Arthritis Care & Research
Vol. 63, No. S11, November 2011, pp S98–S111
DOI 10.1002/acr.20598

Measures of Systemic Sclerosis (Scleroderma)

Health Assessment Questionnaire (HAQ) and Scleroderma HAQ (SHAQ), Physician- and Patient-Rated Global Assessments, Symptom Burden Index (SBI), University of California, Los Angeles, Scleroderma Clinical Trials Consortium Gastrointestinal Scale (UCLA SCTC GIT) 2.0, Baseline Dyspnea Index (BDI) and Transition Dyspnea Index (TDI) (Mahler's Index), Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR), and Raynaud's Condition Score (RCS)

JANET POPE

INTRODUCTION

Outcome measurements are important in evaluating patients with systemic sclerosis (SSc; scleroderma) and in research such as clinical trials, and many patient-reported outcomes can be useful for monitoring SSc patients seen in practice.

Previous research and consensus exercises have demonstrated important domains that may be useful in SSc clinical trials. These include skin, musculoskeletal, cardiac, pulmonary, cardiopulmonary, gastrointestinal, renal, Raynaud's phenomenon and digital ulcers, health-related quality of life and function, global health, and biomarkers (1,2). This review will focus on the Health Assessment Questionnaire (HAQ) disability index in SSc (3) and the Scleroderma HAQ (SHAQ) (4), physician and patient global assessments, the University of California, Los Angeles, Scleroderma Clinical Trials Consortium Gastrointestinal Scale (UCLA SCTC GIT) instrument for gastrointestinal involvement in SSc (5), the Raynaud's Condition Score (RCS) (6), Symptom Burden Index (7), the Cambridge Pulmonary Hypertension Outcome Review scale for pulmonary arterial hypertension (8), and some dyspnea scales for SSc lung disease, which have partially been validated in SSc.

Fatigue scales, general quality of life measurements, and several hand scales (Cochin Hand Function Scale, the Duruoz Hand Index, the Disabilities of the Arm, Shoulder, and Hand Questionnaire, Arthritis Hand Function Index, Italian Hand Mobility Scale, and the Delta Finger-to-Palm measure) will be discussed elsewhere in this supplement and are not part of the SSc review. In addition, outcomes used in the assessment of SSc for research and/or clinical care such as skin scores, pulmonary function tests, echocardiogram, functional class, 6-minute walk distance, renal measurements, digital ulcer burden, pulmonary imaging, inflammatory markers, joint counts, time to clinical worsening, and disease activity and damage scales are excluded from this review. Measurement of depression and comorbidities are value added in certain circumstances within SSc. All these instruments may be important to consider in a complete review of SSc measurement scales. The durometer and digital ulcer outcomes were not reviewed.

A literature search was performed on July 1, 2011 using PubMed for key words including validity, reliability, and questionnaire, and combining with HAQ, SHAQ, global assessments, SSc or scleroderma, GI outcomes, UCLA GIT, quality of life (QOL), Raynaud's Condition Score (RCS), and dyspnea scales for determining the characteristics of the outcome measurement tools that were within the scope of the study. The abstracts of original and review articles were read, and those thought to be relevant to this topic were fully read to extract instrument characteristics such as reliability, validity, and minimal important difference or minimum clinically important difference. Scales that are reported elsewhere in this supplement, such as general quality of life and other versions of the HAQ, were not reviewed, except for some modifications for SSc. Many clinical measures that are important in SSc, such as skin score and disease and damage indices, were not reviewed.

The results of the search included the following: n = 41 for scleroderma and questionnaire and validity, n = 29 for scleroderma and Health Assessment Questionnaire and validity, n = 14 for HAQ and scleroderma and reliability, n = 9 for HAQ and SSc and reliability, n = 0 for SHAQ and SSc and reliability, n = 1 for SHAQ and scleroderma and reliability, n = 14 for Scleroderma HAQ and scleroderma and reliability, n = 3 for UCLA GIT and SSc, n = 61 for

Janet Pope, MD, MPH, FRCPC: University of Western Ontario, London, Ontario, Canada.
Address correspondence to Janet Pope, MD, MPH, FRCPC, 268 Grosvenor Street, London, Ontario, N6A 4V2, Canada. E-mail: janet.pope@sjhc.london.on.ca.
Submitted for publication March 2, 2011; accepted in revised form August 2, 2011.

pain assessment and SSc, n = 100 for pain assessment and scleroderma, n = 52 for global assessment and scleroderma, n = 38 for global assessment and SSc, n = 12 for Raynaud condition score, and n = 19 for Raynaud's condition score. Symptom burden index and scleroderma resulted in 3 articles. Articles that discussed validation of the selected instruments were included if they were within the scope of this review and not reported elsewhere in this supplement.

MEASURES OF FUNCTION IN SYSTEMIC SCLEROSIS (SSc; SCLERODERMA)

HEALTH ASSESSMENT QUESTIONNAIRE (HAQ) AND SCLERODERMA HAQ (SHAQ)

This section will concentrate on the HAQ disability index (HAQ DI) within the context of SSc and SSc modifications.

Description

The HAQ DI is a self-reported questionnaire in 8 domains. The SHAQ consists of the HAQ (8 domains) and also includes the following scales: pain, patient global assessment, vascular, digital ulcers, lung involvement, and gastrointestinal involvement (4).

Purpose. The HAQ measures self-reported function and is one of the most commonly used quality of life measures in SSc. Due to the multisystem nature of SSc, it can greatly impact a patient's functioning and quality of life. Patient-centered outcomes are important in both clinical practice and research studies.

Developers. Fries et al developed the HAQ, which has been used extensively in SSc (3). Steen and Medsger added the visual analog scales (VAS) to the HAQ to create the SHAQ (4). An excellent review on the measurement properties of the HAQ and SHAQ has been published (9). The HAQ and SHAQ have been extensively studied for clinimetric properties in SSc (4,6,9,10–33).

Scoring. *HAQ.* Scoring is fast and each question is scored 0–3 (where 0 = without difficulty and 3 = unable to do) (3). There are 8 categories and the maximum from each category is added together and divided by the number of categories completed. There is an added point, to a maximum of 3, in each category if aids/devices are checked as being used (if the score is already at 3 or "unable to do," then the score cannot increase further).

SHAQ. The SHAQ is scored like the HAQ, and the other domains are continuous VAS instruments that are measured and then changed to a 0–3 scale. Each area is scored, and the scores are not added together for the VAS components. Therefore, a score can be 1.25 on the HAQ, 0.5 on pain, and separate scores for each of the other items.

Reliability. A therapist observed and graded the activities of the HAQ in patients with SSc. The therapist versus patient intraclass correlation coefficients (ICCs) ranged from 0.38–0.76, and there were significant differences between the observer and subjects' responses for 4 items, whereas the other items had moderate to good agreement (13). Since the HAQ and SHAQ are self-reported, the agreement between observed versus the patient does not need to be high since the instrument is not meant to be scored by an observer. Merkel et al demonstrated reliability of the HAQ and other VAS measures including Raynaud's phenomenon–specific scales in an analysis of data from a large Raynaud's phenomenon trial in SSc (6). Very good within-patient test–retest reliability, if stable SSc patients completed the HAQ once and then 2 days later, has been documented (33). Therefore, overall the HAQ in SSc seems reliable.

Validity. Cole et al compared the HAQ in SSc and early rheumatoid arthritis (ERA), and structural validity was demonstrated (25). HAQ scores have also been compared between established RA and SSc, where the HAQ is on average higher in SSc (21). Convergent and construct validity have been shown with a strong correlation (r = 0.9) between the HAQ and the UK Scleroderma Functional Score (UKFS) (17). There is face and content validity. The mean HAQ is higher in diffuse cutaneous (dcSSc) than limited cutaneous (lcSSc) (32). HAQ is correlated with skin scores, joint pain, tendon rubs, contractures, grip strength, thumb abduction, wrist extension, and motion of the index and middle fingers, and in some studies, the presence versus absence of digital ulcers, but not another (4,6,10,32). Higher HAQ scores are related to more work disability (21). HAQ and pain were found to be related to the physical component score of the Medical Outcomes Study Short Form 36 (SF-36; R = 0.70) in a cross-sectional study of 89 patients with SSc (27).

Predictive validity. A low baseline HAQ score is predictive of improving skin scores over the next year in 2 early dcSSc trials. There was a 1.5 to 5-fold chance of improvement in skin scores and patient global assessments if there was a low baseline HAQ at trial entry in early dcSSc (18). A low HAQ score is also predictive of improved patient global assessment the following year in SSc patients followed in a clinic setting (31). Another study found the HAQ to be correlated with skin, cardiac and renal involvement, tendon friction rubs, hand contractures, proximal muscle strength, and survival in SSc longitudinally (4).

Two large cohorts (one with ERA and the other established SSc) were compared, and there was structural validity in comparing the HAQ scores between the 2 groups (25). SSc patients with joint involvement had higher HAQ scores than in psoriatic arthritis, whereas pain was higher in SSc than RA (26).

Ability to detect change. In a 6-month randomized controlled trial of dcSSc, there was good agreement with the HAQ if skin score improved by at least 30%, with ICCs ranging from 0.69–0.91 (good to excellent). The SF-36 had a larger magnitude of responsiveness for physician and patient global assessments compared to the HAQ, whereas the HAQ was more responsive for skin score and the forced vital capacity on pulmonary function testing (19).

A change in digital ulcers status was related to a change in the HAQ and this was statistically significant when 2 digital ulcer trial results were combined in an exploratory analysis (29). A nonvalidated HAQ subscale score, which contained items primarily asking about finger function, demonstrated an improvement with treatment used to pre-

vent digital ulcers, but this was a post hoc analysis and using a part of the validated scale may not be appropriate, so this subscale needs further validation (28).

The minimum important difference (MID) of the HAQ has been calculated in clinical practice combining patients with lcSSc and dcSSc. MID estimates for improvement and worsening, respectively, were −0.0125 (for the mean, which is well below any measurement that is detectable, or a change of −0.125 if the 75th percentile was used, which has a more reasonable estimate) and 0.042 (for the mean change in the worsened group or worsening by 0.217 if the 75th percentile was used) within one SSc clinic between followup visits. In the Canadian Scleroderma Research Group, where patients have data collected annually, MID estimates for improvement and worsening were −0.037 (−0.250, 75th percentile) and 0.140 (0.375, 75th percentile), respectively (22). This method of MID calculation gives a HAQ that is well below a minimal change in the instrument (which is 0.125, so the 75th percentile was studied). This could be the case due to patients getting worse or better in areas that are not related to function, such as dyspnea or GI symptoms, and this would not be expected to affect the HAQ. Whereas, the MIDs in early dcSSc from the D-penicillamine trial were −0.10 for HAQ improvement and 0.14 for worsening (0.15–0.21 effect size) (23). The MID determined by a Delphi of SSc experts was estimated to be 0.2–0.25 for the HAQ (34). Therefore, depending on how the MID is calculated (or in the latter case estimated by experts to be relevant), the results may be different. It is likely that the MID of the HAQ in SSc, when function is changing, is in reality between 0.125 and 0.25 (1- or 2-point differences on the HAQ scale). Disability as measured by the HAQ worsens over time in SSc by 0.039 (95% confidence interval [95% CI] 0.018–0.061) to 0.071 (95% CI 0.048–0.094) per year, or at least on average 0.12 over 3 years (24).

SHAQ validity and reliability. When compared to the UKFS, the SHAQ had concurrent and convergent construct validity (17). It has face validity (4,6). Reliability of the HAQ and SHAQ VAS was demonstrated using data from a Raynaud's phenomenon trial (6).

Alternate scoring of SHAQ. In one study, the HAQ appeared more reliable than the SHAQ if the scales were added, but this is not routinely done. Using the French translation of the SHAQ, Georges et al proposed a combined score, obtained by pooling the 8 domains of the HAQ DI and the 5 VAS scales, and called it the SSc HAQ. However, this approach has not yet been widely accepted (35).

SHAQ predictive validity. The VAS subscales of the SHAQ were significantly correlated with objective parameters (4). Regarding convergent and construct validity, the SHAQ should have further face and content validity over the HAQ since it includes SSc-specific manifestations (9).

SHAQ responsiveness to change. The SHAQ was responsive to change in a cohort and in a Raynaud's phenomenon trial in SSc (4,6).

HAQ compared to SHAQ. The HAQ was compared to VAS scales of the SHAQ and the UKFS (17). The HAQ and UKFS were strongly correlated (r = 0.9) and both tools were significantly related to other clinical measures. Not surprisingly, the correlations with the VAS were not as strong since they were compared with a functional scale, which would not be as relevant to VAS scales of various organs or symptoms (17).

The SHAQ was no better than the HAQ in discriminating between lcSSc and dcSSc (20). The lung VAS has incremental concurrent validity over the HAQ as an outcome measure evaluating SSc lung disease (36).

Validation of HAQ for SSc and SHAQ in other languages. The SHAQ has been validated in French-speaking SSc patients for structural and convergent validity, with strong coefficients between the HAQ and the physical component score of the SF-36 ($r = -0.74$, $P < 0.0001$). Discriminant validity was found as the HAQ separated dcSSc and lcSSc (worse in the former). The test–retest reliability was excellent (r = 0.98) (36). The HAQ has been translated into Japanese. In Japanese SSc patients, the HAQ was related to many other clinical variables, especially hand extension. The mean HAQ was lower than what has been reported in US patients with SSc (37). The HAQ has face validity in Japanese (38). The HAQ has been translated into Italian, and significant differences in the HAQ were found in those with higher versus lower modified Rodnan skin thickness scores (above and below 14 units, mean ± SD HAQ in former of 1.158 ± 0.176 versus 0.652 ± 0.076; $P < 0.001$). When present, other clinical features (contractures, myopathy, and digital ulcers) had higher HAQ scores than if absent (39).

Critical Appraisal of Overall Value to the Rheumatology Community

Scoring the HAQ includes adding aids and devices. However, in SSc a change from a low HAQ denoting little or no disability to moderate disability occurs when aids/devices are scored in the total HAQ score in the latter but not the former scenario. This must be taken into account when describing cohorts or trials and the method of scoring (with or without aids and devices should be stated) (40). However, the usual scoring is adding the aids and devices as mentioned above.

The HAQ is widely used, inexpensive, and takes only a couple of minutes to complete and score. It has been translated into many languages, with some validation in SSc in other languages. The HAQ is somewhat outdated and may not apply to patients in different countries (such as opening a milk carton, lifting a certain amount in pounds, taking a tub bath, etc.).

OTHER MEASURES OF FUNCTION IN SSc

There are strong correlations between the UK Scleroderma Functional Score (UKFS) and the Health Assessment Questionnaire (HAQ) in a cross-sectional SSc sample ($\rho = 0.90$, $P < 0.0001$) and prospectively with change over time comparing UKFS and HAQ ($\rho = 0.59$, $P < 0.0001$) (41). The Functional Index, which is an 11-item scale, has not been widely used in SSc (42).

The Scleroderma Assessment Questionnaire is a self-assessed measure ranging from 0–3 for several questions including vascular, respiratory, gastrointestinal, musculo-

skeletal, and overall disease status with 23 questions divided into 4 groups. The questions include the HAQ and other questions (43,44). It has face validity and is sensitive to change, but currently is not commonly used. Another measure that has not been validated is the Systemic Sclerosis Questionnaire, which includes general, organ-specific, and musculoskeletal complaints (45). There are other proposed functional scales (46–48).

GLOBAL ASSESSMENTS IN SYSTEMIC SCLEROSIS (SSc; SCLERODERMA)

PHYSICIAN- AND PATIENT-RATED GLOBAL ASSESSMENTS

Description

Global assessments are rated by the observer (usually a physician) or the patient. They may be rated from 0–100 with a continuous visual analog scale (VAS) or a Likert scale with, for example, a 10-point rating or a change scale, such as a 7-point scale (from −3 to +3 including 0 in the center for no change). There is no standardization for the scale, but usually a low number indicates less disease activity. They can also rate severity, damage, or overall disease. Neither the question or the recall period is standardized (e.g., a global assessment may ask the patient to rate overall disease activity or the way that SSc has affected her/him over the last month, week, or today since the last visit).

Purpose. The global assessments are very easy and are used in both clinical practice and research studies to quantify the disease activity or severity (or whatever is being asked). The most frequent rating is SSc overall disease activity.

Content. There may be one scale (e.g., an overall global assessment) or several scales (e.g., organ areas or Raynaud's phenomenon symptoms).

Developer. None.

Number of items in scale. There may be 1 global assessment each for the patient and physician to complete, and/or an additional series of global assessments of each organ system. There may be questions with respect to activity, damage, and severity.

Scoring. Scoring is easy and may be a 10-cm or 15-cm VAS and converted to 0–100 for the former or 0–3 for the latter. Likert scales do not usually have measurement in between the numbers provided such as 0,1,2–9,10. It takes only a few seconds to complete and to score.

Reliability. Test–retest reliability has been calculated for physician and patient global assessments (49). Measures with inherent interpretation such as global assessments and skin scores were found to have increased variability than easily-performed measurements such as grip strength and oral opening. However, there was good reproducibility within observers, but moderate between observers' intraclass correlation coefficients (ICCs). The patient global assessment has very good reproducibility (33). The relatedness of physician global assessments in SSc for disease activity, severity, and damage has been calculated ranging from $\rho = 0.77$ for severity and activity to $\rho = 0.61$ for damage and activity (50).

Validity. There is some face validity since a patient or rater is asked to determine overall disease activity or severity. The question may be open to interpretation. The ratings are different between the patient and the physician, so they are measuring different things. There are higher scores in general for diffuse cutaneous SSc (dcSSc) versus limited cutaneous SSc (lcSSc) (50). A large Canadian database (Canadian Scleroderma Research Group) and a Michigan database demonstrated that there is discordance between the patient and physician global assessments. Patients perceived greater disease severity than physicians (mean ± SD difference 0.78 ± 2.65). The agreement between patient and physician assessments of disease severity was modest (ICC 0.38, weighted $\kappa = 0.38$). Both patient and physician scales were related to skin scores, breathlessness, and pain, but the relative importance of these predictors differed. Patients were also influenced by other subjective symptoms, while physicians were also influenced by disease duration and creatinine. The predictors explained 56% of the deviance in the patient global assessments and 29% in the physician assessments (51). This makes sense as they are not measuring the same things and are both necessary end points for measuring the status of a patient with SSc. Disease activity was rated higher for dcSSc (especially early dcSSc) compared to lcSSc (50).

Sensitivity to change. Low Health Assessment Questionnaire (HAQ) scores predicted improvement in the physician global assessment in clinical practice over 1 year (31). In the D-penicillamine trial, multivariate logistic regression demonstrated that the physician's global assessment of improvement was best explained by a model with skin score and HAQ ($R^2 = 0.46$) (52). Skin scores and patient global assessments were correlated with improvement in 2 early dcSSc trials (r = between 0.25 and 0.35) (18). The minimal important differences for patient global assessment in SSc have been calculated in clinical practice and are very small (4–6.7 on 100-mm VAS) (22). Minimum clinically relevant important differences from a physician's perspective, obtained by expert opinion and Delphi consensus, were 3–7.5 units of the modified Rodnan skin thickness score, 8–13 for physician global (out of 100), and, similarly, 10–12 for patient global assessment (out of 100) (34).

Critical Appraisal of Overall Value to the Rheumatology Community

Global measurements are important and easy to use. They have some validity and give a rating to what the patient perceives as important for a patient-reported global assessment. It is difficult for a patient living with a complex disease such as SSc to know the difference between disease activity and damage, so the physician and patient global assessments may be very different (51). The global assessment question depends on what is asked, since standardization is lacking and different questions may have variable sensitivity to change, i.e., severity may or may not change, damage will potentially worsen over time but not

improve, and disease activity within a study can change for the better or worse or remain stable. In addition, patients and physicians are not measuring the same things or they are weighing them differently and therefore differences between patient and physician global ratings do occur (51). The physician assessor may score serious organ involvement that is active as worse than mild complications, whereas a patient rates what they are feeling and may score a digital ulcer or gastrointestinal problem higher than a serious organ involvement, especially if the latter has minimal symptoms. Continuous and Likert scales are not completely interchangeable.

PAIN ASSESSMENTS IN SSc

Description

Pain in SSc is most often assessed by visual analog scales (VAS), Likert, or change scales. Questions are not different from other generic pain scales except there could be attribution, such as overall pain from SSc, Raynaud's phenomenon (RP) pain, or digital ulcer pain, or there is no attribution (such as overall pain compared to attribution asking about pain from SSc or from a specific problem such as RP, digital ulcers, skin involvement, arthritis, or gastrointestinal [GI] problems).

Purpose. Self-administered scale to rate SSc-related pain.

Content. Scale or scales on pain.

Developer. None.

Number of items in scale. Single item, part of other scales (such as overall disease, overall pain, or digital ulcer pain), or part of a multiquestion pain questionnaire. The Raynaud's Condition Score (RCS) and Symptom Burden Index contain pain questions. There may be subscales of various pain areas (from RP, ulcers, GI, overall, etc.).

Scoring. The Health Assessment Questionnaire (HAQ) pain scale is 15 cm with a conversion from 0 (no pain) to 3 (100% pain). The measured number on the scale is divided by 5 cm to make it be scaled appropriately for the HAQ pain scale. Other scales could be numbers, Likert, or 100-mm VAS. Completion should be less than a minute. Scoring is also fast.

Reliability. Pain is correlated with other patient-reported outcomes of disease activity in RP with SSc (6). Pain is very common in SSc with mean pain of ~40 out of 100 in a clinic setting (22). In a large SSc study, 83% of patients had pain, half of whom had mild pain (1–4 out of 10); one quarter had moderate pain (5–7), and 10 had severe pain (8–10). More frequent RP attacks, active ulcers, worse synovitis, and GI symptoms were associated with pain. Overall pain was worse in diffuse cutaneous SSc (53). The modified Rodnan skin thickness score is strongly associated with pain (54). Higher pain scores are also associated with more alteration of body image in SSc (55).

Validity. The pain scale is validated in SSc (content) alone and with the HAQ or RCS (4,6). The mean ± SD pain in SSc patients is 41 ± 26 out of 100 (22).

Sensitivity to change. In many effective therapies for RP in SSc, the pain scale improves (56). The minimum important difference (MID) for pain in SSc clinical practice on a 100-mm VAS is from 3.6–8 for pain (22). Physicians perceived the MID for pain in SSc to be 0.2–0.3 (out of 3) (34).

Critical Appraisal of Overall Value to the Rheumatology Community

Pain is likely under-recognized in SSc and is important to measure. There is a lack of standardization for the time frame and actual question in SSc with respect to pain. There can be overall pain as well as organ-specific pain questions, and scales can be 100-mm, 15-cm, Likert, and even descriptive. The same limitations of global assessments apply to the assessment of pain. Pain and attribution from disease under study or other problems are very difficult for patients, and some pain questions may be about disease-related pain while another scale may be about overall pain. Therefore, mechanical back pain would not be included in the former, but it would in the latter. However, the test–retest reliability should not be affected in either scenario, but the attribution to SSc is not necessarily present in a question that asks about overall pain. In addition, even if asked about SSc-specific pain, many patients rank all their pain as they do from their disease. It has been found that patients have problems distinguishing SSc from other comorbidities (57). Pain can be from disease activity or damage and therefore may not be responsive to treatment.

FATIGUE

Fatigue is a very common complaint in SSc, but there are no specific SSc fatigue scales. As with other rheumatic diseases, pain and fatigue are significant determinants of quality of life in SSc (58). The minimal important difference for fatigue in SSc clinical practice is from 3.8–10.0 out of 100, and a sleep problems visual analog scale was from 5.9–18.5 (22). A detailed review of fatigue scales has recently been completed (59).

SYSTEMIC SCLEROSIS (SSc; SCLERODERMA)–SPECIFIC MEASURES OF QUALITY OF LIFE

SYMPTOM BURDEN INDEX (SBI)

This is a self-reported questionnaire for SSc.

Description

The SBI was developed to determine the effects of SSc in several domains that impact quality of life (QOL) (7). The SBI has 8 major symptomatic areas (skin, hand mobility, calcinosis, shortness of breath, eating, bowel, sleep, and pain) (7).

Purpose. The SBI determines the effects of SSc from a patient's perspective in several domains beyond physical function and generic health-related QOL instruments. It is a patient-reported instrument, measuring burden of illness in SSc (7).

Content. The domains consist of several areas with 5 questions in each domain with Likert scales for each question.

Developer. M. A. Kallen and Maria E. Suarez-Almazor, University of Texas.

Number of items in scale. Eight major symptomatic areas of importance to patients are included (skin, hand mobility, calcinosis, shortness of breath, eating, bowel, sleep, and pain), with 5 items each per area with a 0–10 Likert scale. The questions are based on how much, how often, how much interference, how often interfering, and how important is this to the patient.

Scoring. There are 40 questions (5 questions in each of 8 domains). Scoring is done for each scale with the average burden score reported per problem area (in 8 domains) on an 11-point scale (from 0–10). The SBI also gives the number of patients experiencing each SSc-related problem in a group of SSc patients and the number of problems experienced by each patient.

Reliability. Inter-item and item-total score correlations per item set were all moderate to high, and internal consistency reliability estimates were high. These scale characteristics reflect the small to moderate item score ranges observed per item set from 0.4–2.2 (7). Patients had a mean of 5.7 problems with one-third having 0–5 problems and another one-third having 7 or 8 problems in the total of 8 domains.

Validity. The SBI is partially validated in a single site study with 62 SSc patients. Scores in each domain correlated with the Health Assessment Questionnaire (HAQ) and Medical Outcomes Study Short Form 36 (SF-36). For the HAQ, correlations of each SBI scale ranged from 0.3–0.6 and were statistically significant. For the SF-36, higher SBI scores negatively correlated with the SF-36 for both the mental and physical components (7).

Construct validity. Focus groups were tested in order to develop the domains of importance to patients with SSc (60). Except for a few correlations comparing shortness of breath to other domains, all other domains were statistically significantly related to the other domains with low to moderate correlations. However, the burden scores across problems were relatively independent.

Sensitivity to change. Sensitivity to change has not yet been demonstrated.

Translations/adaptations. This has not yet been done.

Critical Appraisal of Overall Value to the Rheumatology Community

The 8 problem areas each have a score and are somewhat independent from the other problems, and SBI scores correlate with the HAQ and SF-36. Especially pain (localized or generalized), fatigue, and malaise were reported to have a major influence on QOL. Gastrointestinal (GI) symptoms were prevalent and had high scores. This could potentially be used in clinical practice and in research. The SBI should be further studied in other cohorts and sensitivity to change is important to determine if the SBI will be used in the future as an outcome measurement in treatment trials. The Scleroderma HAQ visual analog scales for GI, lung, and pain have not been compared with the SBI, where one would expect very strong correlations with the respective scales.

GASTROINTESTINAL (GI) SCALES

The GI tract is a source of considerable discomfort, morbidity, and mortality in patients with SSc. The approach to GI tract–related outcome measures logically follows the pathogenesis, including dysmotility (dysphagia, early satiety, bloating, small bowel bacterial overgrowth, and malabsorption), patent lower esophageal sphincter with gastroesophageal reflux disease, watermelon stomach causing anemia, and obstipation or constipation from large bowel dysmotility, etc. Measures have been validated for manometry and for esophageal and gastric transit time, but these measures may not change significantly in the timeframe of SSc trials. Often, measures are used that have been successful in other GI diseases.

UNIVERSITY OF CALIFORNIA, LOS ANGELES, SCLERODERMA CLINICAL TRIALS CONSORTIUM GASTROINTESTINAL SCALE (UCLA SCTC GIT) 2.0

Description

Khanna et al have validated and improved upon the SSc GIT 1.0, shortening it to the UCLA SCTC GIT 2.0 instrument, which can potentially be used as an outcome for randomized controlled trials in SSc-associated gastrointestinal (GI) involvement (5,61). This is a 7–multi-item scale with areas of reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning and has been shown to have a good test–retest reliability (5).

Purpose. To have a self-reported GI quality of life (QOL) tool specifically for the range of problems that can occur in SSc and to be able to score the instrument, looking for changes over time or within a trial.

Developer. Dinesh Khanna, et al. University of California, Los Angeles.

Number of items in scale. There are 34 items in the UCLA SCTC GIT 2.0 instrument. The 7 multi-item scales include reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning.

Scoring. Version 2.0 consists of 34 items scored from 0–3, with lower values indicating better health-related (HR) QOL. The total UCLA SCTC GIT 2.0 score averages 6 of 7 scales (excluding constipation) and is scored from 0 (no GI problems) to 3 (most severe).

Reliability. Test–retest reliability estimates were $\geq$0.68 (5).

Validity. Self-rated severity of GI involvement has spanned no symptoms to very mild (39%), mild (21%), moderate (31%), and severe/very severe (9%) (5). It is also related to poor sleep (62).

Discriminant validity. Symptom scales were also able to discriminate subjects with corresponding clinical GI diagnoses. The total UCLA SCTC GIT 2.0 score, developed

by averaging 6 of 7 scales (excluding constipation), was reliable and provided greater discrimination between mild, moderate, and severe self-rated GI involvement than individual scales.

The 2.0 version was developed using the 52 items from the SSC-GIT 1.0 and 1 rectal incontinence item, grouped into 8 scales based on content: reflux, distention/bloating, diarrhea, fecal soilage (to assess rectal incontinence), constipation, pain, emotional well-being, and social functioning (5). Version 2.0 contains 34 items and is scored from 0–3, with lower values indicating better HRQOL. Therefore, in version 2.0, 7 multi-item scales (reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning) are included. The UCLA SCTC GIT 2.0 instrument was correlated with depression (except for the parameter of fecal soilage) (63).

Critical Appraisal of Overall Value to the Rheumatology Community

The UCLA SCTC GIT 2.0 scale has been partially validated. It is unknown if this will be sensitive to change in a GI treatment trial. In patients with very frequent symptoms, a moderate improvement may not be detected since the symptoms could still be frequent even if occurring far less often. No comparison of the UCLA SCTC GIT 2.0 questionnaire and the GI visual analog scale on the Scleroderma Health Assessment Questionnaire was found in the literature search.

OTHER GI SCALES IN SSc

Another scale that is not SSc-specific is the Gastrointestinal Quality of Life Index (64), which is a validated 52-item questionnaire capturing SSc-related gut dysfunction given to more than 400 SSc clinic patients assessing the frequency and impact of 5 categories of symptoms. There was a positive correlation between diarrhea scores and pulmonary fibrosis ($r = 0.13$), but not with other organs. In addition, limited cutaneous SSc and diffuse cutaneous SSc did not score differently; this is expected for GI disease in SSc, which is virtually universal (65).

GASTROESOPHAGEAL REFLUX DISEASE (GERD)

GERD is extremely common in SSc and is often severe. There are scales that have been used in SSc that assess GERD, such as the Frequency Scale for the Symptoms of GERD (FSSG), and visual analog scales (66). One study in severe GERD used gut pH measurements, which did not differentiate active treatment from placebo, but the study was negative with respect to the FSSG and quality of life (67). This study compared ranitidine to placebo on background double-dose proton pump inhibitors for severe GERD in SSc patients. Therefore, we cannot conclude if pH measurements of the gut are useful as an outcome in SSc randomized controlled trials. The testing is invasive and needs training to be performed. The trial was also likely underpowered.

DYSPNEA MEASUREMENTS USED IN SSc-ASSOCIATED INTERSTITIAL LUNG DISEASE (ILD)/PULMONARY FIBROSIS

There is no fully validated dyspnea questionnaire in SSc. In addition, the quality of life (QOL) in SSc patients with ILD may be impacted by cough, which is not captured on questionnaires that have been studied in SSc. Numerous dyspnea scales have been published in other diseases such as chronic obstructive pulmonary disease (COPD) and idiopathic pulmonary fibrosis.

The dyspnea questionnaire by Mahler et al includes the Baseline Dyspnea Index (BDI) and Transition Dyspnea Index (TDI) (68). The Modified Medical Research Council Scale and the Oxygen Cost Diagram are widely used tools for evaluation of limitation of activities due to dyspnea that are used in COPD (68–72) but not SSc-associated ILD. There is an activity of daily living dyspnea scale, the Modified Dyspnea Index, and dyspnea scales from the Scleroderma Health Assessment Questionnaire (HAQ) and the Symptom Burden Index (4,7,73). Dyspnea is significantly related to function and QOL. A model including age, sex, disease duration, disease severity, and dyspnea explained one-third, 10%, 40%, and 30%, respectively, of the variance of the HAQ (74). The BDI and TDI from the questionnaire by Mahler et al (68) will be reviewed more extensively since being used in a SSc lung disease trial of cyclophoshamide, and there has been some validation in SSc of these instruments (75). The Borg Dyspnea Index is a measurement by which dyspnea is assessed following the 6-minute walk and has only been partially validated in ILD and pulmonary hypertension (76). The Modified Borg Dyspnea Scale is numerical and describes the severity of dyspnea.

BASELINE DYSPNEA INDEX (BDI) AND TRANSITION DYSPNEA INDEX (TDI) (MAHLER'S INDEX)

Description

The BDI and TDI measure dyspnea at one point in time and then how it has changed at another time point (68). It can be self- or interviewer administered.

Purpose. To measure the severity of dyspnea at a point in time and in followup to determine if there is change (improving, the same, or worsening), as well as to evaluate the severity of dyspnea as the changes are added to the baseline score.

Content. For Mahler's dyspnea scales, the BDI is designed to rate the severity of dyspnea at a single time point, and the TDI is designed to capture a change (or no change) from the baseline assessment. Each index rates 3 different categories: magnitude of task, magnitude of effort, and functional impairment. Each category has 5 grades ranging from 0 (severe) to 4 (unimpaired) added together for a baseline focal score (range 0–12). At the transition period, changes in dyspnea were rated by 7 grades, ranging from −3 (major deterioration) to +3 (major improvement). The ratings for each of the 3 categories for the TDI were added to form a transition focal score (range −9 to +9) (68).

Developer. D. A. Mahler, D. H. Weinberg, C. K. Wells, and A. R. Feinstein.

Number of items in scale. Each index rates 3 different categories: magnitude of task, magnitude of effort, and functional impairment. Each category has 5 grades ranging from 0 (severe) to 4 (unimpaired). At the transition period, changes in dyspnea were rated by 7 grades, ranging from −3 (major deterioration) to +3 (major improvement).

Scoring. Each category is added together for a baseline score (range 0–12) as a maximum of 4 for each of 3 scales. For the TDI, the ratings for each of the 3 categories were added to form a transition focal score (range −9 to +9). TDI has improvement as major, moderate, or minor corresponding to improvement on the scale as 7–9, 4–6, and 1–3, respectively, and conversely there is deterioration if the scales show worsening (−1 to −3 for minor, −4 to −6 for moderate, and −7 to −9 for severe).

Reliability. This has not been fully tested in SSc, but the instrument was used successfully in a scleroderma lung study (SLS) using cyclophosphamide versus placebo (75).

Validity. The original indices were validated in men, most of whom had chronic obstructive pulmonary disease (68).

Construct validity. Not fully tested in SSc; in the SLS, baseline scores of the BDI and visual analog scale (VAS) for breathing were highly correlated ($r = -0.61$). Medical Outcomes Study Short Form 36 (SF-36) scores were able to differentiate patients with more breathlessness (measured by BDI and VAS for breathing) (77).

Face validity. In the SLS, there was face validity for the Mahler dyspnea scale, since a larger proportion of patients treated with cyclophosphamide obtained at least the minimum important difference (MID) compared to placebo in the TDI (78).

Predictive validity. Not tested in SSc.

Sensitivity to change. Using the SF-36 transition question and defining the MID as patients who rated themselves as a little better or a little worse in the SLS, the MID was estimated for the TDI. TDI improvement and worsening, respectively, ranged from 1.05 to 2.16 (mean 1.5) and from −0.61 to −2.55 (mean −1.5) (79). More patients on cyclophosphamide achieved a MID for the TDI (46% for cyclophosphamide versus 13% for placebo) (78). The mean TDI change was higher in the cyclophosphamide group (75). Other measurements such as changes in fibrosis on high-resolution computed tomography were associated with changes in dyspnea (80).

OTHER LUNG SCALES IN SSc

There is face validity of the Saint George's Respiratory Questionnaire (81) in the evaluation of the health-related quality of life in SSc associated with interstitial lung disease (82). An exercise program in SSc found that a significant proportion of patients with SSc experienced an improvement in the Saint George's Respiratory Questionnaire and exercise tolerance (83).

PULMONARY ARTERIAL HYPERTENSION (PAH) IN SSc

PAH randomized controlled trials often do not measure a dyspnea questionnaire. Therefore, standardized dyspnea questionnaires may or may not be sensitive to change in SSc-associated PAH. A Delphi exercise for PAH in SSc suggested the domains should include lung vascular, exercise testing, cardiac function, dyspnea (as measured by a visual analog scale [VAS]), discontinuation of treatment, quality of life, and physician global assessment. These could be measured by right heart catheterization, echocardiography, exercise oxygen saturation, 6-minute walk distance, Medical Outcomes Study Short Form 36, the Health Assessment Questionnaire (HAQ), and survival as well as adverse events (84).

The New York Heart Association and World Health Organization functional class systems are essentially the same and are divided into 4 categories: no restriction of activities (class I), mild restriction (class II), moderate (class III), and severe inability to do activities of daily living with dyspnea even at rest (class IV) (85,86). There is a large potential range of severity in class II and III patients, so refining a dyspnea questionnaire would be valuable. There is a lack of correlation between the HAQ in SSc and PAH with respect to functional class at baseline and with treatment (87). The SSc community via a Delphi exercise rated that outcome measurements in SSc PAH should include severity of dyspnea measured on a VAS (84).

PULMONARY ARTERIAL HYPERTENSION (PAH)–SPECIFIC QUALITY OF LIFE INSTRUMENT: CAMBRIDGE PULMONARY HYPERTENSION OUTCOME REVIEW (CAMPHOR)

Description

CAMPHOR is a PAH-specific quality of life (QOL) measure and not specifically for SSc. It is the first pulmonary hypertension–specific instrument for assessing patient-reported symptoms, functioning, and QOL, with scales including overall symptoms (made up of energy, breathlessness, and mood subscales), functioning and QOL. This has not been validated specifically in SSc-associated PAH (8).

Purpose. This instrument is to be used in PAH to assess QOL. It should quantify the effects of PAH on QOL, assessing impairment, disability, and needs-based QOL.

Content. Questions about symptoms, function, energy, mood, breathlessness, and QOL.

Developer. Galen Research (S. P. McKenna, N. Doughty, D. M. Meads, L. C. Doward, and J. Pepke-Zaba).

Number of items in scale. CAMPHOR has 3 scales including overall symptoms, functioning, and quality of life with 65 items (8). The overall symptoms category has the subscales of energy, breathlessness, and mood. The instrument consists of 25 items for impairment, 15 for functioning, and 25 for QOL.

Reliability. The CAMPHOR scales have good reproducibility (0.86–0.92 for test–retest correlations) when tested in idiopathic PAH (8), but it has not been tested in SSc.

Validity. The CAMPHOR scales have very good internal consistency (α = 0.90–0.92) (7).

Face validity. The CAMPHOR utility score appears better able to distinguish between World Health Organization functional classes (II and III) than the EuroQol 5-domain and Short Form 6D (88).

Construct validity. The CAMPHOR scales have convergent, divergent, and known-groups validity (8).

Predictive validity. Patients remaining in the New York Heart Association (NYHA) class III experienced, on average, a significant improvement (CAMPHOR Utility Index and functioning), which exceeded the minimum important difference (MID) when PAH was treated (89).

Sensitivity to change. The CAMPHOR Utility Index has face validity and is responsive to change in PAH, but is not SSc specific. Patients can experience meaningful improvement even if they do not improve on functional class (which could require a larger change in status), and the CAMPHOR Utility Index distinguished between adjacent NYHA classes and correlated with 6-minute walking test (6MWT) results. CAMPHOR subscales and utility were as responsive as the 6MWT (effect sizes range 0.31–0.69 for the CAMPHOR and 0.16–0.34 for the 6MWT). The within-group MID for the CAMPHOR Utility Index is 0.09 (89).

Translations/adaptations. CAMPHOR has been validated in the US (90). It has also been adapted to French Canadian and English Canadian (91). There is also a version for English in Australia.

RAYNAUD'S PHENOMENON (RP) AND DIGITAL ULCERS

Raynaud's Condition Score has been validated in RP associated with SSc and will be discussed in detail, whereas there will only be a brief overview of DU.

DIGITAL ULCERS (DU)

It has been suggested that core measures for Raynaud's phenomenon (RP) in SSc clinical trials contain the Raynaud's Condition Score, physician and patient global assessments of RP activity, a DU measure, measures of disability and pain (Health Assessment Questionnaire), and measures of psychological function/generic quality of life (Arthritis Impact Measure 2 or Medical Outcomes Study Short Form 36) (6). Outcomes in DU are not standardized. There is no consensus on which DU measurements should be included in SSc DU trials. DU assessments may include visual analog scale (VAS) for RP, DU pain scales, number of digital ulcers, size of DU, burden of DU, healing or partial healing of DU (all or a cardinal ulcer), prevention of new DU, proportion of patients affected by DU, mean number of DU per patient, and VAS for physician and patient global assessments (92). As would be expected, within a 3-month SSc ulcer study there were significant correlations between ulcer dimension and pain VAS (r = 0.42, $P < 0.001$) (93).

RAYNAUD'S CONDITION SCORE (RCS)

Description

The RCS is a self-reported global assessment of Raynaud's phenomenon (RP) activity using a 0–10 ordinal scale, which incorporates the cumulative daily frequency, duration, severity, and impact of RP attacks. A composite score from daily measures is then calculated (6).

Purpose. To estimate the overall effect of RP.

Content. The RCS is a daily self-assessment of RP activity using a 0–10 ordinal scale that incorporates the cumulative daily frequency, duration, severity, and impact of RP attacks.

Developer. Peter Merkel, et al. Boston University.

Number of items in scale. One item with an 11-point ordinal scale (0–10), completed on a daily basis, and then an overall summary score is calculated for a defined period of time. There are no subscales.

Scoring. The number on the ordinal scale completed daily is added and divided by the number of days it has been completed to get a mean RCS for a period of time.

Reliability. It was found to be reliable when analyzing data from a trial (94).

Validity. In a recent randomized controlled trial (RCT) of tadalafil versus placebo, improvement in the frequency and duration of RP, RCS, healing and number of new digital ulcers (DUs), Scleroderma Health Assessment Questionnaire (SHAQ), and patient and physician global assessments significantly improved with active treatment (94).

Construct validity. Merkel et al have demonstrated the construct, content, criterion, and discriminant validity of the RCS, HAQ, and 12 visual analog scales (VAS) for RP in scleroderma using data from a RP RCT (6). There were relevant associations between the outcome measures and the patient and physician global assessments of RP activity.

Predictive validity. RCS can discriminate between those with and without DUs (6).

Sensitivity to change. RCS has been studied to determine the change needed to be clinically relevant in an RP trial. The minimum important difference score for the RCS for improvement is from −13.9 to −14.3 points (95). The patient acceptable symptom state was 34 (scale 0–100).

Critical Appraisal of Overall Value to the Rheumatology Community

Many clinicians and even researchers do not routinely use or interpret the RCS. In trials, it is often performed in addition to recording the frequency, severity, and duration of attacks. Therefore, the added value of the RCS is not fully determined. There is an advantage if a day of data are missing in an RP trial, since the score can still be calculated with the data that are completed, whereas if a day is missed then the frequency and duration of RP over 2 weeks cannot be calculated. There is a theoretical advantage to having a single scale that incorporates the impact of RP. Confusion between a 0–10 ordinal scale summary score of RCS and other Raynaud's scales may occur; how-

ever, the RCS is labeled accordingly whereas the other scales are often continuous VAS scores for RP.

HAND FUNCTION INSTRUMENTS FOR SSc

A detailed review for hand function scales has been performed elsewhere (96). However, some SSc studies related to hand function will be briefly reviewed. The Cochin Hand Function Scale has had good construct validity and its total score explained 75% of the variance of the Health Assessment Questionnaire in SSc patients (20). The Duruoz Hand Index was studied for test–retest reliability and intraclass correlation coefficients were very good (0.81–0.97) (48). The UK SSc Functional Score (17) and the Michigan Hand Questionnaire (97) also measure hand function. The latter may not be very useful for SSc hand function and has been rarely used for digital ulcer assessment (92).

THE MOUTH HANDICAP SCALE IN SSc

The majority of patients with SSc have oral problems including reduced oral opening, difficulties with dry mouth, and functional impairment with respect to oral hygiene. Mouthon et al have published the Mouth Handicap Scale in Systemic Sclerosis (98). It has 12 items with each scored from 0–4 (total score range 0–48). The mean ± SD total score of the scale was 20.3 ± 9.7. The test–retest reliability was 0.96. Divergent validity was confirmed for global disability (Health Assessment Questionnaire; r = 0.33), hand function (Cochin Hand Function Scale; r = 0.37), interincisor distance (r = −0.34), handicap (McMaster-Toronto Arthritis questionnaire; r = 0.24), depression and anxiety using the Hospital Anxiety and Depression (r = 0.26 and r = 0.17 for depression and anxiety, respectively). Three factors within the scale could explain two-thirds of the variance (98).

DISCUSSION

Many important instruments were not discussed in this review. Also, some articles may have been missed by the search strategy. Validation and reliability testing varied, where in some instruments (such as Raynaud's Control Score [RCS]) it was tested within a randomized controlled trial (RCT). For others there was cross-sectional testing at a single site (Symptom Burden Index). The University of California, Los Angeles, Scleroderma Clinical Trials Consortium Gastrointestinal Scale will likely be used within a treatment trial to determine its sensitivity to change. The Cambridge Pulmonary Hypertension Outcome Review needs validation in systemic sclerosis (SSc) if it is to be used for pulmonary arterial hypertension in SSc. However, for the instruments that were included, many have been partially validated in SSc, which is important for future research. Some lack testing for sensitivity to change. The global assessments (as in any rheumatic disease) do not have standardized questions or time frames but have been found to be sensitive to change within studies. There are also validated measurements that are not completed by the patients that are valuable in routine care and trials, such as the Modified Rodnan Skin Thickness Score.

There are also differences in minimum important differences (MIDs) when comparing how they were derived, such as in the Health Assessment Questionnaire (HAQ) (22,23,34). The MID in a trial of early diffuse cutaneous SSc (dcSSc) is not the same as expert determined. In the clinic with limited cutaneous SSc and dcSSc patients, many of whom may have been relatively stable, and in the latter methodology, the mean change in HAQ did not make sense since it was below the limit of the scale to detect change (22). This could also illustrate that patients may be worse with SSc that is unrelated to worsening function (and due to symptoms in other domains such as lung, gastrointestinal, Raynaud's phenomenon, etc.).

In addition, for use as outcomes in clinical trials, the sample size calculations can be different for instruments such as the HAQ, functional index, and physician global assessment due to variability in the measures in a group of SSc patients (99). This is important when selecting outcome measurements in clinical trials since some may be more apt to change within a given sample size.

AUTHOR CONTRIBUTIONS

Dr. Pope drafted the article, revised it critically for important intellectual content, and approved the final version to be published.

REFERENCES

1. Khanna D, Lovell DJ, Giannini E, Clements PJ, Merkel PA, Seibold JR, et al. Development of a provisional core set of response measures for clinical trials of systemic sclerosis. Ann Rheum Dis 2008;67:703–9.
2. Khanna D, Distler O, Avouac J, Behrens F, Clements PJ, Denton C, et al. Measures of response in clinical trials of systemic sclerosis: the Combined Response Index for Systemic Sclerosis (CRISS) and Outcome Measures in Pulmonary Arterial Hypertension related to Systemic Sclerosis (EPOSS). J Rheumatol 2009;36:2356–61.
3. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
4. 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.
5. Khanna D, Hays RD, Maranian P, Seibold JR, Impens A, Mayes MD, et al. Reliability and validity of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. Arthritis Rheum 2009;61:1257–63.
6. Merkel PA, Herlyn K, Martin RW, Anderson JJ, Mayes MD, Bell P, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. Arthritis Rheum 2002;46:2410–20.
7. Kallen MA, Mayes MD, Kriseman YL, de Achaval SB, Cox VL, Suarez-Almazor ME. The symptom burden index: development and initial findings from use with patients with systemic sclerosis. J Rheumatol 2010;37:1692–8.
8. McKenna SP, Doughty N, Meads DM, Doward LC, Pepke-Zaba J. The Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR): a measure of health-related quality of life and quality of life for patients with pulmonary hypertension. Quality Life Res 2006;15:103–15.
9. Johnson SR, Hawker GA, Davis AM. The Health Assessment Questionnaire disability index and Scleroderma Health Assessment Questionnaire in Scleroderma Trials: an evaluation of their measurement properties. Arthritis Rheum 2005;53:256–62.
10. 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.
11. Clements PJ, Lachenbruch PA, Seibold JR, Zee B, Steen VD, Brennan P, et al. Skin thickness score in systemic sclerosis: an assessment of interobserver variability in 3 independent studies. J Rheumatol 1993;20:1892–6.

12. 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.
13. 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.
14. Malcarne VL, Greenbergs HL. Psychological adjustment to systemic sclerosis. Arthritis Care Res 1996;9:51–9.
15. Roca RP, Wigley FM, White B. Depressive symptoms associated with scleroderma. Arthritis Rheum 1996;39:1035–40.
16. Clements PJ, Wong WK, Hurwitz EL, Furst DE, Mayes M, White B, 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.
17. Smyth AE, MacGregor AJ, Mukerjee D, Brough GM, Black CM, Denton CP. A cross-sectional comparison of three self-reported functional indices in scleroderma. Rheumatology (Oxford) 2003;42:732–8.
18. Sultan N, Pope JE, Clements PJ, Scleroderma Trials Study Group. The health assessment questionnaire (HAQ) is strongly predictive of good outcome in early diffuse scleroderma: results from an analysis of two randomized controlled trials in early diffuse scleroderma. Rheumatology (Oxford) 2004;43:472–8.
19. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, et al. Responsiveness of the SF-36 and the Health Assessment Questionnaire disability index in a systemic sclerosis clinical trial. J Rheumatol 2005; 32:832–40.
20. Rannou F, Poiraudeau S, Berezne A, Baubet T, Le-Guern V, Cabane J, et al. Assessing disability and quality of life in systemic sclerosis: construct validities of the Cochin Hand Function Scale, Health Assessment Questionnaire (HAQ), Systemic Sclerosis HAQ, and Medical Outcomes Study 36-Item Short Form Health Survey. Arthritis Rheum 2007;57:94–102.
21. Ouimet JM, Pope JE, Gutmanis I, Koval J. Work disability in scleroderma is greater than in rheumatoid arthritis and is predicted by high HAQ scores. Open Rheumatol J 2008;2:44–52.
22. Sekhon S, Pope J, Canadian Scleroderma Research Group (CSRG), Baron M. The minimally important difference (MID) for patient centered outcomes including Health Assessment Questionnaire (HAQ), fatigue, pain, sleep, global VAS and SF-36 in scleroderma (SSc). J Rheumatol 2010;37:591–8.
23. Khanna D, Furst DE, Hays RD, Park GS, Wong WK, Seibold JR, et al. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. Ann Rheum Dis 2006;65:1325–9.
24. Schnitzer M, Hudson M, Baron M, Steele R. Disability in systemic sclerosis: a longitudinal observational study. J Rheumatol 2011;38:685–92.
25. Cole JC, Khanna D, Clements PJ, Seibold JR, Tashkin DP, Paulus HE, et al. Single-factor scoring validation for the Health Assessment Questionnaire-disability index (HAQ-DI) in patients with systemic sclerosis and comparison with early rheumatoid arthritis patients. Qual Life Res 2006;15:1383–94.
26. Johnson SR, Glaman DD, Schentag CT, Lee P. Quality of life and functional status in systemic sclerosis compared to other rheumatic diseases. J Rheumatol 2006;33:1117–22.
27. Georges C, Chassany O, Toledano C, Mouthon L, Tiev K, Meyer O, et al. Impact of pain in health related quality of life of patients with systemic sclerosis. Rheumatology (Oxford) 2006;45:1298–302.
28. Korn JH, Mayes M, Cerinic M, Rainisio M, Pope J, et al, for the RAPIDS-1 Study Group. Digital ulcers in systemic sclerosis: prevention by treatment with bosentan, an oral endothelin receptor antagonist. Arthritis Rheum 2004;50:3985–93.
29. Zelenietz C, Pope J. Differences in disability as measured by the Health Assessment Questionnaire (HAQ) between patients with and without digital ulcers in systemic sclerosis: a post hoc analysis of pooled data from two randomized controlled trials in digital ulcers using bosentan. Ann Rheum Dis 2010;69:2055–6.
30. Brower LM, Poole JL. Reliability and validity of the Duruoz Hand Index in persons with systemic sclerosis (scleroderma). Arthritis Rheum 2004;51:805–9.
31. Lawrence E, Pope J, Al Zahraly Z, Lalani S, Baron M. The relationship between changes in self-reported disability (measured by the Health Assessment Questionnaire-HAQ) in scleroderma and improvement of disease status in clinical practice. Clin Exp Rheumatol 2009;27 Suppl 54:32–7.
32. Khimdas S, Harding S, Bonner A, Canadian Scleroderma Research Group, Zummer B, Baron M, Pope J. Associations with digital ulcers (DU) in a large cohort of systemic sclerosis (SSc). Arthritis Care Res 2010:63:142–9.
33. Stevens A, Pope JE. Retest reliabilities and variability among scleroderma patients for 4 tests of disability: support for a better measure. Presented at the Canadian Rheumatology Association Meeting, 1995. J Rheumatol 1995;22:1603.
34. Gazi H, Pope JE, Clements P, Medsger TA, Martin RW, Merkel PA, et al. Outcome measurements in scleroderma: results from a Delphi exercise. J Rheumatol 2007;34:501–9.
35. Georges C, Chassany O, Mouthon L, Tiev K, Toledano C, Meyer O, et al. Validation of French version of the Scleroderma Health Assessment Questionnaire (SSc HAQ). Clin Rheumatol 2005;24:3–10.
36. Johnson SR, Baron M, Hudson M, Taillefer S, Hirsch A, and the Canadian Scleroderma Research Group. Lung VAS has incremental concurrent validity over the HAQ-DI was an outcome measure evaluating scleroderma lung disease. American College of Rheumatology Annual Meeting, San Diego, 2005. Arthritis Rheum 2005;52 Suppl:S584.
37. Kuwana M, Sato S, Kikuchi K, Kawaguchi Y, Fujisaku A, Misaki Y, et al. Evaluation of functional disability using the health assessment questionnaire in Japanese patients with systemic sclerosis. J Rheumatol 2003;30:1253–8.
38. Morita Y, Muro Y, Sugiura K, Tomita Y, Tamakoshi K. Results of the Health Assessment Questionnaire for Japanese patients with systemic sclerosis: measuring functional impairment in systemic sclerosis versus other connective tissue diseases. Clin Exp Rheumatol 2007;25:367–72.
39. La Montagna G, Cuomo G, Chiarolanza I, Ruocco L, Valentini G. HAQ-DI Italian version in systemic sclerosis. Reumatismo 2006;58: 112–5.
40. Khanna D, Clements PJ, Postlethwaite AE, Furst DE. Does incorporation of aids and devices make a difference in the score of the health assessment questionnaire-disability index? Analysis from a scleroderma clinical trial. J Rheumatol 2008;35:466–8.
41. Serednicka K, Smyth AE, Black CM, Denton CP. Using a self-reported functional score to assess disease progression in systemic sclerosis. Rheumatology (Oxford) 2007;46:1107–10.
42. Guillevin L, Ortonne JP. Treatment of scleroderma. Ann Med Interne (Paris) 1983;134:754–65. In French.
43. Ostojic P, Damjanov N. The Scleroderma Assessment Questionnaire (SAQ): a new self-assessment questionnaire for evaluation of disease status in patients with systemic sclerosis. Z Rheumatol 2006;65:168–75.
44. Ostojic P, Damjanov N. Indices of the Scleroderma Assessment Questionnaire (SAQ) can be used to demonstrate change in patients with systemic sclerosis over time. Joint Bone Spine 2008;75:286–90.
45. Sigl T, Ewert T, Stucki G. Patient-centered assessment of functional health in systemic sclerosis: where are we now? Z Rheumatol 2004;63: 463–9.
46. Silman A, Akesson A, Newman J, Henriksson H, Sandquist G, Nihill M, et al. Assessment of functional ability in patients with scleroderma: a proposed new disability assessment instrument. J Rheumatol 1998;25: 79–83.
47. 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.
48. Danieli E, Airo P, Bettoni L, Cinquini M, Antonioli CM, Cavazzana I, et al. Health-related quality of life measured by the Short Form 36 (SF-36) in systemic sclerosis: correlations with indexes of disease activity and severity, disability, and depressive symptoms. Clin Rheumatol 2005; 24:48–54.
49. Pope J, Baron M, Bellamy N, Campbell J, Carette S, Chalmers I, et al. The variability of skin scores and clinical measurements in scleroderma. J Rheumatol 1995;22:1271–6.
50. Fan X, Pope J, Canadian Scleroderma Research Group (CSRG), Baron M. What Is the relationship between disease activity, severity and damage in a large Canadian systemic cohort? Results from the Canadian Scleroderma Research Group (CSRG). Rheumatol Int 2010;30:1205.
51. Hudson M, Impens A, Baron M, Seibold JR, Thombs BD, Walker JG, et al. Discordance between patient and physician assessments of disease severity in systemic sclerosis. J Rheumatol 2010;37:2307–12.
52. Clements PJ, Hurwitz EL, Wong WK, Seibold JR, Mayes M, White B, 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.
53. Schieir O, Thombs BD, Hudson M, Boivin JF, Steele R, Bernatsky S, et al. Prevalence, severity, and clinical correlates of pain in patients with systemic sclerosis. Arthritis Care Res (Hoboken) 2010;62:409–17.
54. Malcarne VL, Hansdottir I, McKinney A, Upchurch R, Greenbergs HL, Henstorf GH, et al. Medical signs and symptoms associated with disability, pain, and psychosocial adjustment in systemic sclerosis. J Rheumatol 2007;34:359–67.
55. Benrud-Larson LM, Heinberg LJ, Boling C, Reed J, White B, Wigley FM, et al. Body image dissatisfaction among women with scleroderma: extent and relationship to psychosocial function. Health Psychol 2003; 22:130–9.
56. Thompson AE, Shea B, Welch V, Fenlon D, Pope JE. Calcium channel blockers for Raynaud's phenomenon in progressive systemic sclerosis. Arthritis Rheum 2001;44:1841–7.
57. Hudson M, Bernatsky S, Taillefer S, Fortin PR, Wither J. Patients with

systemic autoimmune diseases could not distinguish comorbidities from their index disease. J Clin Epidemiology 2008;61:654–62.
58. Sandusky SB, McGuire L, Smith MT, Wigley FM, Haythornthwaite JA. Fatigue: an overlooked determinant of physical function in scleroderma. Rheumatology (Oxford) 2009;48:165–9.
59. Hewlett S, Dures E, Almeida C. Measures of fatigue. Arthritis Care Res (Hoboken) 2011;63:S263–86.
60. Suarez-Almazor ME, Kallen MA, Roundtree AK, Mayes M. Disease and symptom burden in systemic sclerosis: a patient perspective. J Rheumatol 2007;34:1718–26.
61. Khanna D, Hays RD, Park GS, Braun-Moscovici Y, Mayes MD, McNearney TA, et al. Development of a preliminary scleroderma gastrointestinal tract 1.0 quality of life instrument. Arthritis Rheum 2007;57: 1280–6.
62. Frech T, Hays RD, Maranian P, Clements PJ, Furst DE, Khanna D. Prevalence and correlates of sleep disturbance in systemic sclerosis: results from the UCLA scleroderma quality of life study. Rheumatology (Oxford) 2011;50:1280–7.
63. Bodukam V, Hays RD, Maranian P, Furst DE, Seibold JR, Impens A, et al. Association of gastrointestinal involvement and depressive symptoms in patients with systemic sclerosis. Rheumatology (Oxford) 2011; 50:330–4.
64. Eypasch E, Williams JI, Wood-Dauphinee S, Ure BM, Schmulling C, Neugebauer E, et al. Gastrointestinal Quality of Life Index: development, validation and application of a new instrument. Br J Surg 1995; 82:216–22.
65. Thoua NM, Bunce C, Brough G, Forbes A, Emmanuel AV, Denton CP. Assessment of gastrointestinal symptoms in patients with systemic sclerosis in a UK tertiary referral centre. Rheumatology (Oxford) 2010; 49:1770–5.
66. Muro Y, Sugiura K, Nitta Y, Mitsuma T, Hoshino K, Usuda T, et al. Scoring of reflux symptoms associated with scleroderma and the usefulness of rabeprazole. Clin Exp Rheumatol 2009;27 Suppl 54:15–21.
67. Janiak P, Thumshirn M, Menne D, Fox M, Halim S, Fried M, et al. Clinical trial: the effects of adding ranitidine at night to twice daily omeprazole therapy on nocturnal acid breakthrough and acid reflux in patients with systemic sclerosis: -a randomized controlled, cross-over trial. Aliment Pharmacol Ther 2007;26:1259–65.
68. Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea: contents, interobserver agreement and physiologic correlates of two new clinical indexes. Chest 1984;85;751–8.
69. Mahler D, Wells C. Evaluation of clinical methods for rating dyspnea. Chest 1988;93:580–6.
70. Elliott MW, Adams L, Cockcroft A, Macrae KD, Murphy K, Guz A. The language of breathlessness. Am Rev Respir Dis 1991;144:826–32.
71. Stoller JK, Ferranti R, Feinstein AR. Further specification and evaluation of a new clinical index for dyspnea. Am Rev Respir Dis 1986;134: 1129–34.
72. Chhabra SK, Gupta AK, Khuma MZ. Evaluation of three scales of dyspnea in chronic obstructive pulmonary disease. Ann Thorac Med 2009;4:128–32.
73. Yoza Y, Ariyoshi K, Honda S, Taniguchi H, Senjyu H. Development of an activity of daily living scale for patients with COPD: the Activity of Daily Living Dyspnea scale. Nihon Kokyuki Gakkai Zasshi 2009;47: 858–64.
74. Baron M, Sutton E, Hudson M, Thombs B, Markland J, Pope J, et al. The relationship of dyspnoea to function and quality of life in systemic sclerosis. Ann Rheum Dis 2008;67:644–50.
75. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 2006;354:2655–66.
76. Borg G. Psychophysical basis of perceived exertion. Med Sci Sports Exerc 1982;14:377–81.
77. Khanna D, Clements PJ, Furst DE, Chon Y, Elashoff R, Roth MD, et al. Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with systemic sclerosis and active alveolitis: results from the Scleroderma Lung Study. Arthritis Rheum 2005;52:592–600.
78. Khanna D, Yan X, Tashkin DP, Furst DE, Elashoff R, Roth MD, et al. Impact of oral cyclophosphamide on health-related quality of life in patients with active scleroderma lung disease: results from the scleroderma lung study. Arthritis Rheum 2007;56:1676–84.
79. Khanna D, Tseng CH, Furst DE, Clements PJ, Elashoff R, Roth M, et al, for Scleroderma Lung Study Investigators. Minimally important differences in the Mahler's Transition Dyspnoea Index in a large randomized controlled trial: results from the Scleroderma Lung Study. Rheumatology (Oxford) 2009;48:1537–40.
80. Goldin J, Elashoff R, Kim HJ, Yan X, Lynch D, Strollo D, et al. Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. Chest 2009;136:1333–40.
81. Incalzi RA, Bellia V, Catalano F, Scichilone N, Imperiale C, Maggi S, et al. Evaluation of health outcomes in elderly patients with asthma and COPD using disease-specific and generic instruments: the Salute Respiratoria nell'Anziano (Sa.R.A.) Study. Chest 2001;120:734–42.
82. Beretta L, Santaniello A, Lemos A, Masciocchi M, Scorza R. Validity of the Saint George's Respiratory Questionnaire in the evaluation of the health-related quality of life in patients with interstitial lung disease secondary to systemic sclerosis. Rheumatology (Oxford). 2007;46:296–301.
83. Antonioli CM, Bua G, Frige A, Prandini K, Radici S, Scarsi M, et al. An individualized rehabilitation program in patients with systemic sclerosis may improve quality of life and hand mobility. Clin Rheumatol 2009;28:159–65.
84. Distler O, Behrens F, Pittrow D, Huscher D, Denton CP, Foeldvari I, et al. Defining appropriate outcome measures in pulmonary arterial hypertension related to systemic sclerosis: a Delphi consensus study with cluster analysis. Arthritis Rheum 2008;59:867–75.
85. Hatano S, Strasser T. Primary pulmonary hypertension. World Health Organization Geneva: WHO; 1975.
86. AHA medical/scientific statement. 1994 revisions to classification of functional capacity and objective assessment of patients with diseases of the heart. Circulation 1994;90:644–5.
87. Chow S, Pope JE, Mehta S. Lack of correlation of the health assessment questionnaire disability index with lung parameters in systemic sclerosis associated pulmonary arterial hypertension. Clin Exp Rheumatol 2008;26:1012–7.
88. McKenna SP, Ratcliffe J, Meads DM, Brazier JE. Development and validation of a preference based measure derived from the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) for use in cost utility analyses. Health Qual Life Outcomes 2008;6:65.
89. Meads DM, McKenna SP, Doughty N, Das C, Gin-Sing W, Langley J, et al. The responsiveness and validity of the CAMPHOR Utility Index. Eur Respir J 2008;32:1513–9.
90. Gomberg-Maitland M, Thenappan T, Rizvi K, Chandra S, Meads DM, McKenna SP. United States validation of the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR). J Heart Lung Transplant 2008;27:124–30.
91. Coffin D, Duval K, Martel S, Granton J, Lefebvre MC, Meads DM, et al. Adaptation of the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) into French-Canadian and English-Canadian. Can Respir J 2008;15:77–83.
92. Matucci-Cerinic M, Seibold JR. Digital ulcers and outcomes assessment in scleroderma [review]. Rheumatology (Oxford) 2008;47 Suppl 5:v46–7.
93. Toffolo SR, Furtado RN, Klein A, Watanabe S, Andrade LE, Natour J. Measurement of upper limb ulcers in patients with systemic sclerosis: reproducibility and correlation with pain, function, and quality of life. Nurs Res 2008;57:84–92.
94. Shenoy PD, Kumar S, Jha LK, Choudhary SK, Singh U, Misra R, Agarwal V. Efficacy of tadalafil in secondary Raynaud's phenomenon resistant to vasodilator therapy: a double-blind randomized cross-over trial. Rheumatology (Oxford) 2010;49:2420–8.
95. Khanna PP, Maranian P, Gregory J, Khanna D. The minimally important difference and patient acceptable symptom state for the Raynaud's condition score in patients with Raynaud's phenomenon in a large randomised controlled clinical trial. Ann Rheum Dis 2010;69:588–91.
96. Poole JL. Measures of hand function. Arthritis Care Res (Hoboken) 2011;63:S189–99.
97. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am 1998;23:575–87.
98. Mouthon L, Rannou F, Berezne A, Pagnoux C, Arene JP, Fois E, et al. Development and validation of a scale for mouth handicap in systemic sclerosis: the Mouth Handicap in Systemic Sclerosis scale. Ann Rheum Dis 2007;66:1651–5.
99. Pope JE, Bellamy N. Sample size calculations in scleroderma: a rational approach to choosing outcome measurements in scleroderma trials. Clin Invest Med 1995;18:1–10.

Summary Table for Systemic Sclerosis Measures*

Scale	Purpose/content	Method of administration	Respondent burden	Administrative burden	Score interpretation	Reliability evidence	Validity evidence	Ability to detect change	Strengths	Cautions
HAQ HAQ DI (3)	Measures function in 8 domains	Self-report	5 minutes to complete Questions not always relevant to patient, e.g., opening milk carton	Hand scoring; <1 minute	8 subscales scored each in worst answer from 0–3 Adds 1 point if aids used unless score already = 3 If 8 scales answered, then total score divided by 8 (divide total score by no. scales completed) 0 = no disability, 3 = worst score. Moderate impairment is often >1.0 Observation that adding aids in SSc; scleroderma may overestimate disability (40)	Reliability demonstrated in RP RCT (6) Good test–retest reliability when repeated after 2 days (33) ICCs range 0.38–0.76 for therapist watching activities and patient self-reporting (13)	Convergent/construct validity shown (17) Construct validity shown (10) Predictive validity shown in early dcSSc trials/clinical practice (18,31) Content validity not tested in SSc (9) Concurrent validity shown (13,30) Predictive low HAQ was associated with improved MD global 1 year later in clinical practice (31) A low HAQ predicted improved skin/patient global assessments in early dcSSc RCTs (1.5–5 times more likely to improve) (18) HAQ explained change in skin score in dcSSc RCT (R^2 = 0.57) (53) Face validity as mean HAQ worse in dcSSc than lcSSc subset (32) HAQ worsens over time in SSc in a year from 0.039 to 0.071, and in 3 years by ≥0.12 (24) HAQ and pain both related to SF-36 (PCS) (R = 0.70) (27) Structural validity between RA and SSc patients (25)	Responsiveness shown (4,6) MID in early dcSSc 0.10–0.14 (ES 0.15–0.21) (23) Clinical practice not very well proven For improvement was 0.0125; well below a change in measurement, but at 75th percentile it was 0.125, which is a minimum change in scoring of HAQ For worsening was 0.042 (below the minimal measurement change) and 0.217 for 75th percentile so MID for HAQ in clinical practice may be lacking (22) MID for HAQ 0.2–0.25 by MD Delphi consensus (34) HAQ related to work disability in SSc (21) If skin score improved then HAQ improved in active dcSSc RCT (ICC < 0.69) (19)	Measures function easily Inexpensive Compares to other diseases such as RA Used widely in RCTs and clinical practice Predicts mortality/morbidity and if very low HAQ at baseline in early dcSSc, predicts higher chance of improving skin score over next year (18)	Not predictive of all organ systems Not related to PAH changes (88) but dyspnea explained some variance of HAQ (52) Primarily measuring MSK-related function Mostly tested in dcSSc early trials Most patients in clinical practice worsen function so MIDs in clinical practice have wide range and may lack some face validity when change is less than the minimum change in measurement
SHAQ (4)	Measures function in 8 domains and separately VAS scales for GI, lung, vascular, digital ulcers and overall	Self-report	5 minutes to complete	1 minute Hand scoring	HAQ DI scored from 0–3 and 5 other VAS scales converted to 0–3 (GI, vascular, digital ulcers, lung, overall)	Yes (6)	Convergent construct (17) Construct not tested Predictive not tested Content validity, not tested Concurrent (6) VAS subscales significantly correlated with objective parameters (4)	Responsive (4,6) MIDs for VAS scales Not tested fully, but VAS for RP on SHAQ has ES 0.20 and SRM 0.15 (6)	HAQ DI for function and adds other areas of problems in RP (easy to measure and important to patients) Scope in SSc is more inclusive due to addition of subscales Fast, inexpensive, feasible	Not utilized as much as HAQ DI Some validity of the subscales is lacking Sometimes papers report the subscales scores added together and this is not the intended scoring of the SHAQ
Physician global assessment	Outcome assessor to estimate overall disease activity Also used for severity and damage	Physician-administered after reviewing patient and often labs/results of other organ function tests	Takes a few seconds to complete (but after a review of the patient's specific data)	Easy scoring (10-mm VAS or Likert change score)	Higher number is a worse score	Test–retest reliability; interobserver reliability tested (very good ICCs) (50) Agreement between MD and patient global assessments is ICC 0.38 (52)	Content reliability (50) SF-36 related to MD global (19)	MID by MD Delphi is 8–13 on 100-mm VAS for MD global (34) Responsive with other parameters: skin scores/breathlessness Low HAQ associated with improved MD global 1 year later in clinical practice (31) Improved MD global associated with skin score and HAQ (R^2 = 0.46) (53)	Gives overall number for SSc disease Sensitive to change and good face validity	No standardized question No standardization of measurement before global assessment completed Important to include both MD and patient global assessments since not strongly correlated and measure different perspectives
Patient global assessment	Overall assessment of SSc by patient	Self-report	<1 minute	Hand scored in <1 minute	Higher number is worse Score of 0 considered no activity (sometimes scale is overall where 100 = best and 0 = worst ever, but not usually)	Cannot test for interobserver variability as scale is completed by patient Test–retest reliability modest (33) Agreement between MD and patient global assessments ICC 0.38 (52)	Patient global improvement and skin score improvement related in 2 early dcSSc trials (r = ≥0.25) (18) SF-36 related to patient global (19)	MID tested in clinical practice for patient global is from 4 to 7 mm on 100-mm VAS (22) Responsive with other parameters: skin scores, breathlessness (52) MID by physician Delphi is 10–12 on 100-mm VAS, for patient global (34)	Patient reported overall SSc disease burden	Does not correlate strongly with MD global assessment Patient may be measuring overall disease burden (activity and damage or severity) and not disease activity alone Patients rate higher disease severity than physicians (52) Important to include both MD and patient globals as they are not very strongly correlated and thus measuring different perspectives
Patient VAS of SSc	Measures SSc pain	Self-report	≤1 minute	~2 seconds to hand score	Higher rating implies more pain Often 0–100-mm VAS or Likert scale	Test–retest very good (33)	Pain related to more RP, active ulcers, arthritis, GI symptoms. Worse pain with dsSSc (22) SSc pain higher than RA (in one study and not in another) (26,97,22) HAQ and pain both related to SF-36 (PCS) (R = 0.70) (27) Pain and fatigue related to physical function (59)	MID for pain in SSc clinical practice is 4–8 on 100-mm VAS (22) MID by expert opinion of pain is 0.2–0.3 out of 3 for pain VAS (34)	Easy to do/important to test pain in SSc as it is common and likely under-recognized	Pain can be attributed to many causes (skin with pruritis, ulcers, RP, GI, arthritis, etc.), so if wanting to intervene with pain, specific organ pain or symptoms may need to be asked Pain questionnaire is not standardized with respect to how it is asked
SBI (7)	Determine SSc QOL in 8 domains (skin, hand mobility, calcinosis, shortness of breath, eating, bowel, sleep, and pain) 5 questions per domain	Self-report	A couple of minutes	A few minutes	Questions answered in Likert scale Scoring for each domain 0–10. Number of domains with problems noted	Good reliability between items Distribution of scores ~equal between low/medium/high Patients had mean of 5.7 out of 8 problems	Significantly correlated with HAQ (0.3–0.6) Negatively correlated with SF-36 (PCS and MCS) Construct validity Focus groups tested to develop domains (61) Domains independent but significantly related except shortness of breath (between domain correlations low to moderate, desired for different domains)	Not tested	Easy to complete Fills gap for patient symptoms not fully included in other scales such as SHAQ, UCLA SCTC GIT Has some common domains that are not identical	Only tested in single site with 62 patients, more work on instrument needed Easy to complete Value-added above other QOL (generic) instruments/SSc tools Adding all 5 questions in each domain may not give comparable score between domains as questions are how much, how often, how much interference and thus a rare problem that interferes a lot May score the same as a common problem interferes a little

(continued)

Summary Table for Systemic Sclerosis Measures* *(Cont'd)*

Scale	Purpose/content	Method of administration	Respondent burden	Administrative burden	Score interpretation	Reliability evidence	Validity evidence	Ability to detect change	Strengths	Cautions
UCLA SCTC GIT 2.0 (5,62)	7 multi-item scales with areas of reflux, distention/ bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning For total score, 6 of 7 scales averaged (excluding constipation) 34-item revised instrument 2.0 revised from 52-item 1.0	Self-report	A couple of minutes	A minute	Items scored 0–3 with lower values showing better HRQOL Higher score indicates more GI problems Correlated with self-rated GI severity (r = 0.60) Discriminant validity Detects differences among no symptoms, very mild, mild, moderate, severe/ very severe (F = 31)	Overall, test–retest reliability ICC 0.81; coefficient α = 0.71 Internal consistency shown (Cronbach's α 0.67–0.91) (5) Test–retest reliability 0.68–0.89 (5) Severity scales/subscales (r = 0.2–0.5) and overall r = 0.6 Higher scores in subscales with MD diagnosis of each of GERD, gastroparesis, bacterial overgrowth, pseudo-obstruction, diarrhea, rectal incontinence, constipation	Face validity Symptoms correlated with GI diagnoses (5,62) Correlated with worse sleep and depressive symptoms (63,64) Social functioning/emotional well-being scales correlated with those of SF-36 (r = 0.36) Convergent and divergent validity shown (5)		Only SSc-specific GI scale validated for SSc No floor effect Small ceiling effect	May not be responsive in a GI SSc trial (ceiling effect) If symptoms are severe and improve by 30% they still may be considered frequent by the patient
RCS (6)	Daily self-assessment of RP activity using 0–10 ordinal scale that incorporates cumulative daily frequency, duration, severity, and impact of RP attacks Therefore in general a part of RP attack diary	Self-report	Moderate respondent burden as done in conjunction with recording daily attacks (frequency, severity, and duration) Diary must be completed daily	Scoring adds each day that diary is completed then divided by number of days recorded May take long time if RP trial is long, but is total number of attacks, duration, and severity that takes scoring time	Scored 0–10 ordinal scale Higher score is worse	Higher if digital ulcers vs no ulcers; patient globals did not differentiate (but MD globals did) (6)	Construct, content, criterion, and discriminant validity present (tested in RP trial) (6) Change associated with improved frequency and RP duration, improved digital ulcers (new and healing), SHAQ, and MD and patient global assessments. (95)	Sensitive to change (tested in RP trial) (6) MID for RCS improvement is 14 points on 100-mm VAS (96) ES 0.6 SRM 0.64 These are same or better than other RP VAS scales (6)	Used in clinical trials and supposed to take into account the entire RP impact for patient	Most clinicians and even researchers are not too certain about what the score means and what is clinically important regarding baseline score and change Still performed in addition to frequency, severity, and duration of attacks daily Confusion between 0–10 ordinal scale and 100-mm VAS may occur as RCS is an ordinal scale and overall severity or activity of RP is often measured as continuous VAS
CAMPHOR (7)	QOL scale for PAH Designed/validated in idiopathic PAH Overall symptoms (energy, breathlessness, mood subscales) Functioning and QOL	Self-report	10 minutes	At least a few minutes to score	Test–retest correlations 0.86–0.92 (7)	Good internal consistency (α = ≥0.9) (7)	Convergent, divergent validity Responsive to change Correlated with change in functional class (7,89,90)	Subscales and utility were as responsive as 6MWT (ES 0.31–0.69 for CAMPHOR and 0.16–0.34 for 6MWT) (90) MID for utility index estimated as ~0.09 (90)	PAH specific QOL scale Validated in idiopathic PAH, but not in SSc	Not tested in PAH of SSc but in idiopathic PAH and alone may be insufficient to reflect all QOL aspects of SSc as it is weighted especially towards the impact of dyspnea (not tight skin, pain, GI problems, arthritis, etc.).
BDI TDI (69)	Measure severity and change in dyspnea Each index rates 3 categories: magnitude of task, magnitude of effort, and functional impairment Each category has 5 grades from 0 (severe) to 4 (unimpaired) Categories added together for BDI (range 0–12) TDI has 7 grades from −3 (major deterioration) to 3 (major improvement) Ratings for TDI scales added for transition score (−9 to 0)	Self-report or interviewer-administered	10 minutes	A few minutes	12 = no impairment and 0 = the worst for BDI 9 is the best, 0 is no change and − is the worst for TDI TDI has improvement as major, moderate, or minor corresponding to improvement on scale as 7–9, 4–6, and 1–3 and conversely there is deterioration if worsening	Not fully tested	Face validity (correlated with VAS for breathing (r = −0.6) and pulmonary function (FVC and DLco) (78) Other validity not fully tested in SSc but related to other lung parameters in SSc lung study (79)	MID for TDI was 1.05–2.16 (mean 1.5) for improvement and from −0.61 to −2.55 (mean −1.5) for worsening. (80) RR >3 for achieving at least MID for transitional dyspnea index score in cyclophosphamide vs placebo in ILD (79)	Face validity and sensitive to change in scleroderma lung study (thus at least partially validated) (76)	Not proven for PAH in SSc No consensus on which dyspnea scale to use for ILD studies in SSc

* HAQ DI = Health Assessment Questionnaire disability index; SSc; scleroderma = systemic sclerosis; RP = Raynaud's phenomenon; RCT = randomized controlled trial; ICC = intraclass correlation coefficient; dcSSc = diffuse cutaneous SSc; MD = physician; lcSSc = limited cutaneous SSc; SF-36 = Medical Outcomes Study Short Form 36; PCS = physical component score; RA = rheumatoid arthritis; MID = minimal important difference; ES = effect size; PAH = pulmonary arterial hypertension; MSK = musculoskeletal; SHAQ = scleroderma HAQ; VAS = visual analog scale; GI = gastrointestinal; SRM = standardized response mean; SBI = Symptom Burden Index; QOL = quality of life; MCS = mental component score; UCLA SCTC GIT = University of California, Los Angeles, Scleroderma Clinical Trials Consortium Gastrointestinal Scale; HRQOL = health-related QOL; GERD = gastroesophageal reflux disease; RCS = Raynaud's Condition Score; CAMPHOR = Cambridge Pulmonary Hypertension Outcome Review; 6MWT = 6-Minute Walk Test; BDI = Baseline Dyspnea Index; TDI = Transition Dyspnea Index; FVC = forced vital capacity; DLco = diffusing capacity for carbon monoxide; RR = relative risk; ILD = interstitial lung disease.