurement of clinical change in progressive systemic sclerosis: a 1 year double-blind placebo-controlled trial of N-acetylcysteine. *Ann Rheum Dis* 1979;38:356-61.
 153. Clements PJ, Furst DE, Campion DS, et al. Muscle disease in progressive systemic sclerosis: diagnostic and therapeutic considerations. *Arthritis Rheum* 1978;21:62-71.
 154. Ringel RA, Brick JE, Brick JF, Gutmann L, Riggs JE. Muscle involvement in the scleroderma syndromes. *Arch Intern Med* 1990;150:2550-2.
 155. Russell ML, Hanna WM. Ultrastructural pathology of skeletal muscle in various rheumatic diseases. *J Rheumatol* 1988;15:445-53.
 156. Kaufman LD, Gruber BL, Gerstman DP, Kaell AT. Preliminary observations on the role of magnetic resonance imaging for polymyositis and dermatomyositis. *Ann Rheum Dis* 1987;46:569-72.
 157. Park JH, Vital TL, Ryder NM, et al. Magnetic resonance imaging and P-31 magnetic resonance spectroscopy provide unique quantitative data useful in the longitudinal management of patients with dermatomyositis. *Arthritis Rheum* 1994;37:736-46.
 158. Lofberg M, Liewendahl K, Lamminen A, Korhola O, Somer H. Antimyosin scintigraphy compared with magnetic resonance imaging in inflammatory myopathies. *Arch Neurol* 1998;55:987-93.
 159. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27:S217-32.
 160. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
 161. Poole JL, Steen VD. The use of the Health Assessment Questionnaire (HAQ) to determine physical disability in systemic sclerosis. *Arthritis Care Res* 1991;4:27-31.
 162. Moser DK, Clements PJ, Brecht ML, Weiner SR. Predictors of psychosocial adjustment in systemic sclerosis. The influence of formal education level, functional ability, hardiness, uncertainty, and social support. *Arthritis Rheum* 1993;36:1398-405.
 163. Poole JL, Williams CA, Bloch DA, Hollak B, Spitz P. Concurrent validity of the Health Assessment Questionnaire Disability Index in scleroderma. *Arthritis Care Res* 1995;8:189-93.
 164. Malcarne VL, Greenbergs HL. Psychological adjustment to systemic sclerosis. *Arthritis Care Res* 1996;9:51-9.
 165. Merkel PA. Measurement of functional status, self-assessment, and psychological well-being in scleroderma. *Curr Opin Rheumatol* 1998;10:589-94.
 166. Clements PJ, Wong WK, Hurwitz EL, et al. Correlates of the disability index of the Health Assessment Questionnaire: a measure of functional impairment in systemic sclerosis. *Arthritis Rheum* 1999;42:2372-80.
 167. 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.
 168. Ruof J, Bruhlmann P, Michel BA, Stucki G. Development and validation of a self-administered systemic sclerosis questionnaire (SySQ). *Rheumatology (Oxford)* 1999;38:535-42.
 169. Silman A, Akesson A, Newman J, et al. Assessment of functional ability in patients with scleroderma: a proposed new disability assessment instrument. *J Rheumatol* 1998;25:79-83.
 170. Guillevin L, Ortonne JP. [Treatment of scleroderma]. *Ann Med Interne (Paris)* 1983;134:754-65.
 171. Friedman A, Fischbach M, Goel N, Aguilar M, Wan T, Reveille J. Predictors of self-reported function in a multi-ethnic scleroderma cohort [abstract]. *Arthritis Rheum* 1999;42 Suppl:187.
 172. Cossutta R, Masserini AB, Colombelli P. Evaluation of quality of life in patients with systemic sclerosis by the SF-36 questionnaire. Preliminary results [abstract]. *Arthritis Rheum* 1999;42 Suppl:186.
 173. Suarez-Almazor ME, Ahn C, Friedman AW, et al. Performance of outcome measures in systemic sclerosis [abstract]. *Arthritis Rheum* 2000;43 Suppl:319.