



Published in final edited form as:

*Ann Rheum Dis.* 2008 May ; 67(5): . doi:10.1136/ard.2007.078923.

## Development of a provisional core set of response measures for clinical trials of systemic sclerosis

D Khanna<sup>1</sup>, D J Lovell<sup>2</sup>, E Giannini<sup>2</sup>, P J Clements<sup>1</sup>, P A Merkel<sup>3</sup>, J R Seibold<sup>4</sup>, M Matucci-Ceric<sup>5</sup>, C P Denton<sup>6</sup>, M D Mayes<sup>7</sup>, V D Steen<sup>8</sup>, J Varga<sup>9</sup>, D E Furst<sup>1</sup>, and for the Scleroderma Clinical Trials Consortium co-authors

<sup>1</sup>Division of Rheumatology, Department of Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California, USA

<sup>2</sup>Division of Rheumatology, Cincinnati Children's Medical Center, Cincinnati, Ohio, USA

<sup>3</sup>Section of Rheumatology and the Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Massachusetts, USA

<sup>4</sup>University of Michigan Scleroderma Program, Ann Arbor, Michigan, USA

<sup>5</sup>University of Florence, Florence, Italy

<sup>6</sup>Royal Free Hospital, London, UK

<sup>7</sup>Division of Rheumatology and Clinical Immunogenetics, The University of Texas Houston Medical School, Houston, Texas, USA

<sup>8</sup>Division of Rheumatology, Georgetown University Medical Center, Washington DC, USA

<sup>9</sup>Division of Rheumatology, Northwestern University Medical School, Chicago, Illinois, USA

### Abstract

**Objective**—To develop a provisional core set of response measures for clinical trials of systemic sclerosis (SSc).

**Methods**—The Scleroderma Clinical Trials Consortium (SCTC) conducted a structured, 3-round Delphi exercise to reach consensus on a core set of measures for clinical trials of SSc. Round 1 asked the SCTC investigators to list items in 11 pre-defined domains (skin, musculoskeletal, cardiac, pulmonary, cardio-pulmonary, gastrointestinal, renal, Raynaud phenomenon and digital ulcers, health-related quality of life and function, global health, and biomarkers) for SSc clinical trials. Round 2 asked respondents to rate the importance of the chosen items and was followed by a meeting, during which the Steering Committee discussed the feasibility, reliability, redundancy and validity of the items. Round 3 sought to obtain broader consensus on the core set measures. Members also voted on items that had data on feasibility but lacked data on reliability and validity, but may still be useful research outcome measures for future trials.

**Results**—A total of 50 SCTC investigators participated in round 1, providing 212 unique items for the 11 domains. In all, 46 (92%) participants responded in round 2 and rated 177 items. The ratings of 177 items were reviewed by the Steering Committee and 31 items from the 11 domains were judged to be appropriate for inclusion in a 1-year multi-centre clinical trial. In total, 40

---

Correspondence to: D Khanna, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine, 1000 Veterans Avenue, Rm 32-59 Rehabilitation Building, Los Angeles, California 90095, USA; dkhanna@mednet.ucla.edu.

**Competing interests:** None declared.

SCTC investigators completed round 3 and ranked 30 of 31 items as acceptable for inclusion in the core set. The Steering Committee also proposed 14 items for a research agenda.

**Conclusion**—Using a Delphi exercise, we have developed a provisional core set of measures for assessment of disease activity and severity in clinical trials of SSc.

Systemic sclerosis (scleroderma, SSc) has the highest case-specific mortality of any rheumatic disease, is associated with substantial morbidity,<sup>1</sup> and has many detrimental effects on health-related quality of life.<sup>2</sup> There has been significant progress over the past 10 years in the development and validation of outcome measures<sup>3</sup> and refinement of trial methodology in SSc.<sup>4</sup> This is paralleled by an increased understanding of the pathogenesis of SSc<sup>5,6</sup> and development of targeted therapies.<sup>7</sup> In order to make progress in therapy for SSc, researchers need a standardised core set of response measures that encompasses the complexity and clinical heterogeneity of SSc. Once developed, future research will be needed to determine the most appropriate algorithm that incorporates these core set measures into a feasible, reliable, and valid composite response index for defining clinically important improvement and worsening. Identification of core set measures is prerequisite to the development of a response index.

We conducted a structured Delphi exercise, with participation of the Scleroderma Clinical Trials Consortium (SCTC) membership, to reach consensus on a provisional core set of measures for SSc clinical trials.

## METHODS

A multisystem process was used to develop the core set of measures, which are described below and summarised in fig 1.

This project is an extension of work previously performed under the auspices of the Outcome Measures in Rheumatology Clinical Trials (OMERACT).<sup>8–10</sup> The project was conducted using well recognised consensus formation methodology, namely the Delphi technique and the Nominal Group technique.<sup>11</sup> The Delphi technique is a method for systematic solicitation and collation of judgments on a particular topic through a set of carefully designed sequential questionnaires interspersed with summarised information and feedback of opinions derived from earlier responses. It is a group process that utilises written responses as opposed to bringing individuals together. By employing a process of soliciting information in questionnaires, aggregating the information, and repeating the questionnaires to gain agreement among the participants, the Delphi technique reaches consensus in an elegant and efficient fashion. The Nominal Group technique<sup>11</sup> is a structured discussion of each idea that is led by a facilitator, after which a voting process takes place to decide upon the ideas. The Delphi and Nominal Group techniques have previously been used successfully to develop outcome measures in other rheumatic diseases.<sup>12–14</sup>

### Steering Committee members meeting

The Steering Committee (listed in the Appendix) for the development of provisional core set measures consisted of nine international scleroderma experts (seven from the USA, two from Europe). Before the first Delphi exercise was conducted, Steering Committee members listed the following 11 domains of illness relevant to people with SSc: skin, musculoskeletal, cardiac, pulmonary, cardio-pulmonary, gastrointestinal, renal, Raynaud phenomenon and digital ulcers, health-related quality of life and function, global health, biomarkers (serological markers, immune function, vascular injury, disease activity, and fibrosis), and others (items not captured by the aforementioned domains).

## Structured Delphi exercise

We invited the SCTC investigators to participate in the development of a scleroderma response index for a 1-year multi-centre randomised clinical trial in SSc. The SCTC is a non-profit organisation dedicated to advancing clinical trial methodology and conduct to develop better treatments for SSc (<http://www.sctc-online.org>). The SCTC has an international membership of over 50 centres in the USA, Canada, South America, Asia, and Europe. At the time of the first Delphi exercise, there were 65 member institutions in the SCTC.

### Round 1

We initiated a Delphi exercise via e-mail in December 2005 and asked the 65 investigators of the SCTC to list the items in the 11 domains that could be used in development of a scleroderma combined response index for a 1-year multi-centre clinical trial. We did not differentiate between the limited and diffuse cutaneous forms of SSc. Two reminder e-mails were sent to the SCTC investigators, 3 weeks apart. Responses were either e-mailed or faxed.

Steering Committee members were asked to consider the proposed items for their usefulness in a hypothetical 1-year multi-centre clinical trial and assess their feasibility, reliability, and validity (face, content, construct, and criterion validity) based on the OMERACT filters.<sup>15</sup> Items considered either to not be feasible in multi-centre clinical trial or to lack reliability or validity data (based on recent observational studies and clinical trials) were dropped.

### Round 2

The investigators of the SCTC who responded to the first request were e-mailed a second Delphi exercise and were asked to rate the importance of the chosen items to be included in a definition of a scleroderma response index for a hypothetical 1-year multi-centre clinical trial. They were asked to rate each one on a scale of 1 (extremely inappropriate for a combined measure) to 9 (extremely appropriate for a combined measure). Two reminder e-mails were sent to the SCTC investigators.

### Follow-up meeting

Steering Committee members conducted a 1-day nominal group meeting in early August 2006 and discussed the issues of feasibility, reliability, redundancy, and validity, with special emphasis on responsiveness to change, based on the OMERACT filters.<sup>15</sup> Although all proposed items were considered, the Steering Committee concentrated on items from round 2 that had a median score of 6 or greater on a scale of 1 to 9. The Steering Committee was provided with key articles summarising the current available data on outcome measures.<sup>291016-19</sup> After comprehensive discussions, the Steering Committee voted on each item in the core set. A priori, at least 66% consensus was required (six out of nine members present) to accept or reject an item for the core set. We chose 66% instead of the traditional 80% as there were only nine members in the Steering Committee and an 80% consensus would have required agreement among eight of nine experts. This latter was considered an unrealistic and overly stringent requirement. For purposes of practicality and easy of use, the experts were also instructed to choose two to four core set items for each domain. The group also agreed to include items that appear highly promising and feasible for use in future clinical trials but lacked published data regarding reliability and validity.

### Round 3

We sought to obtain broader consensus on the core set measures from the investigators of the SCTC. A third round of the Delphi exercise was conducted to ensure that scleroderma

experts around the world agreed with the proposed core set items. This Delphi exercise was sent in October 2006 and investigators of the SCTC were asked to provide their agreement on a scale of 1 (complete disagreement on the core set measure for a combined measure) to 9 (complete agreement on the core set measure for a combined measure). All relevant articles were attached with the Delphi questionnaire.

### Statistical analysis

Statistical analyses were descriptive in nature. The results are reported as median (25th–75th percentile). Based on the RAND Corp./University of California, Los Angeles (RAND/UCLA) Appropriateness Method,<sup>20</sup> median scores in the 1–3 range during the third round of the Delphi were considered as inappropriate (complete lack of consensus), those in the 4–6 range were considered as uncertain (some consensus), and those in the 7–9 range were considered as appropriate (good to excellent consensus).

## RESULTS

Of the 65 SCTC investigators, 50 (77% response rate, 30 from the USA, 13 from Europe, 5 from Canada, 1 from Asia, and 1 from South America) provided their responses to Delphi round 1 exercise. The participants provided 212 unique items for the 11 domains. The Steering Committee members were asked to consider the 212 items for their usefulness in a 1-year multi-centre clinical trial and assess their feasibility, reliability, redundancy, and validity. Of the 212 items, the Steering Committee decided to retain 177 items; 20 were dropped due to lack of feasibility (eg, ultrasound for skin thickness) and 15 were dropped either due to lack of reliable or valid data in recent observational studies and clinical trials (eg, oral aperture measurement) or they did not have an available instrument to capture the underlying domain (eg, elasticity of the skin; data summarised in Furst *et al* (2005, 2007), Merkel *et al* and Khanna *et al*).<sup>8–10,21</sup> The combined cardio-pulmonary domain was dropped as the items overlapped completely (were redundant) with the separate cardiac and pulmonary domains. The previously combined Raynaud phenomenon/digital ulcer domain was separated into two domains: Raynaud phenomenon and digital ulcer.

Of the 50 participants, 46 (92%) responded in round 2 with their ratings of the 177 items within each domain. The ratings of the 177 items were provided to the Steering Committee members, and the 60 items with a median score of 6 or greater were discussed further.

Out of 60 items, some of the separate items were measured by the same outcome measure and were combined and resulted in 45 items. For example, echocardiogram with Doppler had four separate items—measurement of left ventricular parameters, diastolic function parameters, right ventricular parameters, and pulmonary artery pressure. These were combined into one item under cardiac domain (table 1).

Similarly, we combined items that were descriptive of right heart catheterisation (pulmonary artery pressure, pulmonary vascular resistance, wedge pressure, and cardiac output/cardiac index), pulmonary function tests (forced vital capacity, diffusion capacity, and total lung capacity), and dyspnoea indices (University of California San Diego Dyspnoea Questionnaire, St. George Respiratory Questionnaire, and Mahler Dyspnoea Index). Certain items, such as 24-h pH monitoring, were not considered to be feasible in a multi-centre study. Other items, such as palmar crease to digital tip measurement, were not responsive-to-change in recent clinical trials.<sup>22</sup>

The Steering Committee decided on 31 (out of 45) items in 11 domains (table 1). All chosen items were judged to be feasible for a 1-year multi-centre clinical trial. Of the 31 core set items, 4 were selected that were particularly suited for organ-specific trials: right heart

catheterisation, 6-min walk test, and Borg Dyspnoea Index were specifically encouraged for clinical trials where a drug is being assessed specifically for pulmonary hypertension; high-resolution computer tomography (HRCT) was encouraged for trials focusing on lung parenchymal disease. Some of the other core items such as tendon friction rubs, visual analogue scale to assess skin activity from patient's and doctor's perspectives, and scleroderma transition items were considered as extremely important by the Steering Committee during nominal group technique (although they lack complete published data on reliability and validity) and were also rated highly by the SCTC members during round 3 of the Delphi assessment.

A total of 14 of the 45 items were feasible<sup>231827-29</sup> but lacked sufficient data on reliability and validity and were included in the research agenda (table 2). In all, 40 SCTC investigators returned the Delphi 3 exercise (87%) via e-mail. The median score for 30 of 31 items was 7 (out of maximum 9) or greater (appropriate based on the RAND/UCLA Appropriateness Method).<sup>20</sup> The exception was tender joint count that received a score of 6. There was higher variability in the median score for the items chosen for the research agenda. The median score ranged from 5, for a measure of telangiectasia, to 8, for a scleroderma gastrointestinal tract instrument (table 2).

## DISCUSSION

Using Delphi and Nominal Group techniques, we have developed a provisional core set of measures for assessment of disease activity and severity in clinical trials of SSc. The proposed core measures should be applied to future observational and multi-centre clinical trials in SSc. These measures have been validated to approximately the same degree as were the items in the core set for the American College of Rheumatology Response Criteria for rheumatoid arthritis when those criteria were first published.<sup>3031</sup> The SSc core set of measures will facilitate the standardisation, conduct, reporting, and interpretation of clinical trials, and facilitate future metaanalyses and comparisons of therapies. In addition, core set measures are the first step in development of a fully validated composite response index.<sup>13143031</sup>

Nearly all the core set of measures are considered "standard of care" and are frequently used to assess patients with early SSc in clinical practice.<sup>103233</sup> For example, the modified Rodnan skin score, echocardiogram with Doppler, pulmonary function test, Scleroderma-Health Assessment Questionnaire Disability Index, and acute phase reactants are considered standard of care in the management of early SSc by many scleroderma experts.<sup>1032-34</sup> Other measures, such as the durometer, 6-min walk test, Raynaud Condition Score, HRCT of the lungs, tendon friction rubs, and visual analogue scale (VAS)/Likert scores to assess disease activity, have been found to be feasible, reliable, valid, and responsive to change in recent SSc multi-centre clinical trials.<sup>162335-40</sup> Additionally, the Steering Committee chose 11 domains because SSc is a multisystem disease involving different organ systems, including a significant detrimental effect on health-related quality of life (HRQOL) and functional disability.

One of the primary goals of the exercise was to avoid domains/items that cannot be measured in a multi-centre study. For example, measures such as MRI to assess pulmonary hypertension or ultrasound to assess skin thickness were not included due to the cost associated with these procedures and inter-reader variability associated with the techniques, respectively. Indeed, for items where there is a high inter-reader variability (such as echocardiogram and HRCT), central reading was strongly recommended. All items chosen in the provisional core set were considered to be feasible in a 1-year multi-centre study. For

example, although HRCT is associated with high cost, it was shown to be feasible in two recent SSc multi-centre studies, one in the US<sup>40</sup> and one in Europe.<sup>41</sup>

The patient and doctor-reported outcomes were included to be measured as either a VAS or a Likert scale. Different analyses have shown that VAS and Likert responses are highly correlated and yield similar results in patients with chronic diseases.<sup>42–44</sup> Since Likert responses are easier to administer and interpret, they may be preferable to a VAS. In addition, the Likert scales can be external anchors to assess minimally important differences.<sup>45</sup>

During round 3, there was good to excellent consensus on 30 of 31 core set measures in the 11 domains. The exception was tender joint count that received a score of 6 out of 9. We decided to keep the tender joint count in the core set measures as joint pain affects a majority of patients with SSc.

The Steering Committee also discussed a research agenda; measures that were not included in the core measures that appear very promising, but are not yet ready for clinical trials and hence may be useful outcomes measures for future clinical trials. Most of the items in the research agenda have been determined to be feasible, but reliability and validity—including responsiveness to change—has not been determined for multi-centre clinical trials. Although cardiac MRI to assess cardiac function and pulmonary artery is expensive, the data so far is encouraging<sup>46</sup> and so it was included in our research agenda. If MRI is shown to be a reliable and valid tool in future studies, it may be a useful outcome measure despite its cost (similar to HRCT). This list of items can be incorporated in future observational and clinical trials and, if found responsive, can be considered in future versions of the scleroderma composite response index.

Our study has both strengths and limitations. First, we were able to conduct a Delphi exercise using member institutions of the SCTC who conduct clinical treatment trials of new medications in SSc. Second, our response rate was good: 77% for round 1, 92% for round 2 and 80% for round 3. The principal limitations were, firstly, that the Delphi and Nominal Group techniques are consensus building approaches rather than data-driven approaches. Secondly, we chose a cut-off of 6 during rounds 1 and 2 as the majority of items received a median score of 6 during these rounds and we wanted to incorporate large numbers of items for the Steering Committee and SCTC members to vote on; a cut-off of 7 was chosen during round 3 to present items with good to excellent agreement. Thirdly, we chose a consensus of 66% instead of the traditional 80% during our Nominal Group technique assessment as only members of the Steering Committee (n=9) participated in this. Acknowledging this limitation, we obtained broader consensus on the core set measures from the investigators of the SCTC using a Delphi exercise where 30 of 31 items were rated as having good to excellent agreement regarding appropriateness for a 1-year multi-centre clinical trial.

We are currently performing a prospective longitudinal observational clinical study to gather information on the measurement characteristics of the items in the preliminary core set. This data will be used to further develop a reliable, valid and responsive set of SSc measures and also to develop a validated response index. This is being accomplished through prospective, data-driven, consensus building techniques to develop and quantitatively evaluate candidate definitions for a combined response index for SSc, as has been performed in other rheumatic diseases.<sup>12–14</sup>

## Acknowledgments

**Funding:** is project was funded by a National Institutes of Health Award (NIAMS U01 AR055057-01). DK was supported by the Scleroderma Foundation (New Investigator Award), a National Institutes of Health Award (NIAMS K23 AR053858-01A1), and a grant from the Scleroderma Clinical Trial Consortium. PAM is supported by a Mid-Career Clinical Investigator Award (NIAMS K24 AR2224-01A1).

## APPENDIX

### Scleroderma Clinical Trials Consortium steering committee

Co-Chairs: Dinesh Khanna; Daniel E Furst.

Committee members (in alphabetical order): Philip J Clements, Christopher P Denton, Edward Giannini, Daniel E Lovell, Maureen D Mayes, Marco Matucci-Cerinic, Peter A Merkel, James R Seibold and Virginia D Steen.

Co-authors (in alphabetical order): Baron, Murray, Jewish General Hospital, Montreal, Quebec, Canada. Csuka, Mary Ellen, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. Berezne, Alice, Hôpital Cochin, Paris, France. Briet, Samuel N, University of New South Wales, Sydney, Australia. Brühlmann, Pius, University Hospital, Zurich, Switzerland. Buch, Maya H, University of Michigan, Ann Arbor, Michigan, USA. Catoggio, Luis, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina. Collier, David, University of Colorado Health Sciences Center, Denver, Colorado, USA. Crofford, Leslie, University of Kentucky, Lexington, Kentucky, USA. Czirják, László, University of Pécs, Pécs, Hungary. Derk, Chris T, Jefferson Medical College, Philadelphia, Pennsylvania, USA. Distler, Oliver, University Hospital Zurich, Zurich, Switzerland. Doyle, Mittie Kelleher, Centocor, Malvern, Pennsylvania, USA. Farge-Bancel, Domonique, Hopital Saint-Louis, Paris, France. Fessler, Barri, University of Alabama at Birmingham, Alabama, USA. Foeldvari, Ivan, Klinikum Eilbek, Hamburg, Germany. Goldberg, Avram, Albert Einstein College of Medicine, Bronx, New York, USA. Gran, Jan Tore, Rikshospitalet, Oslo, Norway. Grau, Raffael, Indiana University, Indianapolis, Indiana, USA. Griffing, W Leroy, Mayo Clinic College of Medicine, Scottsdale, Arizona, USA. Hayat, Samina, Louisiana State University, Shreveport, Louisiana, USA. Herrick, Ariane L, University of Manchester, Manchester, UK. Hsu, Vivien, RWJ Medical School, New Brunswick, New Jersey, USA. Hummers, Laura K, Johns Hopkins University, Baltimore, Maryland, USA. İnanç, Murat, University of Istanbul, Istanbul, Turkey. Johnson, Sindhu, University of Toronto, Toronto, Ontario, Canada. Kahaleh, M Bashar, Medical University of Ohio, Toledo, Ohio, USA. Lafyatis, Robert A, Boston University Medical Campus, Boston, Massachusetts, USA. Lee, Peter, Mount Sinai Hospital, Toronto, Ontario, Canada. Mahmud, Tafazzul H, Shaikh Zayed Postgrad Med Institute, Lahore, Pakistan. Malcarne, Vanessa, San Diego State University, San Diego, California, USA. McHugh, Neil J, Royal National Hospital, Bath, UK. Martin Richard W, College of Human Medicine, Michigan State University, Grand Rapids, Michigan, USA. McKown, Kevin, University of Wisconsin Medical School, Madison, Wisconsin, USA. Medsger, Thomas A, Jr, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. Moreland, Larry, University of Alabama at Birmingham, Alabama, USA. Pope, Janet E, St Joseph Health Care London, London, Ontario, Canada. Rich, Eric, Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada. Rothfield, Naomi F, University of Connecticut, Farmington, Connecticut, USA. Schioppa, Elena, University of Michigan, Ann Arbor, Michigan, USA. Scorza, Raffaella, University of Milan, Milan, Italy. Senécal, Jean-Luc, Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada. Shanahan, Joseph, Duke University Medical Center, Durham, North Carolina, USA. Simms, Robert W, Boston University school of Medicine, Boston, Massachusetts, USA. Strand, Vibeke, Stanford University, Portola Valley, California, USA.

Silver, Richard M, Medical University of South Carolina, Charleston, South Carolina, USA. Sweiss, Nadera, University of Chicago, Chicago, Illinois, USA. Valentini, Gabriele, Second University of Napoli, Napoli, Italy. van den Hoogen, Frank H J, St Maartenskliniek and Radboud University Medical Centre, Nijmegen, The Netherlands. Veale, Douglas, St Vincent's University Hospital, Dublin, Ireland. Voskuyl, Alexander E, VU University Medical Center, Amsterdam, The Netherlands. Wigley, Fred, Johns Hopkins University, Baltimore, Maryland, USA. Wollheim, Frank A, Lund University Hospital, Lund, Sweden.

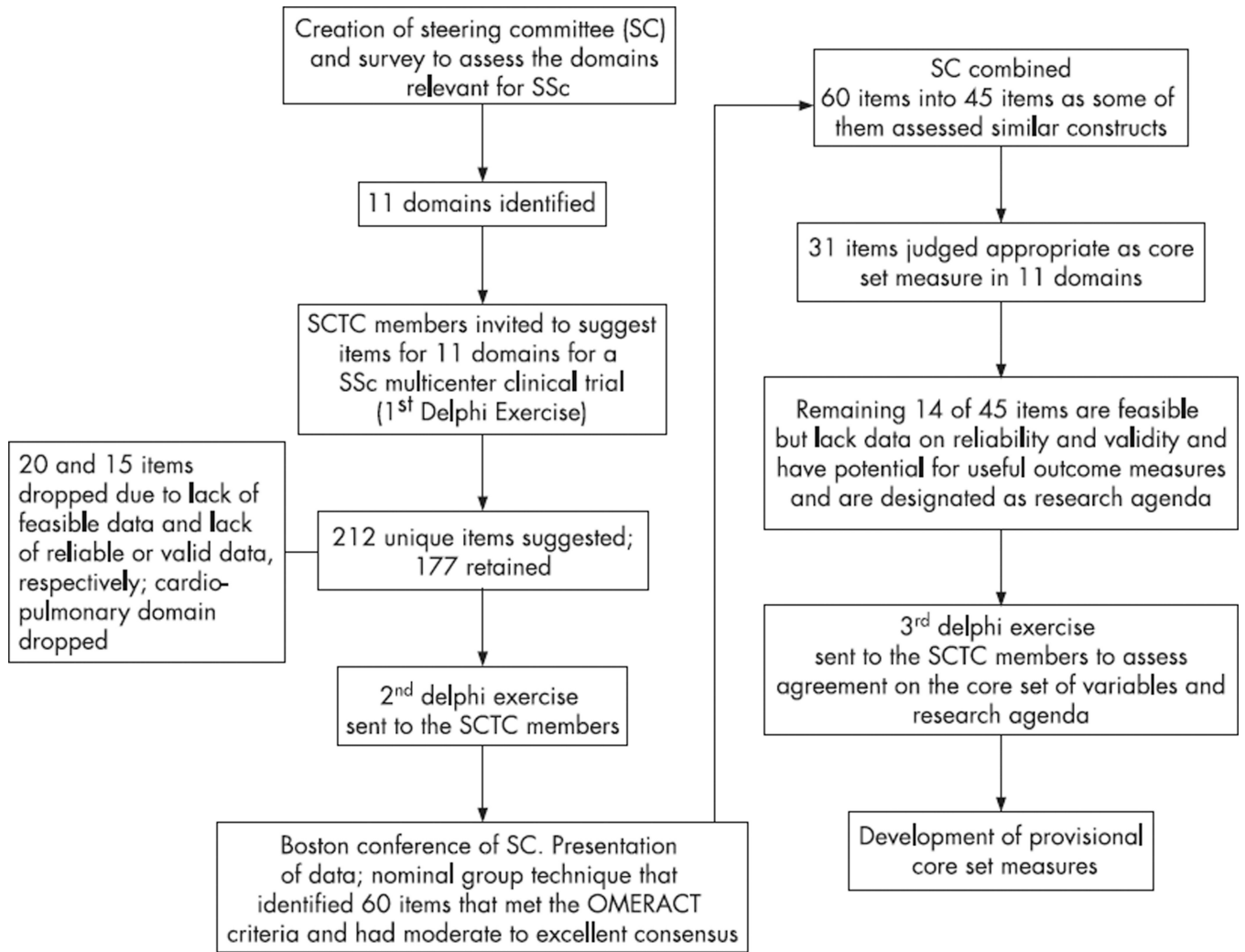
## REFERENCES

1. Altman RD, Medsger TA Jr, Bloch DA, Michel BA. Predictors of survival in systemic sclerosis (scleroderma). *Arthritis Rheum.* 1991; 34:403–413. [PubMed: 1901491]
2. Khanna D, Ahmed M, Furst DE, Ginsburg SS, Park GS, Hornung R, et al. Health values of patients with systemic sclerosis. *Arthritis Rheum.* 2007; 57:86–93. [PubMed: 17266070]
3. Khanna D, Hays RD, Park GS, Braun-Moscovici Y, Mayes MD, McNearney TA, et al. Development of the Scleroderma Gastrointestinal Tract 1.0 (SSc-GIT 1.0) Quality of Life Instrument. *Arthritis Rheum.* 2007; 57:1280–1286. [PubMed: 17907224]
4. White B, Bauer EA, Goldsmith LA, Hochberg MC, Katz LM, Korn JH, et al. Guidelines for clinical trials in systemic sclerosis (scleroderma). I. Disease-modifying interventions. The American College of Rheumatology Committee on Design and Outcomes in Clinical Trials in Systemic Sclerosis. *Arthritis Rheum.* 1995; 38:351–360. [PubMed: 7880189]
5. Charles C, Clements P, Furst DE. Systemic sclerosis: hypothesis-driven treatment strategies. *Lancet.* 2006; 367:1683–1691. [PubMed: 16714190]
6. Varga J. Scleroderma and Smads: dysfunctional Smad family dynamics culminating in fibrosis. *Arthritis Rheum.* 2002; 46:1703–1713. [PubMed: 12124852]
7. Denton CP, Merkel PA, Furst DE, Khanna D, Emery P, Hsu VM, et al. Recombinant human anti-transforming growth factor  $\beta$ 1 antibody therapy in systemic sclerosis: a multicenter, randomized, placebo-controlled phase I/II trial of CAT-192. *Arthritis Rheum.* 2006; 56:323–333. [PubMed: 17195236]
8. Furst D, Khanna D, Matucci-Cerinic M, Clements P, Steen V, Pope J, et al. Systemic sclerosis – continuing progress in developing clinical measures of response. *J Rheumatol.* 2007; 34:1194–1200. [PubMed: 17477486]
9. Furst DE, Khanna D, Matucci-Cerinic M, Silman AJ, Merkel PA, Foeldvari I. Scleroderma – developing measures of response. *J Rheumatol.* 2005; 32:2477–2480. [PubMed: 16331791]
10. Merkel PA, Clements PJ, Reveille JD, Suarez-Almazor ME, Valentini G, Furst DE. Current status of outcome measure development for clinical trials in systemic sclerosis. Report from OMERACT 6. *J Rheumatol.* 2003; 30:1630–1647. [PubMed: 12858472]
11. Delbecq, A.; Van de Ven, A. Group techniques for program planning: a guide to nominal group and delphi processes. Glenview, Illinois, USA: Scott, Foresman and Company; 1975.
12. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum.* 1995; 38:727–735. [PubMed: 7779114]
13. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum.* 1997; 40:1202–1209. [PubMed: 9214419]
14. Ruperto N, Ravelli A, Oliveira S, Alessio M, Mihaylova D, Pasic S, et al. The Pediatric Rheumatology International Trials Organization/American College of Rheumatology provisional criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: prospective validation of the definition of improvement. *Arthritis Rheum.* 2006; 55:355–363. [PubMed: 16739203]
15. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol.* 1998; 25:198–199. [PubMed: 9489805]
16. Kissin EY, Schiller AM, Gelbard RB, Anderson JJ, Falanga V, Simms RW, et al. Durometry for the assessment of skin disease in systemic sclerosis. *Arthritis Rheum.* 2006; 55:603–609. [PubMed: 16874783]



17. 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; 32:832–840. [PubMed: 15868618]
18. Impens AJ, Chung KC, Buch MC, Schioppa E, Kotsis S, Burns P, et al. Validation of the Michigan Hand Outcomes Questionnaire in systemic sclerosis. *Arthritis Rheum.* 2006; 54:5483.
19. Furst DE, Khanna D, Matucci-Cerinic M, Clements P, Steen V, Pope J, et al. Systemic sclerosis – continuing progress in developing clinical measures of response. OMERACT 2006. *J Rheumatol.* 2007; 34:1194–1200. [PubMed: 17477486]
20. Brook, RH. The RAND/UCLA appropriateness method. Rockville, Maryland, USA: Public Health Service; 1994. p. 59-70. Methodology perspectives AHCPH no. 95-0009.
21. Khanna D, Merkel PA. Outcome measures in systemic sclerosis: an update on instruments and current research. *Curr Rheumatol Rep.* 2007;151–157. [PubMed: 17502046]
22. Merkel PA, Silliman NP, Clements P, Denton CP, Furst DE, Mayes M, et al. Patterns and predictors of change in outcome measures in clinical trials in scleroderma. *Arthritis Rheum.* 2005; 52:S282–S283.
23. 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; 40:1984–1991. [PubMed: 9365087]
24. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum.* 1980; 23:137–145. [PubMed: 7362664]
25. Ware, J.; Kosinski, M.; Dewey, J. How to score version two of the SF-36 Health Survey. Lincoln, Rhode Island, USA: QualityMetric Inc; 2000.
26. 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; 46:2410–2420. [PubMed: 12355489]
27. Williams MH, Handler CE, Akram R, Smith CJ, Das C, Smee J, et al. Role of N-terminal brain natriuretic peptide (N-TproBNP) in scleroderma-associated pulmonary arterial hypertension. *Eur Heart J.* 2006; 26:1485–1494. [PubMed: 16682379]
28. Khanna D, Furst DE, Wong WK, Tsevat J, Clements PJ, Park GS, et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. *Qual Life Res.* 2007; 16:1083–1092. [PubMed: 17404896]
29. Medsger TA Jr, Silman AJ, Steen VD, Black CM, Akesson A, Bacon PA, et al. A disease severity scale for systemic sclerosis: development and testing. *J Rheumatol.* 1999; 26:2159–2167. [PubMed: 10529133]
30. Felson DT. Choosing a core set of disease activity measures for rheumatoid arthritis clinical trials. *J Rheumatol.* 1993; 20:531–534. [PubMed: 8478865]
31. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum.* 1993; 36:729–740. [PubMed: 8507213]
32. Sterz, M.; McCloskey, D. Management: holistic approach to systemic sclerosis. Nurse’s approach. In: Clements, PJ.; Furst, DE., editors. Systemic sclerosis. Philadelphia, Pennsylvania, USA: Lippincott, Williams and Wilkins; 2004. p. 385-392.
33. Wigley, FM.; Hummers, L. Management: holistic approach to systemic sclerosis. Physician’s approach. In: Clements, PJ.; Furst, DE., editors. Systemic sclerosis. Philadelphia, Pennsylvania, USA: Lippincott, Williams and Wilkins; 2004. p. 371-384.
34. Khanna D. Health-related quality of life: a primer with focus on scleroderma. *Scler Care Res.* 2006; 3:3–13.
35. Fries R, Shariat K, von Wilmowsky H, Bohm M. Sildenafil in the treatment of Raynaud’s phenomenon resistant to vasodilatory therapy. *Circulation.* 2005; 112:2980–2985. [PubMed: 16275885]
36. Lederer DJ, Arcasoy SM, Wilt JS, D’Ovidio F, Sonett JR, Kawut SM. Six-minute-walk distance predicts waiting list survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2006; 174:659–664. [PubMed: 16778159]

37. Buch MH, Denton CP, Furst DE, Guillevin L, Rubin LJ, Wells AU, et al. Sub-maximal exercise testing in the assessment of interstitial lung disease secondary to systemic sclerosis: reproducibility and correlations of the six minute walk test. *Ann Rheum Dis.* 2006 Jul 25. Published Online First:
38. Wigley FM, Korn JH, Csuka ME, Medsger TA Jr, Rothfield NF, Ellman M, et al. Oral iloprost treatment in patients with Raynaud's phenomenon secondary to systemic sclerosis: a multicenter, placebo-controlled, double-blind study. *Arthritis Rheum.* 1998; 41:670–677. [PubMed: 9550476]
39. Merkel PA, Silliman NP, Denton CP, Clements P, Emery P, et al. Validity, reliability, and feasibility of durometer measurements of scleroderma skin disease in a multicenter treatment trial. *Arthritis Rheum.* 2004; 50:S419.
40. 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; 354:2655–2666. [PubMed: 16790698]
41. Hoyles RK, Ellis RW, Wellsbury J, Lees B, Newlands P, Goh NS, et al. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis Rheum.* 2006; 206:3962–3970. [PubMed: 17133610]
42. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis.* 1987; 40:1129–1133. [PubMed: 3680471]
43. Pfenning L, Cohen L, van der PH. Preconditions for sensitivity in measuring change: visual analogue scales compared to rating scales in a Likert format. *Psychol Rep.* 1995; 77:475–480. [PubMed: 8559872]
44. Grant S, Aitchison T, Henderson E, Christie J, Zare S, McMurray J, et al. A comparison of the reproducibility and the sensitivity to change of visual analogue scales, Borg scales, and Likert scales in normal subjects during submaximal exercise. *Chest.* 1999; 116:1208–1217. [PubMed: 10559077]
45. 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; 65:1325–1329. [PubMed: 16540546]
46. Hsu V, Shinnar M, Moreyra A, Shindler D, Wilson A, Wilson J, et al. Assessment of pulmonary hypertension in systemic sclerosis: a comparison of doppler echocardiography, cardiac MRI and right heart catheterization. *Arthritis Rheum.* 2004; 50:S424.



**Figure 1.** Process of choosing core set items. OMERACT, Outcome Measures in Rheumatology Clinical Trials; SC, Steering Committee; SCTC, Scleroderma Clinical Trials Consortium; SSc, scleroderma (systemic sclerosis).

**Table 1**

Core set items selected for 11 domains

		Agreement from round 3 Median (25th–75th percentile)
Skin:		
1	Modified Rodnan skin score (range 0–51)	9 (9–9)
2	Visual analogue scale (VAS)/Likert score of patient's global assessment for skin activity*	7 (5.5–9)
3	VAS/Likert score of doctor's global assessment for skin activity*	7 (6–9)
4	Durometer	7 (6–8)
Musculoskeletal:		
1	Tender joint count	6 (5–8)
2	Tendon friction rubs assessed by the doctor*	8 (7–9)
3	Serum creatinine phosphokinase, aldolase	7 (6–9)
Cardiac:		
1	(1)Cardiac echocardiogram with Doppler <sup>‡</sup>	9 (8–9)
2	(2)Right heart catheterisation <sup>‡</sup>	8 (6–9)
3	6-min walk test <sup>‡</sup>	8 (7–9)
4	Borg dyspnoea instrument <sup>‡</sup>	7 (6–9)
Pulmonary:		
1	(3)Pulmonary function testing	9 (9–9)
2	(A)Validated measure of dyspnoea	8 (7–9)
3	Breathing VAS from the Scleroderma Health Assessment Questionnaire (S-HAQ) <sup>23</sup>	7 (6–9)
4	High resolution computer tomography (HRCT) of the lungs: quantifiable scale <sup>‡</sup>	9 (7–9)
Renal:		
1	Calculated creatinine clearance based on serum creatinine (Cockcroft–Gault or Modification of Diet in Renal Disease (MDRD) formula)	9 (7–9)
2	Pre-defined renal crisis (presence or absence)	9 (8–9)
Gastrointestinal:		
1	Body mass index (BMI)	7 (6–9)
4	Validated gastrointestinal (GI) tract VAS scale (part of S-HAQ) <sup>23</sup> or other SSc-validated GI questionnaire	8 (7–9)
Health-related quality of life and function:		
1	Health Assessment Questionnaire-Disability Index (HAQ-DI) <sup>24</sup>	9 (7–9)
2	VAS pain from the HAQ-DI <sup>24</sup>	7 (6–9)
3	Short form-36 (SF-36) version 2 <sup>25</sup>	8 (7–9)
Global health:		
1	VAS/Likert patient global severity	8 (6.5–9)
2	VAS/Likert doctor global severity	8 (6–9)
3	(B)Scleroderma-related health transition by patient	7 (6–9)

		<b>Agreement from round 3 Median (25th–75th percentile)</b>
4	(C)Scleroderma-related health transition by doctor <sup>¶</sup>	7 (5–8)
Raynaud phenomenon:		
1	Raynaud condition score <sup>26</sup>	8 (7–9)
2	VAS Raynaud (part of S-HAQ) <sup>23</sup>	8 (6.5–9)
Digital ulcers:		
1	Active digital tip ulcer count on the volar surface	9 (7–9)
2	VAS digital ulcer (part of S-HAQ) <sup>23</sup>	8 (6–9)
Biomarkers:		
1	Acute phase reactant(s): erythrocyte sedimentation rate (ESR) and/or C-reactive protein (CRP)	8 (6–9)

Parameters to be measured for the items in the core set:

1. Echocardiogram with Doppler: measure pulmonary artery systolic pressure, left and right ventricular systolic and diastolic parameters, pericardial effusion, and chamber size.
2. Right heart catheterisation: pulmonary artery pressure, pulmonary vascular resistance, wedge pressure, cardiac output/cardiac index.
3. Pulmonary function test: forced vital capacity (FVC), diffusion capacity (DLCO), total lung capacity (TLC).
  - A. Validated dyspnoea questionnaires, eg, University of California San Diego (UCSD) Dyspnoea Questionnaire, St. George Respiratory Questionnaire, Mahler Dyspnoea Index.
  - B. Scleroderma-related health transition by patient; compared to 1 year ago, how do you rate your overall scleroderma: much better, a little better, no change, a little worse, much worse?
  - C. Scleroderma-related health transition by doctor; compared to 1 year ago, how do you rate your patient's overall scleroderma: much better, a little better, no change, a little worse, much worse?

\* Items were based on Steering Committee and Scleroderma Clinical Trials Consortium (SCTC) consensus despite lack of full validation.

<sup>†</sup> Standardised central reading mechanism strongly encouraged,

<sup>‡</sup> if relevant to the study.

**Table 2**

## Proposed items for future research

	<b>Research agenda</b>	<b>Median (IQR)</b>
Skin:		
1	Measure of telangiectasia	5 (3–7)
2	Visual analogue scale (VAS) or/Likert patient global assessment for pruritus	7 (5–8)
Musculoskeletal:		
1	Large joint contracture	7 (4–8)
2	Michigan Hand Questionnaire <sup>18</sup>	7 (6–8)
Cardiac:		
1	Serum pro-brain natriuretic peptide (Pro-BNP) or NT-Pro-BNP	7 (6–8)
2	Non-invasive measures of cardiac function, eg, cardiac magnetic resonance imaging, tissue Doppler	7 (6–8)
Gastrointestinal:		
1	Scleroderma gastrointestinal 1.0 questionnaire <sup>3</sup>	8 (7–9)
2	Gastric emptying time and/or 24-h small bowel transit time	7 (6–8)
Health-related quality of life and function:		
1	Measure of health utility eg, Short Form-6D (SF-6D), EuroQol, Quality of Well Being Scale, time trade-off, standard gamble	7 (7–9)
2	Measure of fatigue, eg, functional assessment of chronic illness therapy (FACIT)-fatigue	7 (5–8)
3	Measure of depression/anxiety, eg, Beck Depression Inventory, Center for Epidemiologic Depression Scale	6 (5–8)
Global health:		
1	Medsker Severity Index <sup>29</sup> Digital ulcers	7 (5–8)
1	Development of digital ulcer condition score that captures activity, severity, and impact	7 (6.5–9)
Biomarkers:		
1	Markers of the collagen breakdown, eg, soluble interleukin (IL)2 receptor levels, procollagen I and III aminopropeptide, connective tissue growth factor (CTGF) levels, serum collagen I carboxyterminal telopeptide, urinary pyridinoline cross-link compounds of collagen	7 (5–8)

IQR, interquartile range.