Ross Laura (Orcid ID: 0000-0003-4636-729X)
Khanna Dinesh (Orcid ID: 0000-0003-1412-4453)

Running head: Patient and physician global assessments in systemic sclerosis

Patient and physician global assessments of disease status in systemic sclerosis

Authors: Laura Ross[1,2] (MBBS), Mandana Nikpour[1,2] (MBBS, PhD), Julie D'Aoust[3] (MDCM), Dinesh Khanna[4] (MD, MSc), Peter A. Merkel[5] (MD, MPH), John D. Pauling[6] (BMedSci, BMBS, PhD), Murray Baron[7] (BSc, MDCM)

Author degrees:

LR: MBBS, FRACP

MN: MBBS, FRACP, FRCPA, PhD

JD: MDCM, FRCP(C)

DK: MD, MSc

PM: MD, MPH

JP: BMedSci, BMBS, FRCP, PhD

MB: BSc, MDCM, FRCP(C)

Affiliations:

1. Department of Medicine, The University of Melbourne at St Vincent's Hospital, Melbourne Australia
2. Department of Rheumatology, St Vincent's Hospital, Melbourne Australia
3. Division of Rheumatology, The Ottawa Hospital, University of Ottawa, Ottawa, Canada
4. University of Michigan Scleroderma Program, Division of Rheumatology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acr.25056

5. Division of Rheumatology, Department of Medicine, Division of Epidemiology, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA
6. Department of Rheumatology, North Bristol NHS Trust, Bristol, UK
7. Division of Rheumatology, Jewish General Hospital, McGill University, Montreal Canada

Corresponding author:

Dr Laura Ross, Department of Rheumatology, St Vincent's Hospital, Melbourne 41 Victoria Parade, Fitzroy VIC 3065, Australia; Ph: +61 3 9231 2211; Fax: +61 3 9231 3841; Email: laura.ross@svha.org.au

Funding:

No financial support was provided for this study.

LR is supported by an Arthritis Australia – Australian Rheumatology Association (Victoria) Fellowship. MN holds a National Health and Medical Research Council of Australia Investigator Grant (GNT 1176538).

Conflict of interest:

LR: No conflict of interest

MN: Research support from Actelion, Astra Zeneca, BMS, BSK, Janssen and UCB and has received honoraria from Actelion, Boehringer-Ingelheim, Eli Lilly, GSK, Janssen, Pfizer and UCB outside of the submitted work.

JD: No conflict of interest

DK: Consulting fees from Actelion, Acceleron, Amgen, Bayer, Boehringer-Ingelheim, Chemomab, CSL Behring, Genentech/Roche, Horizon, Paracrine Cell Therapy, Mitsubishi Tanabe, Prometheus, and owns stock options in Eicos Sciences, Inc.

PM: Consulting: AbbVie, AstraZeneca, Boehringer-Ingelheim, Bristol-Myers Squibb, ChemoCentryx, CSL Behring, Dynacure, EMDSerono, Forbius, Genentech/Roche, Genzyme/Sanofi, GlaxoSmithKline, Immagene, InflaRx, Janssen, Kiniksa, Kyverna, Magenta, MiroBio, Mitsubishi, Neutrolis, Novartis, NS Pharma, Pfizer, Regeneron, Sparrow, Takeda, Talaris.

Research Support: AbbVie, AstraZeneca, Boehringer-Ingelheim, Bristol-Myers Squibb, ChemoCentryx, Eicos, Electra, Forbius, Genentech/Roche, Genzyme/Sanofi, GlaxoSmithKline, InflaRx, Sanofi, Takeda.

Stock options: Kyverna.

Royalties: UpToDate.

JDP: Personal support and / or undertaken consultancy work for Janssen, Astra Zeneca, Permeatus Inc, Boehringer-Ingelheim and Sojournix Pharma outside of the submitted work.

MB: No conflict of interest

Word count: 4,199 words

Abstract

Global assessments of disease by both patients and physicians are widely used in clinical studies of systemic sclerosis (SSc). They are commonly secondary endpoints in randomised controlled trials (RCTs) and are considered important items in composite measures of treatment response. A comprehensive literature review was conducted of the formats, wording, and clinimetric properties of the patient global assessments (PtGA) and physician global assessments (PhGA) used in RCTs of SSc. Marked heterogeneity was found in the wording and measurement scales of the global assessments applied in RCTs. These instruments were not developed using rigorous methodology and have not been fully validated. There is a pressing need for standardisation and validation of patient and physician global assessment tools in SSc to enable universal application of these measures to across RCTs in SSc.

Key words: systemic sclerosis, outcome measurement, global assessment, clinimetrics

Significance and innovation

- There is marked heterogeneity in the application of global assessments of systemic sclerosis across randomised controlled trials.
- No global assessment has been developed using rigorous methodology to prove the content and construct validity of the instrument.
- There is an urgent need to standardise and validate patient and physician global assessments in systemic sclerosis to improve the quality of clinical trials.

Background

Patient assessment of disease status was first incorporated into rheumatology practice in the 1970s when a patient self-assessment of pain was used to evaluate rheumatoid arthritis (RA) (1). Over the past 50 years, the use of patient-reported assessments of overall health and disease activity and severity have become commonplace in the evaluation of disease states in most rheumatic diseases. It is unclear when or with what justification the physician global assessment (PhGA) was first used to assess patients with systemic sclerosis (SSc), but it has been included in most clinical trials since the 1990s (2-5). In contrast to multi-question patient-reported outcomes measuring organ manifestations or health-related quality of life, the PtGA and PhGA are presented as a single question and ask respondents to rate global disease status, either via a visual analogue scale (VAS), numeric rating scale, or Likert scale. For at least the past 25 years, it has been recommended that both a patient global assessment (PtGA) and PhGA be included in clinical trials of SSc (6, 7). Newer multi-system outcome measures for use in randomised controlled trials (RCTs) in SSc such as the American College of Rheumatology (ACR) Composite Response Index for Systemic Sclerosis (CRISS) include both a PtGA and PhGA as part of a composite measure of response to treatment (8).

As the PtGA and PhGA each capture disease status from different perspectives, they have complementary roles in measuring disease response in RCTs and observational studies (9, 10). Patient self-assessment of the burden of disease and change in disease status are essential aspects of the assessment of the efficacy and tolerability of novel therapies. Patient assessment may be more likely to reflect a combination of physical function, psychological well-being, pain, fatigue, and severity of organ involvement. A

PtGA can be a direct measure of change in how patients feel and function, important aspects considered in the process of regulatory approval of new therapies. The physician assessment is more likely to consider poor prognostic factors, organ damage and physicians give greater weighting to physical and laboratory findings such as the modified Rodnan Skin Score (mRSS) and renal function when rating disease severity (9, 10). However, there is no strict division between patient and physician global assessments, and these measures almost certainly assess several overlapping domains, and which factors contribute most highly to a response to treatment depends not only on the evaluator, but also on the nature of the trial and physiologic effects of the treatment under study.

Given the widespread use of both PtGA and PhGA in the assessment of SSc, this study sought to assess the PtGA and PhGAs applied in SSc research. A comprehensive literature review was performed to evaluate the clinimetric properties of existing global outcome measures, with reference to the OMERACT filter of truth, discrimination, and feasibility (Figure 1) (11). To ascertain whether any limitations of the validity of global assessments in SSc are unique to the disease or more widely observed, the clinimetric properties of the PtGA in other rheumatic diseases were also reviewed.

Methods

A comprehensive literature search was performed with a particular emphasis on identifying papers concerning either RCTs or observational studies that included data assessing the face, content or construct validity and discriminant validity of the PtGA and PhGA. The nature of this review was not amenable to a systematic or standardised scoping literature search because of the need to identify a broad range of sources,

including cross-sectional, observational studies and RCTs, as well as identify analyses within papers that included the PtGA and PhGA. A standardised search using PubMed (performed 7 March 2022) with the following search criteria: ('systemic sclerosis' or scleroderma or CREST) and 'global assessment' and ('reliability' or 'validity' or 'minimal clinical important difference'). In addition, a hierarchical literature search, that included hand and expert opinion searches, was performed to identify key publications, a methodology previously used to explore patient-reported outcomes in RA (12). The wording and response items of global assessments used in 16 RCTs were extracted from published manuscripts, clinical trial data registered at *clinicaltrials.gov* (accessed 7 March 2022), or study protocols, where available, to identify the variability in PtGA and PhGA as they are currently used in RCTs in SSc. Two authors (LR, MB) independently screened all titles and abstracts for relevance, with disagreements resolved by consensus. A hierarchical literature search strategy was applied to identify relevant publications assessing the clinimetric properties of the PtGA and PhGA in RA, psoriatic arthritis (PsA), ANCA-associated vasculitis (AAV) and systemic lupus erythematosus (SLE).

Results

The search strategy for SSc-related publications identified 75 citations, of which 27 articles, 16 RCTs and 11 observational studies, were included in the final review of the clinimetric properties of the PtGA and PhGA, with results summarised in Table 1. The wording of the global assessment and response anchors for the PtGA and PhGA were extracted from the 16 RCT publications or study protocols and are summarised in Table 2.

Face and content validity

Expert consensus has concluded that both the PtGA and PhGA should be included as part of the core set of outcome measures for use in RCTs of SSc (7). No formal evaluation of the patient opinion of global assessments has been performed so no assessment could be made regarding the appropriateness and relevance of the instruments to assess its measurement of overall or global status in SSc. Indeed, any assessment of content validity is challenging as a conceptual framework for either instrument has never been established and defined. Poorly measured variables and inconsistently defined measures make it extremely difficult, if not impossible, to evaluate the properties of clinical outcomes (13).

Many studies included PtGA and PhGA either alone as secondary outcomes or as part of a composite measure to improve the content validity of the overall instrument (8). However, the methods for the derivation of the questions and anchors used in the global assessments were not documented in any of the studies found and thus it was not possible to ascertain the content validity of these specific measures. There was also no evidence that patient representatives were involved in the development of any PtGA or that physicians beyond the core investigator group for a specific RCT were involved in developing any global assessment (14). One of the few studies that did employ specific methodology to develop global assessments was that of Steen and Medsger, in the development of the modified health assessment questionnaire (HAQ) for SSc (15). After generating the patient global assessment they asked 11 patients how they would describe the features and impact of SSc on their daily lives and found that the question 'In the past week how much have your overall scleroderma problems interfered with your activities?' incorporated the language patients used to assess their SSc (15).

Significant heterogeneity exists in the types of global assessments, the wording of the questions, and the Likert items or VAS anchors used in RCTs (Table 2). Unlike the work of Steen and Medsger (15), no such exercise was performed to ascertain the most appropriate wording of any global assessment of disease activity, damage, overall health, or other global domains. It is not known which global assessment may best reflect patients' assessment of their own health or disease status.

Construct validity

Recently used global assessments measure a variety of constructs, including disease activity, severity, overall health, and overall SSc health across RCTs (see Table 2). Only low inter-rater agreement between expert-rated disease activity has been observed and there is significant within physician correlation between assessments of disease activity, severity, and damage (16, 17). This suggests that physicians incorporate an assessment of multiple disease constructs when asked to evaluate global disease status, and inclusion of multiple constructs into one measure. Patients and physicians likely have difficulty distinguishing disease activity from damage, and it has been suggested that patient assessments more strongly correlate with particular organ manifestations rather than overall disease (9, 10, 15, 17, 18).

There is poor to moderate correlation between a PhGA of activity and SSc-specific disease activity indices (17, 19), however the PhGA of global health has been shown to correlate with mRSS, a commonly used surrogate measure of overall disease burden (10). Both the PtGA and PhGA of disease severity have been strongly associated with physician-rated improvement over a 12-month period (8). Observational data have shown associations between PtGA and health-related quality of life and overall disease

severity (15, 20), however there are inconsistent reports about the association of PtGA and function as measured by the HAQ (20, 21).

The lack of independent gold standard measures of either activity, damage, or disease severity in SSc limits any evaluation of the criterion validity of both PtGA and PhGA and as such it remains unproven. Somewhat paradoxically, despite the unproven content and construct validity of PtGA and PhGA, it is often these global assessments that are used to anchor and validate other new outcome measures, such as the EUSTAR Activity Index and ACR CRISS (8, 22).

Discriminant validity

Review of the performance of the PhGA in RCTs in SSc has shown that the PhGA can identify a net treatment benefit in active treatment arms in some studies despite the failure of many RCTs to reach their primary endpoint (23-26). Repeated PhGAs were shown to be a useful measure of disease status in the D-penicillamine RCT, correlating with mRSS and HAQ scores (27). Notably, the PtGA did not distinguish between treatment and placebo groups in any recent randomised trial with published results, however these were clinical trials that did not reach their primary endpoint. The phase 2 and 3 trials of tocilizumab in SSc (28, 29) did not find a meaningful change in PtGA or PhGA or between group differences between the treatment and placebo study arms, despite meaningful improvement in other secondary outcomes.

Sensitivity to change & minimal clinical important difference

The sensitivity to change of global assessments in SSc remains unknown and clinically meaningful thresholds of change (MCID) are yet to be defined. Observational cohort

data used to identify items that predict overall improvement over a 12-month period defined a standard error measure of 0.26 and 1.75 for PtGA and PhGA, respectively (8). Attempts have been made to define the minimal clinical important difference of global VAS, using observational data (21) and a Delphi exercise of experts (30), but no data from prospective randomised trials have been used to define these thresholds.

Reliability

Reliability has yet to be established for the patient and physician global assessments in SSc. Data collected from the post-procedure observation phase of a RCT for Raynaud's phenomenon found that both the PtGA and PhGA of disease severity were unreliable measures with a standard deviation of difference between 18-25% (31). Increased variability of PtGA and PhGA scores compared to other measures of disease status was also observed in observational cohort data and significant inter-rater variability of PtGA has been described (16, 32).

Feasibility

Both the PtGA and PhGA are quick and simple to perform without any great burden to either the patient or physician in studies of SSc. Both assessments are well accepted by both patients and physicians (12, 31, 33). Feasibility was formally assessed during the development of the ACR CRISS, and global assessments were found to be feasible measures of disease both in the derivation study, where they achieved a pre-specified threshold of feasibility, and by subjective evaluation by the ACR CRISS steering committee (8).

Measurement of global assessments in other rheumatic diseases

Assessment of both the PtGA and PhGA in other rheumatic diseases shows similar limitations to those revealed by our evaluation of global assessments in SSc. Global assessments are used widely across many multi-system autoimmune diseases, with the derivation of each global assessment largely undocumented and highly variable application of both the PtGA and PhGA across studies. The PtGA has been incorporated into many of the major disease activity indices to measure disease status in RA, such as Disease Activity Score 28 (DAS28) (34), Simplified Disease Activity Index (SDAI) (35), the Clinical Disease Activity Index (CDAI) (36), and the Routine Assessment of Patient Index Data (RAPID3) (37). The PtGA is reported in almost half of all RA clinical trials (1). However, the scales used to rate the PtGA are often not specified, and the RA PtGA has poor content validity due to the lack of patient involvement via focus groups, qualitative studies, or evaluation of the meaning of a given PtGA (1). There is heterogeneity of concepts measured by different PtGAs and variation in the specificity of PtGAs to RA, as some are measures of global health status meaning that patient responses may not be comparable between studies (1). A PtGA is one of the domains of the PsA core domain set and defined as a patient global assessment of disease-related health status (38). Similar to SSc, little is understood about what a PtGA should look like and there remains no standardised application of PtGA across PsA clinical trials (39-41). However, meaningful improvements in the PtGA are frequently observed in PsA RCTs of effective therapies (39-41). The PsA PtGA is considered a reliable measure of changes in disease status; however, it is influenced by the presence of co-morbidities and overall pain (42, 43). There are no studies that detail the development of the PtGA, so it is therefore not possible to assess the face and content validity of any of the PsA PtGAs. Recent systematic review of the global assessments used in clinical trials of psoriasis found significant heterogeneity of

the implementation of these instruments and concluded that all psoriasis PtGA and PhGAs as they are currently used are lacking in content validity and feasibility (44).

AAV is another complex multi-system disease with multiple outcome measures used across clinical trials. The clinimetric properties of the PhGA in AAV remain unassessed (45), but interestingly, as we observed in SSc, it has been used to validate other composite outcome measures such as the Birmingham Vasculitis Activity Score (BVAS) (45) and the Patient-Reported Outcome Measurement Information System (PROMIS) instruments (46). The PtGA is considered to lack validity and reliability in AAV and has untested responsiveness to change (45). The PtGA in AAV has only poor to moderate correlation with BVAS scores and other measures of health-related quality of life (45). No standardised measures of patient or physician global assessment exist for large vessel vasculitis (47).

It is notable that in SLE a PtGA is not used to assess disease status in clinical trials. In contrast, a PhGA is widely used as part of composite outcome measures such as the Lupus Activity Index (48), the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) (49) and the British Isles Lupus Assessment Group (BILAG) index (50). Similar to SSc, there is no consistent application of the PhGA across SLE trials, including discrepancies in the application of the types of visual analogue scales applied, time frames of assessment and whether the PhGA should incorporate investigation findings (51). There are currently significant international efforts underway to standardize the application of the SLE PhGA in recognition that the reliability of the instrument can only be truly established when it is consistently applied across studies.(52)

The patient experience of using a PtGA has been explored in individuals diagnosed with RA. Three overarching problems have been identified: difficulty understanding the meaning and purpose of the PtGA and measurement difficulties (53). Wording of global assessments and the phrasing of response anchors requires consistency as changes to the wording of either outcome anchors or item responses alters the construct measured by a particular global assessment tool (1, 54). In RA, up to half of patients can be confused by the PtGA (55). A potential floor effect exists when asking patients to rate disease activity or the impact of disease. Patients with RA will rarely score zero on a PtGA, commenting that activity or impact of RA 'can't be zero', given the persistent impact of disease damage and co-morbidities, wanting to encompass the varying nature of chronic illness in any self-assessment and concern about withdrawal of treatment if disease is reported to be under excellent control (53). It has been identified that in RA there is a pressing need to reckon with these limitations of the content validity of PtGA (56). Patient-reported outcome assessments should permit accurate documentation of patients' disease status, and the current PtGAs used in RA do not seem to adequately give voice to patients' experiences of their disease (56). The inclusion of the PtGA in many measures of disease remission risks over-treatment because the floor effect of the PtGA and therefore risks incorporating this same effect into composite measures of low disease activity and, or remission.

Discussion

There is a lack of consensus as to what concept of disease the PtGA and PhGA should be measuring in clinical trials of SSc. Disease constructs such as activity, damage, and severity, as well as overall health, are not directly interchangeable and it is unknown

which construct best measures patient status; this issue applies to both the patient and physician global assessment. Furthermore, the heterogeneity in what 'global' or 'disease' means to individuals may reduce the precision of any global assessment tool and therefore any composite outcome measure that includes a global assessment. Global constructs of disease status are open to interpretation by both the patient and physician, and responses are highly dependent on the respondent's comprehension of the underlying construct being measured. An example of this is the absence of meaningful change of the PhGA and PtGA in the RCTs of tocilizumab (28, 29), despite important changes in pulmonary domains of disease. It also illustrates an inherent limitation of both the PtGA and PhGA, in that positive effects of therapy on an individual domain may by diluted in a global assessment because of the absence of treatment efficacy on other domains of disease. Demonstrating an absence of statistical difference of an unclear end point, such as the PtGA and PhGA as they are currently used, does not rule out a meaningful effect size in favour of either the test or the control intervention or an absence of a clinically meaningful difference to patients (13). Therefore, it seems reasonable to question whether several domains should be included in a global assessment instrument, such as the accepted patient-reported assessment for cryopyrin-associated periodic syndromes that incorporates a severity rating of individual disease-related symptoms (57).

The authors of this manuscript have been involved in a task force assessing various measures that assess the global aspects of SSc. The concept of measuring a global disease was defined by consensus as measuring the effect of the disease on more than one organ or body system and it should reflect how a patient feels or functions without reference to one specific body system. The focus of legacy global measures can be

classified into several domains: health-related quality of life, function, psychosocial, pain, fatigue, disease activity, disease severity, disease damage, frailty, and composite indices, any of which likely cross multiple disease domains. It is thus not surprising that there are many different patient and physician global assessments that ask different questions. No PtGA or PhGA has been systematically designed for use in SSc or has undergone full validation according to the OMERACT filter of truth, discrimination, and feasibility (11). No study has attempted to identify which global disease construct should be assessed by either the PtGA or PhGA nor ascertain which of the commonly used constructs of disease activity, severity, damage, or overall health is the most informative when measuring response to therapy. None of the global assessments currently used in research in SSc including patient involvement in the development of the question asked or design of item responses.

Review of the performance of global assessments in clinical studies of SSc has not shown substantial sensitivity to change of either PtGA or PhGA. This has implications for the use of global assessments in clinical trials where responsiveness to change is key to being able to identify therapeutic benefit. However, the majority of RCTs in SSc have failed to reach their primary endpoint, thus the lack of difference between study arms may speak to a failure of therapy rather than the insensitivity of an outcome measure to detect change in health status. In a condition such as SSc, effect sizes and minimal clinically important differences are influenced by burden of disease as well as disease duration (9, 58). Potential floor and ceiling effects are influenced by the wording of both the question anchors and potential item responses. Changes in any of these elements of a global assessment tool affects the sensitivity to change of the item in use (54).

The heterogeneity of SSc provides unique challenges concerning patient global health assessments. The gradual accrual of disease-related morbidity could result in shifting patient perceptions of global disease severity being influenced by recent disease events. For example, digital ulcer occurrence may lead to a patient concluding that their previous global disease severity ratings should not have been so high in the time preceding their first digital ulcer. A response shift of global assessments has been observed, particularly if status is compared to a previous visit rather than a baseline study visit (13). Conversely, habituation and adaptation may favourably impact longitudinal global disease severity assessment despite the disease manifestations remaining unchanged or potentially worsening. Both the coping strategies and styles of patients can influence their global rating of disease status (59) and conceivably the concept of 'global' disease may vary depending on the disease subtype, duration of disease and particular organ involvement a patient had experienced. Many of these issues could also impact physician global assessments.

Future directions: a new research agenda

There is an urgent imperative to develop fully valid and universally applied global assessments in clinical trials of SSc. A suggested research methodology to first define and then validate both the patient and physician global assessments is outlined in Figure 1. The most important phase is the initial phase to define the construct to be measured and then developing appropriate wording of the question and scale anchors to best reflect the domain under consideration and thus improve the content validity of global assessments. This step is essential to be able to demonstrate construct validity of the novel instrument. Without face and content validity, any presumed construct validity

demonstrated by a correlation of the novel instrument with other outcome measures hypothesised to assess a similar disease construct may well be spurious. Perhaps multiple PtGA and PhGA assessments should be developed to measure each of several different domains of disease and that consensus methodology be used to decide which domain can be used alone and which should be included in composite indices.

Rigorous prospective studies are required to test the discriminative capacity and determine the sensitivity to change and minimal clinically important difference of any standardised global assessment. Successful ascertainment of these values will require testing of the novel PtGA and PhGA in groups of patients treated with efficacious therapies compared to placebo. Assessment of the test-retest reliability, particularly of the PhGA, will require major resources, including significant physician time, to have multiple physicians clinically reassess individual patients with stable disease after an appropriate interval. Careful consideration needs to be paid to varying thresholds that may need to apply according to individual patients' disease subtype, duration, baseline values, and perhaps even specific organ manifestations.

Both the patient and physician global assessment are attractive outcome measures as they provide complimentary information that directly reflects the patient experience of their disease in the case of the patient global, and direct evaluation by the physician caring for the patient. Additionally, they can capture elements of disease that have proven elusive to measure by more objective measures of disease status through investigation findings or biomarkers. Currently, however, this review of both the work to date in SSc as well as other rheumatic diseases suggests that the heterogeneity of the wording of global assessments, and the lack of rigorous methodology in their

development, leads to an absence of content validity of these assessments and limits their validity as outcome measures for use in RCTs. The lack of content validity of these global measures across multiple rheumatic diseases suggests that this may be a global issue rather than specific to a particular disease. This raises concerns about the use of multi-item measures that include upon both patient and physician assessments to generate a final overall score. The lack of consistency of global assessments between studies undermines the results of validation studies of composite measures that include the PtGA or PhGA and risk negatively affecting these findings and spuriously limiting the use of novel composite outcomes as valid study endpoints.

In conclusion, future work needs to address the most valid constructs to measure using the PtGA and PhGA in RCTs in SSc. Once constructs that best capture aspects of disease important to both patients and physicians are defined, then methods to define appropriate questions, anchors, and item responses must be developed and employed in creating PtGA and PhGAs before considering the reliability, construct validity, and sensitivity to change of these global assessments. Without consistently applied, valid PtGA and PhGAs across RCTs in SSc, there is an ongoing risk of 'negative' trials due to a measurement failure rather than a failure of therapy.

Table 1: Clinimetric properties of the patient and physician global assessment in systemic sclerosis

		Truth			**Discrimination**			
	Feasibility	*Face validity*	*Content validity*	*Construct validity*	*Discriminant validity*	*Sensitivity to change*	*MCID*	*Reliability*
PtGA	Yes	Not done	No	No	No	No	P	No
PhGA	Yes	Not done	No	No	P	No	P	No

Abbreviations: P: partial; PhGA: physician global assessment; PtGA: patient global assessment; MCID: minimal clinical important difference

Table 2: Summary of types of patient and physician global scores and frequency of application in randomised controlled trials of systemic sclerosis

Features of *patient* global assessments (n=16)	**Frequency of use in RCTs**	**References**	**Features of *physician* global assessments (n=16)**	**Frequency of use in RCTs**	**References**
RCTs using VAS	7 / 16 (43.8%)	(5, 26, 28, 29, 31, 60, 61)	*RCTs using VAS*	6 / 16 (37.5%)	(26, 28, 29, 31, 60, 61)
100mm VAS	5 / 7 (71.4%)	(5, 26, 28, 29, 60)	100mm VAS	4 / 6 (66.7%)	(26, 28, 29, 60)
RCTs using NRS	5 / 16 (31.3%)	(62-66)	*RCTs using NRS*	5 / 16 (31.3%)	(62-66)
11 point NRS	5 / 5 (100%)	(62-66)	11 point NRS	5 / 5 (100%)	(62-66)
RCTs using Likert scale	4 / 16 (25%)	(25, 60, 65, 67)	*RCTs using Likert scale*	4 / 16 (25%)	(2, 25, 65, 67)
11 point Likert	2 / 4 (50%)	(25, 67)	11 point Likert	2 /4 (50%)	(25, 67)
7 point Likert	0 / 4 (0%)		7 point Likert	1/4 (25%)	(2)
5 point Likert	2 / 4 (50%)	(60, 65)	5 point Likert	1/4 (25%)	(65)
Assessment of disease activity	3 / 16 (18.8%)	(61, 62, 64)	Assessment of disease activity	4 / 16 (25%)	(61, 63, 64, 67)
Assessment of disease severity or overall	2 / 16 (12.5%)	(31, 63)	Assessment of disease severity	3 / 16 (18.8%)	(28, 31, 62)

impact of disease			/ overall impact of disease		
Assessment of current SSc status / overall disease	4 / 16 (25%)	(25, 28, 29, 67)	Assessment of current SSc status / overall disease	2 / 16 (12.5%)	(25, 29)
Overall health / general wellbeing	3 / 16 (18.8%)	(5, 66, 68)	Global health assessment	3 / 16 (18.8%)	(2, 66, 68)
Health transition question	2 / 16 (12.5%)	(60, 65)	Health transition question	1 / 16 (6.3%)	(65)
Global assessment not otherwise specified	2 / 16 (12.5%)	(26, 60)	Global assessment not otherwise specified	2 /16 (12.5%)	(26, 60)
Assessment of activity or severity	1 / 16 (6.3%)	(65)	Assessment of activity or severity	1 / 16 (6.3%)	(65)
Pain and disability	1 / 16 (6.3%)	(69)	Pain and disability	1 / 16 (6.3%)	(69)

Abbreviations: NRS: numerical rating scale; RCT: randomised controlled trial; SSc: systemic sclerosis; VAS: visual analogue scale

References

1. Nikiphorou E, Radner H, Chatzidionysiou K, Desthieux C, Zabalan C, van Eijk-Hustings Y, et al. Patient global assessment in measuring disease activity in rheumatoid arthritis: a review of the literature. Arthritis Res Ther. 2016;18:251.

2. Clements P, Furst D, Wong WK, Mayes M, White B, Wigley F, et al. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis: analysis of a two-year, double-blind, randomized, controlled clinical trial. Arthritis Rheum. 1999;42:1194-203.

3. Gruber BL, Kaufman LD. A double-blind randomized controlled trial of ketotifen versus placebo in early diffuse scleroderma. Arthritis Rheum. 1991;34:362-6.

4. Casas JA, Saway PA, Villarreal I, Nolte C, Menajovsky BL, Escudero EE, et al. 5-Fluorouracil in the treatment of scleroderma: a randomised, double blind, placebo controlled international collaborative study. Ann Rheum Dis. 1990;49:926-8.

5. van den Hoogen F, Boerbooms AMT, Swaak AJG, Rasker JJ, van Lier HJJ, van de Putte L. 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.

6. White B, Bauer E, Goldsmith LA, Hochberg M, Katz LM, Korn JH, et al. Guidelines for clinical trials in systemic sclerosis (scleroderma). Arthritis Rheum. 1995;38:351-60.

7. 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.

8. Khanna D, Berrocal VJ, Giannini EH, Seibold J, Merkel PA, Mayes M, et al. The American College of Rheumatology provisional composite response index for clinical trials in early diffuse cutaneous systemic sclerosis. Arthritis Rheum. 2016;68:299-311.

9. Hudson M, Impens A, Baron M, Seibold J, Thombs B, Walker J, et al. Discordance between patient and physician assessments of disease severity in systemic sclerosis. J Rheumatol. 2010;37:2307-12.

10. Wiese AB, Berrocal VJ, Furst DE, Seibold JR, Merkel PA, Mayes MD, et al. Correlates and responsiveness to change of measures of skin and musculoskeletal disease in early diffuse systemic sclerosis. Arthritis Care Res. 2014;66:1731-9.

11. Tugwell P, Boers M, D'Agostino MA, Beaton D, Boonen A, Bingham CO, et al. Updating the OMERACT filter: implications of filter 2.0 to select outcome instruments through assessment of "truth": content, face, and construct validity. J Rheumatol. 2014;41:1000-4.

12. Renskers L, van Uden R, Huis AMP, Rongen SAA, Teerenstra S, van Riel P. Comparison of the construct validity and reproducibility of four different types of patient-reported outcome measures (PROMs) in patients with rheumatoid arthritis. Clin Rheumatol. 2018;37:3191-9.

13. Powers JH, Patrick DL, Walton MK, Marquis P, Cano S, Hobart J, et al. Clinician-reported outcome assessments of treatment benefit: report of the ISPOR clinical outcome assessment emerging good practices task force. Value Health. 2017;20:2-14.

14. Pauling J, Frech T, Domsic R, Hudson M. Patient participation in patient-reported outcome instrument development in systemic sclerosis. Clin Exp Rheumatol. 2017;35:S184-S92.

15. Steen VD, Medsger TA. 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.

16. Valentini G, Bencivelli W, Bombardieri S, D'Angelo S, Della Rossa A, Silman AJ, et al. European Scleroderma Study Group to define disease activity criteria for systemic sclerosis. III. Assessment of the construct validity of the preliminary activity criteria. Ann Rheum Dis. 2003;62:901-3.

17. Fan X, Pope J, Canadian Scleroderma Research Group, Baron M. What is the relationship between disease activity, severity and damage in a large Canadian systemic sclerosis cohort? Results from the Canadian Scleroderma Research Group (CSRG). Rheumatol Int. 2010;30:1205-10.

18. Hudson M, Steele R, Canadian Scleroderma Research Group, Baron M. Update on indices of disease activity in systemic sclerosis. Semin Arthritis Rheum. 2007;37:93-8.

19. Ross L, Stevens W, Wilson M, Huq M, Strickland G, Walker J, et al. Performance of the 2017 EUSTAR activity index in an scleroderma cohort. Clin Rheumatol. 2020;39:3701-5.

20. Harel D, Hudson M, Iliescu A, Baron M, Canadian Scleroderma Research Group, Steele R. Summed and weighted summary scores for the Medsger disease severity scale compared with the physician's global assessment of disease severity in systemic sclerosis. J Rheumatol. 2016;43:1510-8.

21. Sekhon S, Pope J, Canadian Scleroderma Research Group, Baron M. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010;37:591-8.

22. Valentini G, Iudici M, Walker UA, Jaeger VK, Baron M, Carreira P, et al. The European Scleroderma Trials and Research group (EUSTAR) task force for the development of revised activity criteria for systemic sclerosis: derivation and validation of a preliminarily revised EUSTAR activity index. Ann Rheum Dis. 2017;76:270-6.

23. Pope J, Bellamy N. Outcome Measurement in Scleroderma Clinical Trials. Semin Arthritis Rheum. 1993;23:22-33.

24. Khanna D, Huang S, Lin CJF, Spino C. New composite endpoint in early diffuse cutaneous systemic sclerosis: revisiting the provisional American College of Rheumatology Composite Response Index in Systemic Sclerosis. Ann Rheum Dis. 2021;80:641-50.

25. Khanna D, Spino C, Johnson S, Chung L, Whitfield ML, Denton CP, et al. Abatacept in early diffuse cutaneous systemic sclerosis: results of a phase II investigator- initiated, multicenter, double- blind, randomized, placebo- controlled trial. Arthritis Rheumatol. 2019;72:125-36.

26. Pope J, Bellamy N, Seibold J, Baron M, Ellman M, Carette S, et al. A randomized, controlled trial of methotrexate versus placebo in early diffuse scleroderma. Arthritis Rheum. 2001;44:1351-8.

27. Clements P. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis trial: lessons learned. Semin Arthritis Rheum. 2004;33:249-63.

28. Khanna D, Denton CP, Jahreis A, van Laar JM, Frech TM, Anderson ME, et al. Safety and efficacy of subcutaneous tocilizumab in adults with systemic sclerosis (faSScinate): a phase 2, randomised, controlled trial. Lancet. 2016;387:2630-40.

29. Khanna D, Lin CJF, Furst DE, Goldin J, Kim G, Kuwana M, et al. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Respir Med. 2020;8:963-74.

30. 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.

31. 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.

32. Pope J. 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). Arthritis Care Res. 2011;63:S98-111.

33. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual analog scale for pain (VAS pain), numeric rating scale for pain (NRS pain), McGill pain questionnaire (MPQ), short-form McGill pain questionnaire (SF-MPQ), chronic pain grade scale (CPGS), short form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP). Arthritis Care Res. 2011;63:S240-52.

34. Prevoo MLL, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte L, van Riel PL. Modified disease activity scores that include twenty-eight joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38:44-8.

35. Smolen JS, Breedveld FC, Schiff MH, Kalden JR, Emery P, Eberl G, et al. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology (Oxford). 2003;42:244-57.

36. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbiel S, Machold KP, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis Res Ther. 2005;7:R796-R806.

37. Pincus T, Swearingen CJ, Bergman M, Yazici Y. RAPID3 (Routine Assessment of Patient Index Data 3), a rheumatoid arthritis index without formal joint counts for routine care: proposed severity categories compared to disease activity score and clinical disease activity index categories. J Rheumatol. 2008;35:2136-47.

38. Orbai AM, de Wit M, Mease P, Shea JA, Gossec L, Leung YY, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis. 2017;76:673-80.

39. Strand V, Van den Bosch F, Ranza R, Leung YY, Drescher E, Zueger P, et al. Patient-reported outcomes in psoriatic arthritis patients with an inadequate response to biologic disease-modifying antirheumatic drugs: SELECT-PsA 2. Rheumatol Ther. 2021;8:1827-44.

40. Nash P, Kirkham B, Okada M, Rahman P, Combe B, Burmester G-R, et al. Ixekizumab for the treatment of patients with active psoriatic arthritis and an inadequate response to tumour necrosis factor inhibitors: results from the 24-week randomised, double-blind, placebo-controlled period of the SPIRIT-P2 phase 3 trial. Lancet. 2017;389:2317-27.

41. Strand V, Mease PJ, Soriano ER, Kishimoto M, Salvarani C, Saffore CD, et al. Improvement in patient-reported outcomes in patients with psoriatic arthritis treated

with upadacitinib versus placebo or adalimumab: Results from SELECT-PsA 1. Rheumatol Ther. 2021;8:1789-808.

42. Cauli A, Gladman DD, Mathieu A, Olivieri I, Porru G, Tak PP, et al. Patient global assessment in psoriatic arthritis: a multicenter GRAPPA and OMERACT study. J Rheumatol. 2011;38:898-903.

43. Tälli S, Etcheto A, Fautrel B, Balanescu A, Braun J, Cañete JD, et al. Patient global assessment in psoriatic arthritis - what does it mean? An analysis of 223 patients from the psoriatic arthritis impact of disease (PsAID) study. Joint Bone Spine. 2016;83:335-40.

44. Perez-Chada LM, Salame NF, Ford AR, Duffin KC, Garg A, Gottlieb AB, et al. Investigator and patient global assessment measures for psoriasis clinical trials: A systematic review on measurement properties from the International Dermatology Outcome Measures (IDEOM) Initiative. Am J Clin Dermatol. 2020;21:323-38.

45. Berti A, Boleto G, Merkel PA, Tomasson G, Monti S, Quinn KA, et al. Psychometric properties of outcome measurement instruments for ANCA-associated vasculitis: a systematic literature review. Rheumatology (Oxford). 2022.epub ahead of print doi:10.1093/rheumatology/keac175.

46. Tomasson G, Farrar JT, Cuthbertson D, McAlear CA, Ashdown S, Cronholm PF, et al. Feasibility and construct validation of the patient reported outcomes measurement information system in systemic vasculitis. J Rheumatol. 2019;46:928-34.

47. Rimland CA, Quinn KA, Rosenblum JS, Schwartz MN, Bates Gribbons K, Novakovich E, et al. Outcome measures in large vessel vasculitis: Relationship between patient-, physician-, imaging-, and laboratory-based assessments. Arthritis Care Res. 2020;72:1296-304.

48. Furie RA, Petri MA, Wallace DJ, Ginzler EM, Merrill JT, Stohl W, et al. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Rheum. 2009;61:1143-51.

49. Gladman DD, Ibanez D, Urowitz M. Systemic Lupus Erythematosus Disease Activity Index 2000. J Rheumatol. 2002;29:288-91.

50. Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN, et al. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology (Oxford). 2005;44:902-6.

51. Chessa E, Piga M, Floris A, Devilliers H, Cauli A, Arnaud L. Use of physician global assessment in systemic lupus erythematosus: a systematic review of its psychometric properties. Rheumatology (Oxford). 2020;59:3622-32.

52. Piga M, Chessa E, Morand EF, Ugarte-Gil MF, Tektonidou M, van Vollenhoven R, et al. Physician global assessment international standardisation consensus in systemic lupus erythematosus: the PISCOS study. Lancet Rheumatol. 2022;4:e441-e9.

53. Ferreira RJO, de Wit M, Henriques M, Pinto AF, Duarte C, Mateus E, et al. 'It can't be zero!' Difficulties in completing patient global assessment in rheumatoid arthritis: a mixed methods study. Rheumatology (Oxford). 2020;59:1137-47.

54. Pang PS, Collins SP, Sauser K, Andrei AC, Storrow AB, Hollander JE, et al. Assessment of dyspnea early in acute heart failure: patient characteristics and response differences between likert and visual analog scales. Acad Emerg Med. 2014;21:659-66.

55. Hirsh J, Wood P, Keniston A, Peng M, Ramaswami S, Caplan L, et al. Limited health literacy and patient confusion about rheumatoid arthritis patient global assessments and model disease states. Arthritis Care Res. 2019;71:611-9.

56. De Cock D, Hirsh J. The rheumatoid arthritis patient global assessment: improve it or lose it! Rheumatology (Oxford). 2020;59:923-4.

57. U.S Food and Drug Administration Center for Drug Evaluation and Research: Clinical Outcome Assessment (COA) Compendium. 2021 https://www.fda.gov/media/130138/download.

58. Pauling JD, Reilly E, Smith T, Frech TM. Evolving symptoms of Raynaud's phenomenon in systemic sclerosis are associated with physician and patient-reported assessments of disease severity. Arthritis Care Res. 2019;71:1119-26.

59. DiRenzo DD, Smith TR, Frech TM, Shah AA, Pauling JD. Effect of coping strategies on patient and physician perceptions of disease severity and disability in systemic sclerosis. J Rheumatol. 2021;48:1569-73.

60. Pope J, McBain D, Petrlich L, Watson S, Vanderhoek L, de Leon F, et al. Imatinib in active diffuse cutaneous systemic sclerosis: Results of a six-month, randomized, double-blind, placebo-controlled, proof-of-concept pilot study at a single center. Arthritis Rheum. 2011;63:3547-51.

61. Assistance Publique - Hopitaux de Paris. Safety and efficacy of itacitinib in adults with systemic sclerosis (SCLERITA). https://clinicaltrials.gov/ct2/show/NCT04789850?term=itacitinib&cond=Systemic+Sclerosis&draw=1&rank=1.

62. Emerald Health Pharmaceuticals Inc. Evaluation of safety, tolerability and preliminary efficacy of EHP-101 in diffuse cutaneous systemic sclerosis. https://clinicaltrials.gov/ct2/show/NCT04166552?term=NCT04166552&cond=Systemic+Sclerosis&draw=2&rank=1.

63. CSL Behring. Efficacy and safety of IgPro10 in adults with systemic sclerosis (SSc).

https://clinicaltrials.gov/ct2/show/NCT04138485?term=NCT+04138485&cond=Systemic+Sclerosis&draw=2&rank=1.

64. Spiera R, Hummers L, Chung L, Frech TM, Domsic R, Hsu V, et al. Safety and efficacy of lenabasum in a phase II, randomized, placebo-controlled trial in adults with systemic sclerosis. Arthritis Rheumatol. 2020;72:1350-60.

65. Khanna D, Allanore Y, Denton CP, Kuwana M, Matucci-Cerinic M, Pope JE, et al. Riociguat in patients with early diffuse cutaneous systemic sclerosis (RISE-SSc): randomised, double-blind, placebo-controlled multicentre trial. Ann Rheum Dis. 2020;79:618-25.

66. Corbus Pharmaceuticals Inc.. Trial to evaluate efficacy and safety of lenabasum in diffuse cutaneous systemic sclerosis (RESOLVE-1). https://clinicaltrials.gov/ct2/show/NCT03398837?term=lenabasum&cond=Systemic+Sclerosis&draw=2&rank=1.

67. Nagaraja V, Spino C, Bush E, Tsou PS, Domsic RT, Lafyatis R, et al. A multicenter randomized, double-blind, placebo-controlled pilot study to assess the efficacy and safety of riociguat in systemic sclerosis-associated digital ulcers. Arthritis Res Ther. 2019;21:202.

68. University College London. Mycophenolate in limited cutaneous systemic sclerosis (MINIMISE-Pilot) (MINIMISE). https://clinicaltrials.gov/ct2/show/NCT04927390?term=MINIMISE&cond=Systemic+Sclerosis&draw=2&rank=1.

69. Herrick A. Prednisolone in Early Diffuse Systemic Sclerosis (PRedSS). https://clinicaltrials.gov/ct2/show/NCT03708718?term=prednisolone&cond=systemic+sclerosis&draw=2&rank=1.

Figure 1: Proposed methodology for development and validation of patient and physician global assessments according to OMERACT Filter

Abbreviations: PhGA: physician global assessment; PtGA: patient global assessment;

RCT: randomised controlled trial; SSc: systemic sclerosis