Arthritis Care & Research
Vol. 64, No. 3, March 2012, pp 358–367
DOI 10.1002/acr.20684

ORIGINAL ARTICLE

Validation of Potential Classification Criteria for Systemic Sclerosis

SINDHU R. JOHNSON,[1] JAAP FRANSEN,[2] DINESH KHANNA,[3] MURRAY BARON,[4] FRANK VAN DEN HOOGEN,[5] THOMAS A. MEDSGER JR.,[6] CHRISTINE A. PESCHKEN,[7] PATRICIA E. CARREIRA,[8] GABRIELA RIEMEKASTEN,[9] ALAN TYNDALL,[10] MARCO MATUCCI-CERINIC,[11] AND JANET E. POPE[12]

Objective. **Classification criteria for systemic sclerosis (SSc; scleroderma) are being updated jointly by the American College of Rheumatology and European League Against Rheumatism. Potential items for classification were reduced to 23 using Delphi and nominal group techniques. We evaluated the face, discriminant, and construct validity of the items to be further studied as potential criteria.**

Methods. **Face validity was evaluated using the frequency of items in patients sampled from the Canadian Scleroderma Research Group, 1000 Faces of Lupus, and the Pittsburgh, Toronto, Madrid, and Berlin connective tissue disease (CTD) databases. Patients with SSc (n = 783) were compared to 1,071 patients with diseases similar to SSc (mimickers): systemic lupus erythematosus (n = 499), myositis (n = 171), Sjögren's syndrome (n = 95), Raynaud's phenomenon (RP; n = 228), mixed CTD (n = 29), and idiopathic pulmonary arterial hypertension (PAH; n = 49). Discriminant validity was evaluated using odds ratios (ORs). For construct validity, empirical ranking was compared to expert ranking.**

Results. **Compared to mimickers, patients with SSc were more likely to have skin thickening (OR 427); telangiectasias (OR 91); anti–RNA polymerase III antibody (OR 75); puffy fingers (OR 35); finger flexion contractures (OR 29); tendon/bursal friction rubs (OR 27); anti–topoisomerase I antibody (OR 25); RP (OR 24); fingertip ulcers/pitting scars (OR 19); anticentromere antibody (OR 14); abnormal nailfold capillaries (OR 10); gastroesophageal reflux disease symptoms (OR 8); antinuclear antibody, calcinosis, dysphagia, and esophageal dilation (all OR 6); interstitial lung disease/pulmonary fibrosis (OR 5); and anti–PM-Scl antibody (OR 2). Reduced carbon monoxide diffusing capacity, PAH, and reduced forced vital capacity had ORs of <2. Renal crisis and digital pulp loss/acroosteolysis did not occur in SSc mimickers (OR not estimated). Empirical and expert ranking were correlated (Spearman's $\rho = 0.53$, $P = 0.01$).**

Conclusion. **The candidate items have good face, discriminant, and construct validity. Further item reduction will be evaluated in prospective SSc and mimicker cases.**

INTRODUCTION

Systemic sclerosis (SSc; scleroderma) is a heterogeneous disease or possibly a family of closely related diseases characterized by vasculopathy, immune activation, and fibrosis. Its clinical manifestations vary across individuals, resulting in differences in organ system involvement, treatment regimens, and prognosis. In the absence of a

Supported by the American College of Rheumatology and the European League Against Rheumatism. Dr. Johnson's work was supported by a Canadian Institutes of Health Research Clinician Scientist Award and the Norton-Evans Fund for Scleroderma Research. Dr. Khanna's work was supported by the Scleroderma Foundation (New Investigator Award) and an NIH award (National Institute of Arthritis and Musculoskeletal and Skin Diseases K23-AR053858-05).

[1]Sindhu R. Johnson, MD, FRCPC: Toronto Western Hospital, Mount Sinai Hospital, and University of Toronto, Toronto, Ontario, Canada; [2]Jaap Fransen, PhD: Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [3]Dinesh Khanna, MD, MS: University of Michigan Scleroderma Program, Ann Arbor; [4]Murray Baron, MD: Jewish General Hospital, McGill University, Montreal, Quebec, Canada; [5]Frank van den Hoogen, MD, PhD: Sint Maartenskliniek, Nijmegen, The Netherlands; [6]Thomas A. Medsger, Jr., MD: University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; [7]Christine A. Peschken, MD, MSc: University of Manitoba, Winnipeg, Manitoba, Canada; [8]Patricia E. Carreira, MD: Hospital Universitario 12 de Octubre, Madrid, Spain; [9]Gabriela Riemekasten, MD: German Rheumatology Research Center, Leibniz Institute, Berlin, Germany; [10]Alan Tyndall, MD: University of Basel, Basel, Switzerland; [11]Marco Matucci-Cerinic, MD, PhD: University of Florence, Florence, Italy; [12]Janet E. Pope, MD, MPH, FRCPC: St. Joseph's Health Care, University of Western Ontario, London, Ontario, Canada.

Dr. Tyndall has served on the advisory boards for Novartis, Actelion, Roche, and MSD and received a fee (less than $10,000 each). Dr. Matucci-Cerinic has received consultant fees, speaking fees, and/or honoraria (less than $10,000) from Actelion, Pfizer, GSK, AstraZeneca, and BMS.

Significance & Innovations

- The candidate items for systemic sclerosis classification criteria have good face, discriminant, and construct validity.
- This study, a joint collaboration between the American College of Rheumatology and the European League Against Rheumatism, involved a large number of connective tissue disease patients that were recruited from multiple sites in North America and Europe.
- The results justify proceeding with the next phase of criteria development, which is prospective case and control ascertainment.

diagnostic test for SSc, several sets of classification criteria have been developed and used to identify patients with similar features for recruitment into clinical studies (1–4). The use of classification criteria as inclusion criteria for study participation has facilitated comparison of results across studies.

Existing classification criteria for SSc should be updated (5–10). With improved understanding of the disease, the items regarded to be important for SSc have increased (5,11,12). Goetz and Berne were among the first to describe gastrointestinal involvement in SSc, yet they did not incorporate this domain in their criteria (13). It has also been recognized that use of the 1980 American College of Rheumatology (ACR) preliminary criteria (1–3) for recruitment into clinical trials results in the exclusion of up to 20% of patients with either early SSc or the limited cutaneous subtype of SSc (8,9,14). The exclusion of the limited cutaneous SSc patients is likely due to the fact that a disproportionate number of diffuse cutaneous SSc patients were entered into the ACR prospective study. Therefore, the statistical analysis resulted in criteria that identified 100% of diffuse cutaneous SSc patients, but only 80% of limited cutaneous SSc patients. It has been demonstrated recently that the addition of nailfold capillary abnormalities and telangiectasias to the ACR SSc criteria improves their sensitivity (6,8).

Few criteria sets were developed for wide-scale application in classifying patients for clinical research studies (15,16). Most criteria sets were developed for use in the clinic or a study at hand, limiting their generalizability (17,18). Standards for devising classification criteria have evolved since the original criteria sets were proposed (19). The methodologies used to develop previous criteria do not meet current standards (20,21). For example, one previous SSc criteria proposal utilized healthy subjects and rheumatoid arthritis patients as the comparator groups (2). It has been argued that these patients are so different that they can nearly always be differentiated from SSc patients (7). In keeping with the differential diagnosis faced by clinicians in practice, it has been suggested that criteria should be tested against control populations selected because they have SSc-like features (7). Examples include other connective tissue diseases (CTDs), i.e., mixed CTD (MCTD), Sjögren's syndrome, systemic lupus erythematosus (SLE), polymyositis/dermatomyositis, and undifferentiated CTD, other fibrosing syndromes (eosinophilic fasciitis, linear scleroderma, generalized morphea, sclero-myxedema, nephrogenic systemic fibrosis), and Raynaud's disease.

The Subcommittee on Classification and Response Criteria, a subcommittee of the ACR Quality of Care Committee, has published recommendations for the development and validation of new criteria sets based on the current standards of measurement science (19–21) that are complemented by recommendations from the European League Against Rheumatism (EULAR) Standing Committees of Clinical Epidemiology and for International Studies Including Clinical Trials (22,23). Recommendations for modern criteria development include 1) collaboration between clinical experts and clinical epidemiologists in criteria development, 2) evaluation of the psychometric properties of each candidate criterion, and 3) description of the derivation sample (origin of the patients and control subjects) and gold standard (21–23). Ideally, phases of criteria development should have a balance between expert opinion and data-driven methods (21). Yet there should be avoidance of circularity of reasoning (a bias that can occur when the same experts developing the criteria are the ones contributing cases and comparison patients) (19). A joint international collaborative initiative supported by the ACR and EULAR is underway to develop revised classification criteria for SSc where the methodology has considered these issues.

During phase 1 of the development process, potential items for revised SSc classification criteria were generated through 2 independent international consensus exercises performed by the Scleroderma Clinical Trials Consortium and the EULAR Scleroderma Trials and Research group, resulting in a list of 168 potential items (24). A Delphi exercise of 105 international SSc experts reduced the list of potential items to 102 items. The item list was again rated and subjected to a consensus meeting using nominal group technique by a separate group of European and North American SSc experts, further reducing the list to 23 items (24). As recommended, the next phase of SSc criteria development requires evaluation of the psychometric properties of each candidate criterion (21,23). An important psychometric property of criteria is their validity, the degree to which their application corresponds to the truth. In this study we aimed to evaluate the validity of candidate items for revised SSc criteria. In particular, the objective of this study was to evaluate the face, discriminant, and construct validity of the candidate items. This knowledge will inform the subsequent phases of SSc criteria development.

Address correspondence to Sindhu R. Johnson, MD, FRCPC, Division of Rheumatology, Ground Floor, East Wing, Toronto Western Hospital, 399 Bathurst Street, Toronto, Ontario, Canada, M5T 2S8. E-mail: Sindhu.Johnson@uhn.on.ca.

Submitted for publication July 11, 2011; accepted in revised form October 14, 2011.

Table 1. Frequency of positive responses and ORs for candidate items in the CSRG and 1000 Faces of Lupus cohorts*

Criterion	Scleroderma (n = 127)	SLE (n = 127)	OR†
Abnormal nailfold capillary pattern	93/126 (74)	NA	NA
Anticentromere antibody	32/109 (29)	1/126 (0.8)	52
Anti–topoisomerase I antibody	18/103 (17)	1/127 (0.8)	27
Antinuclear antibody	101/109 (93)	125/127 (98)	0.2
Anti–PM-Scl antibody	9/80 (11)	NA	NA
Anti–RNA polymerase III antibody	15/83 (18)	NA	NA
Calcinosis	43/125 (34)	NA	NA
Reduced DLco‡	51/106 (48)	NA	NA
Digital pulp loss or acroosteolysis	55/124 (44)	0/127 (0)	NE
Dysphagia for solids	74/116 (64)	NA	NA
Esophageal dilation	14/125 (11)	NA	NA
Finger flexion contractures	37/127 (29)	NA	NA
Fingertip ulcers or pitting scars	76/126 (60)	2/126 (2)	94
Reduced FVC‡	8/90 (9)	NA	NA
ILD or pulmonary fibrosis§	44/122 (36)	1/125 (0.8)	66
Gastroesophageal reflux disease¶	106/126 (84)	NA	NA
Puffy fingers	65/125 (52)	NA	NA
Pulmonary arterial hypertension#	8/107 (7)	3/124 (2)	3
Raynaud's phenomenon	123/127 (97)	56/127 (44)	39
Renal crisis	6/126 (5)	NA	NA
Scleroderma skin changes	118/124 (95)	NA	NA
Telangiectasias	90/119 (76)	0/127 (0)	NE
Tendon or bursal friction rubs	18/125 (14)	NA	NA

* Values are the number/total (percentage) unless otherwise indicated. OR = odds ratio; CSRG = Canadian Scleroderma Research Group; SLE = systemic lupus erythematosus; NA = not available; DLco = carbon monoxide diffusing capacity; NE = not estimated; FVC = forced vital capacity; ILD = interstitial lung disease.
† Can be read as systemic sclerosis patients have OR times the odds of having candidate criteria than a mimicker patient.
‡ <70% predicted.
§ ILD (CSRG data) was considered present if a high-resolution computed tomography (HRCT) lung was interpreted by an experienced radiologist as showing ILD or, in the case where no HRCT was performed, if either a chest radiograph was reported as showing either increased interstitial markings (not thought to be due to congestive heart failure) or fibrosis and/or if a study physician reported findings indicative of ILD on physical examination.
¶ Gastroesophageal reflux disease was defined as the patient having reported a history of heartburn, regurgitation of acid, and/or nocturnal choking, and/or ever taking gastroprotective agents.
# Pulmonary hypertension was defined as an estimated pulmonary artery systolic pressure of >45 mm Hg (CSRG data).

SUBJECTS AND METHODS

SSc patients and comparison subjects. SSc patients were identified from established longitudinal cohorts that were not developed for the purpose of this study. Item definitions were often cohort specific and not necessarily identical between cohorts. Representatives of cohorts were invited to participate in this study based on geographic representation (North America and Europe), size, use of standardized data collection in both SSc and comparison patients, and willingness to participate. Comparison patients represented a spectrum of rheumatic and nonrheumatic diseases that share clinical manifestations with SSc. Patients with SSc that overlapped with another rheumatic disease were not included. In all cohorts, the diagnoses were based on the local center's physician(s) judgment. Only a subset of each cohort was sampled (10% randomly selected from each database with the exception of the Pittsburgh cohort, which was sampled by year) for this study, leaving the remainder available for future validation studies.

The Canadian Scleroderma Research Group (CSRG) database patients were compared with the 1000 Faces of Lupus database patients. Both are Canadian multicenter cohorts that recruit patients from academic and community settings. The University of Pittsburgh Connective Tissue Disease Database, the Toronto Scleroderma Database and the Toronto Pulmonary Hypertension in the Connective Tissue Diseases Database, the Madrid Scleroderma cohort, and the Berlin Scleroderma cohort are single-center, academic hospital–based cohorts. SSc patients were compared to the patients who did not have SSc but had a disease similar to SSc (non-SSc comparisons) within each database. In the case of the CSRG patients and the 1000 Faces of Lupus patients, the items of interest were compared between the databases, where available.

Candidate items. The 23 candidate items were: anti–topoisomerase I antibody; scleroderma (skin thickening on examination); abnormal nailfold capillary pattern; anticentromere antibody or centromere pattern on antinuclear antibody (ANA) test; anti–RNA polymerase III antibody; fingertip and/or periungal ulcers or pitting scars; Raynaud's phenomenon; interstitial lung disease or pulmo-

Table 2. Frequency of positive responses and ORs for candidate items in the University of Pittsburgh Connective Tissue Disease cohort*

Criterion	Scleroderma (n = 326)	Non-SSc comparisons					
		SLE (n = 113)	PM/DM (n = 118)	SS (n = 95)	Raynaud's disease	Combined non-SSc comparisons	OR†
Abnormal nailfold capillary pattern	18/26 (69)	NA	21/42 (50)	3/5 (60)	20/36 (56)	44/83 (53)	2
Anticentromere antibody	95/313 (30)	2/110 (2)	1/82 (1)	2/51 (4)	11/84 (13)	16/327 (5)	8
Anti–topoisomerase I antibody	63/313 (20)	NA	0/82 (0)	0/51 (0)	4/84 (5)	4/217 (2)	18
Antinuclear antibody	298/313 (95)	64/72 (89)	60/82 (74)	35/51 (69)	63/84 (75)	222/289 (77)	6
Anti–PM-Scl antibody	9/313 (3)	NA	2/82 (2)	0/51 (0)	1/84 (1)	3/217 (2)	2
Anti–RNA polymerase III antibody	81/313 (26)	NA	1/82 (1)	0/51 (0)	0/84 (0)	1/217 (0.5)	75
Calcinosis	35/241 (15)	NA	6/75 (8)	1/15 (7)	0/36 (0)	7/126 (6)	3
Reduced DLco‡	118/190 (62)	2/7 (29)	44/59 (75)	2/11 (18)	6/14 (43)	54/91 (59)	1
Digital pulp loss or acroosteolysis§	12/125 (10)	NA	0/18 (0)	0/11 (0)	0/9 (0)	0/38 (0)	NC
Dysphagia for solids	139/325 (43)	6/106 (6)	23/117 (20)	15/95 (16)	14/93 (15)	58/411 (14)	10
Esophageal dilation¶	106/163 (65)	NA	5/19 (26)	6/25 (24)	6/23 (26)	17/67 (25)	5
Finger flexion contracture	203/324 (63)	8/112 (7)	5/117 (4)	7/95 (7)	3/92 (3)	23/416 (6)	29
Fingertip ulcers or pitting scars	149/324 (46)	4/107 (4)	2/116 (2)	1/95 (1)	12/92 (13)	19/410 (5)	18
Reduced FVC‡	61/204 (30)	NA	25/62 (40)	0/12 (0)	3/15 (20)	28/89 (31)	0.9
ILD or pulmonary fibrosis#	106/259 (41)	1/50 (2)	40/71 (56)	4/46 (9)	2/30 (7)	47/197 (24)	2
Gastroesophageal reflux disease**	234/325 (72)	28/106 (26)	39/117 (33)	30/95 (32)	37/93 (40)	134/411 (33)	16
Puffy fingers	285/325 (88)	14/111 (13)	12/117 (10)	8/95 (8)	13/92 (14)	47/415 (11)	56
Pulmonary arterial hypertension††	28/326 (9)	3/113 (3)	2/118 (2)	0/95 (0)	1/93 (1)	6/419 (1)	6
Raynaud's phenomenon	318/326 (98)	67/113 (59)	44/117 (38)	42/95 (44)	93/93 (100)	246/418 (59)	47
Renal crisis	22/326 (7)	0/113 (0)	0/118 (0)	0/95 (0)	0/93 (0)	0/419 (0)	NA
Scleroderma	314/326 (96)	1/93 (1)	7/109 (6)	0/92 (0)	0/92 (0)	8/386 (2)	1,190
Telangiectasias‡‡	180/325 (55)	2/107 (2)	1/115 (1)	8/95 (8)	2/92 (2)	13/409 (3)	38
Tendon or bursal friction rubs§§	95/323 (29)	0/75 (0)	0/111 (0)	1/91 (1)	0/90 (0)	1/367 (0.3)	153

* Values are the number/total (percentage) unless otherwise indicated. OR = odds ratio; SSc = systemic sclerosis (scleroderma); SLE = systemic lupus erythematosus; PM/DM = polymyositis/dermatomyositis; SS = Sjögren's syndrome; NA = not available; DLco = carbon monoxide diffusing capacity; NC = not calculable; FVC = forced vital capacity; ILD = interstitial lung disease.
† Can be read as SSc patients have OR times the odds of having candidate criteria than a mimicker patient.
‡ <70% predicted.
§ Acroosteolysis on physical examination or radiographically.
¶ Esophageal dysmotility by barium swallow or manometry.
# Pulmonary fibrosis radiographically.
** Heartburn by history.
†† Pulmonary arterial hypertension by clinical features (physical examination, echocardiogram) or right heart catheterization.
‡‡ Telangiectasias at any site that is believed to be due to connective tissue disease finger contractures recorded by the examining physician on physical examination.
§§ One or more rubs, including the following sites: shoulders, olecranon bursae, wrists (flexor or extensors), fingers (flexor or extensor), knees, and ankles (Achilles, peroneal, posterior tibial, or anterior tibial tendons).

nary fibrosis; renal crisis; reduced carbon monoxide diffusing capacity (DLco); reduced forced vital capacity (FVC); dysphagia for solid food by history; esophageal dilation on radiograph, barium swallow, or high-resolution computerized tomography; telangiectasias; finger flexion contractures; ANA; anti–PM-Scl antibody; pulmonary arterial hypertension; puffy fingers; digital pulp loss or acroosteolysis; persistent recurrent gastroesophageal reflux disease by history; calcinosis; and tendon or bursal friction rubs (24). All of the items were defined using the local research protocols and harmonized across the databases, where possible. For example, DLco and FVC abnormalities were defined as either <70% or <80% predicted, depending on the cohort. The same definitions were applied for within-group comparisons. These definitions can be found in the footnotes accompanying each table. Serologies were identified based on local laboratory assays. The response for each candidate item was dichotomized as present or absent.

Validity. Face validity is present if the items measure what they purport to measure (25,26). Typically, this is assessed using expert judgment, but should be complemented by data-driven methods (26). Face validity was evaluated using the occurrence of positive responses to each item in patients with SSc. For items with a dichotomous response, this is the proportion of patients who gave a positive response (having the item in question) (26). It is suggested that items with positive rates of <20% may be eliminated (26). When the majority of patients do not have the item, very little is gained by retaining the item in a criteria set. The item may not improve the psychometric properties of the criteria set and may actually detract from it by making it longer (26). However, a low frequency item may still be retained if it confers other beneficial properties. For example, a low frequency item may differentiate SSc patients from mimicking conditions very well.

Discriminative validity of each item was evaluated using SSc patients and patients with a disease similar to SSc

Table 3. Frequency of positive responses and ORs for candidate items in the Toronto cohorts*

Criterion	Scleroderma (n = 86)	Non-SSc comparisons			Combined non-SSc comparisons (n = 114), %	OR†
		SLE (n = 36)	MCTD (n = 29)	IPAH (n = 49)		
Abnormal nailfold capillary pattern‡	31 (36)	0 (0)	4 (14)	0 (0)	4	16
Anticentromere antibody	14 (16)	0 (0)	4 (14)	0 (0)	4	5
Anti–topoisomerase I antibody	15 (17)	1 (3)	0 (0)	0 (0)	0.9	24
Antinuclear antibody	NA	NA	NA	NA	NA	NA
Anti–PM-Scl antibody	NA	NA	NA	NA	NA	NA
Anti–RNA polymerase III antibody	NA	NA	NA	NA	NA	NA
Calcinosis	23 (27)	0 (0)	3 (10)	0 (0)	3	14
Reduced DLco§	30 (35)	13 (36)	15 (52)	17 (35)	39	0.8
Digital pulp loss or acroosteolysis	NA	NA	NA	NA	NA	NA
Dysphagia or gastroesophageal reflux disease	71 (83)	4 (11)	20 (69)	0 (0)	21	18
Esophageal dilation	NA	NA	NA	NA	NA	NA
Finger flexion contractures	NA	NA	NA	NA	NA	NA
Fingertip ulcers or pitting scars	28 (33)	1 (3)	7 (24)	0 (0)	7	6
Reduced FVC§	11 (13)	5 (14)	8 (28)	4 (8)	15	0.8
ILD or pulmonary fibrosis	32 (37)	6 (17)	15 (52)	1 (2)	19	3
Gastroesophageal reflux disease	NA	NA	NA	NA	NA	NA
Puffy fingers	NA	NA	NA	NA	NA	NA
Pulmonary arterial hypertension¶	27 (31)	36 (100)	22 (76)	49 (100)	94	0.03
Raynaud's phenomenon	83 (97)	15 (42)	26 (90)	3 (6)	39	44
Renal crisis	6 (7)	0 (0)	0 (0)	0 (0)	0	NE
Scleroderma	83 (97)	0 (0)	15 (52)	0 (0)	13	183
Telangiectasias	71 (83)	0 (0)	11 (38)	0 (0)	10	44
Tendon or bursal friction rubs	NA	NA	NA	NA	NA	NA

* Values are the number (percentage) unless otherwise indicated. OR = odds ratio; SSc = systemic sclerosis (scleroderma); SLE = systemic lupus erythematosus; MCTD = mixed connective tissue disease; IPAH = idiopathic pulmonary arterial hypertension; NA = not available; DLco = carbon monoxide diffusing capacity; FVC = forced vital capacity; ILD = interstitial lung disease; NE = not estimated.
† Can be read as SSc patients have OR times the odds of having candidate criteria than a mimicker patient.
‡ Abnormal nailfold capillaries with enlargement or dropout by visual inspection or ophthalmoscope.
§ <70% predicted.
¶ Mimicker patients come from the pulmonary hypertension database, resulting in a high frequency of pulmonary hypertension within 1 cohort. The results shown are therefore conservative.

(non-SSc comparison patients) from the same center. Using the positive rates, the odds ratio (OR) for each item was calculated for each cohort separately and aggregated into a pooled OR. The candidate items were ranked from highest to lowest based on the pooled OR. It has been recommended that items with an OR <2 be eliminated (27). In the setting of classification, ORs >2 provide better accuracy (27).

Construct validity evaluates the relationship of the item to other measures that are believed to be part of the same phenomenon or "construct" (28). In this study, construct validity was assessed using the strength of association between the empirical ranking based on the pooled OR and the ranking based on expert judgment from a previous Delphi exercise (24).

Statistical analysis. Summary statistics were used to describe the data. ORs were calculated to analyze the association between each item with case or comparison status for each cohort separately. Bayesian statistics were used to calculate the pooled mean OR and 95% credible interval (95% CrI). This approach was taken because it provides the reader the interval for which there is a 95% probability that the true OR falls within (29). The Bayesian analyses used an uninformative normal prior distribution with a mean of 0 and a variance of 10,000 and Markov chain Monte Carlo to sample from the posterior distribution of the items. Starting at 3 randomly generated initial values, the chains were run for a 5,000 iteration "burn-in" period where the chain moved from the starting value toward the correct posterior distribution. The Brooks-Gelman-Rubin statistic was used to verify convergence at this point, i.e., that all 3 chains were sampling from the same distribution. Then 10,000 new sampled values were collected and used to estimate the properties of the posterior distribution, i.e., the OR and 95% CrI. Reporting of the analysis and results is in accordance with the ROBUST criteria (30). The code for analyses is available from the corresponding author upon request. The strength of association between the empirical ranking based on the pooled OR and the ranking based on expert judgment was analyzed using the Spearman's rho rank correlation coefficient. Given variation between experts in their rankings and variation in measurement of criteria across cohorts, we hypothesized a priori that a "moderate" correlation ($\rho = 0.4-0.6$) between the 2 rankings would be significant. Analyses were performed using R (version 2.2.1, The R Foundation for Statistical Computing) and WinBUGS (version 1.4.3, Imperial College and Medical Research Council).

Table 4. Frequency of positive responses and ORs for candidate items in the Madrid cohort*

Criterion	Scleroderma (n = 175)	Non-SSc comparisons: SLE (n = 223)	Non-SSc comparisons: Myositis (n = 53)†	Non-SSc comparisons: Raynaud's disease (n = 135)	Combined non-SSc comparisons (n = 411), no./total	OR
Abnormal nailfold capillary pattern	113/137 (83)	NA	NA	3/135 (2)	3/135	NA
Anticentromere antibody	45/167 (27)	3/203 (2)	1/53 (2)	1/135 (1)	5/391	28
Anti–topoisomerase I antibody	59/167 (35)	2/103 (2)	1/53 (2)	0/135 (0)	3/291	24
Antinuclear antibody	158/169 (94)	197/203 (97)	25/49 (51)	18/135 (13)	240/387	9
Anti–PM-Scl antibody	NA	NA	NA	NA	NA	NA
Anti–RNA polymerase III antibody	NA	NA	NA	NA	NA	NA
Calcinosis	10/56 (18)	NA	5/53 (9)	0/135 (0)	5/188	8
Reduced DLco‡	55/93 (59)	NA	12/38 (32)	NA	12/38	3
Digital pulp loss or acroosteolysis	21/48 (44)	NA	NA	NA	NA	NA
Dysphagia for solids	NA	NA	17/53 (32)	NA	17/53	NA
Esophageal dilation	NA	NA	NA	NA	NA	NA
Finger flexion contractures	23/123 (19)	NA	NA	NA	NA	NA
Fingertip ulcers or pitting scars	77/173 (45)	NA	6/53 (11)	3/135 (2)	9/188	16
Reduced FVC‡	24/95 (25)	NA	20/44 (46)	NA	20/44	0.4
ILD or pulmonary fibrosis	49/173 (28)	7/54 (13)	18/48 (38)	1/135 (1)	26/237	3
Gastroesophageal reflux disease	115/174 (66)	NA	16/52 (31)	5/135 (4)	21/187	15
Puffy fingers	73/124 (59)	NA	NA	9/135 (7)	9/135	20
Pulmonary arterial hypertension	36/170 (21)	3/223 (1)	2/52 (4)	NA	5/275	14
Raynaud's phenomenon	168/174 (97)	30/60 (50)	14/52 (27)	135/135 (100)	179/247	14
Renal crisis	13/174 (8)	0/223 (0)	0/53 (0)	0/135 (0)	0/411	NE
Scleroderma skin changes	154/175 (88)	2/200 (1)	NA	4/135 (3)	6/335	402
Telangiectasias	31/65 (48)	NA	NA	NA	NA	NA
Tendon or bursal friction rubs	NA	NA	NA	NA	NA	NA

* Values are the number/total (percentage) unless otherwise indicated. OR = odds ratio; SSc = systemic sclerosis (scleroderma); SLE = systemic lupus erythematosus; NA = not available; DLco = carbon monoxide diffusing capacity; FVC = forced vital capacity; ILD = interstitial lung disease; NE = not estimated.
† Patients with inflammatory myopathy fulfilling the criteria by Bohan and Peter (38,39), excluding those with overlap myopathy with SSc or SLE.
‡ <70% predicted.

RESULTS

Patients and comparison subjects. Data on 783 SSc patients (CSRG n = 127, Pittsburgh cohort n = 326, Toronto cohort n = 86, Madrid cohort n = 175, Berlin cohort n = 69) and 1,071 comparison subjects were evaluated in this study. The comparison subjects included 499 SLE patients (1000 Faces of Lupus cohort n = 127, Pittsburgh cohort n = 113, Toronto cohort n = 36, Madrid cohort n = 223), 171 inflammatory myositis patients (Pittsburgh cohort n = 118, Madrid cohort n = 53), 95 Sjögren's syndrome patients (Pittsburgh cohort), 228 Raynaud's syndrome patients (Pittsburgh cohort n = 93, Madrid cohort n = 135), 29 MCTD patients (Toronto cohort), and 49 idiopathic pulmonary arterial hypertension patients (Toronto cohort).

Face validity. Rates of positive responses for the candidate items in each SSc cohort are shown in Tables 1–5. The presence of renal crisis and digital pulp loss or acroosteolysis each occurred in <20% of all cohorts, where measured. Anti–topoisomerase I antibody, anti–PM-Scl antibody, calcinosis, reduced FVC, pulmonary arterial hypertension, and finger flexion contractures variably had positive occurrence frequencies <20%, depending on the cohort. The other candidate items were consistently positive in >20% of SSc patients.

Discriminant validity. The ORs for candidate items comparing SSc to non-SSc comparison patients are shown in Tables 1–4. The pooled mean ORs and 95% CrIs for the candidate items are shown in Table 6. Pulmonary arterial hypertension (OR 1.9, 95% CrI 1.4–2.4), reduced DLco (OR 1.5, 95% CrI 1.1–2.0), and reduced FVC (OR 0.9, 95% CrI 0.6–1.3) had ORs <2. Renal crisis and digital pulp loss or acroosteolysis did not occur in any of the non-SSc comparison patients in any of the cohorts, and consequently the ORs were not estimated. If an infinitely small numerical adjustment were added to facilitate estimation, the result would be an infinitely large OR.

Some of the Raynaud's syndrome patients from the Pittsburgh cohort had the presence of ANAs, abnormal nailfold capillaries, and positive serology, suggesting that they may represent pre-SSc or pre–other CTDs. The pooled OR analysis was repeated excluding these patients, and there was no substantial difference in the results.

Construct validity. The empirical-based and expert-based rankings of the candidate criteria are shown in Table

Table 5. Frequency of positive responses for candidate items in the Berlin cohort*

Criterion	Scleroderma (n = 69)
Abnormal nailfold capillary pattern	NA
Anticentromere antibody	19 (28)
Anti–topoisomerase I antibody	15 (22)
Antinuclear antibody	63 (91)
Anti–PM-Scl antibody	3 (4)
Anti–RNA polymerase III antibody	4 (6)
Calcinosis	NA
Reduced DLCO†	53 (77)
Digital pulp loss or acroosteolysis	NA
Dysphagia for solids	47 (68)
Esophageal dilation	NA
Finger flexion contractures	32 (46)
Fingertip ulcers or pitting scars	22 (32)
Reduced FVC†	26 (38)
ILD or pulmonary fibrosis	31 (45)
Gastroesophageal reflux disease	52 (75)
Puffy fingers	NA
Pulmonary arterial hypertension	25 (36)
Raynaud's phenomenon	61 (88)
Renal crisis	3 (4)
Scleroderma	56 (81)
Telangiectasias	NA
Tendon or bursal friction rubs	8 (12)

* Values are the number (percentage). NA = not available; DLCO = carbon monoxide diffusing capacity; FVC = forced vital capacity; ILD = interstitial lung disease.
† <80% predicted.

6. There was a moderate correlation between the 2 rankings with a Spearman's rho of 0.53 ($P = 0.01$) (Figure 1).

DISCUSSION

Evaluation of the validity of candidate SSc items is an important and necessary phase of classification criteria development. Our results demonstrate that the candidate SSc items are valid: they have good face, discriminant, and construct validity. Our study results also provide valuable insights that should be considered in the subsequent phases of criteria development that will include collecting item frequencies on SSc and mimickers at multiple sites in North America and Europe, using programs for item reduction, and then testing the validity of the final criteria in databases.

When there is face validity, motivation, cooperation, and satisfaction among classification criteria users increase (26). The demonstration of face validity requires more than peer judgments; empirical evidence is also required to show that a criterion is measuring what is intended (26). In the case of SSc, this has an important pragmatic implication. A proportion of SSc patients (approximately 20%) who have the disease have been excluded from participation in some clinical trials because they do not meet existing classification criteria. This is a problem when a rare disease is being studied and a significant minority is excluded (6,9). It has been argued that important domains of the disease have been left out of previous criteria (such as antibodies and vascular complications). If the revised SSc classification criteria incorporate items that improve the specificity of the criteria, then more SSc patients can be included in studies from which they may derive a benefit. In this study, the majority of candidate items have excellent face validity, with endorsement frequencies of >20%. Renal crisis, finger pulp loss or acroosteolysis, anti–topoisomerase I antibody, anti–PM-Scl antibody, calcinosis, reduced FVC, pulmonary arterial hypertension, and finger flexion contractures had lower endorsement frequencies. The value of retaining these candidate criteria will need to be carefully evaluated in the next phase of criteria development. It is uncertain if combining uncommon features in revised SSc classification criteria would improve the sensitivity and/or specificity. The utility of criteria with negative responses will also need to be considered. There could be criteria where a negative response makes SSc unlikely (such as absence of ANA or Raynaud's phenomenon). Furthermore, the value of including criteria that rarely occur in SSc patients will need to be balanced by the impact of including too many criteria on the feasibility and reliability of the final criteria set. If a criterion is irrelevant, then users may omit it (26).

The majority of candidate criteria have excellent discriminant validity with high pooled ORs. They effectively discriminate patients with SSc from non-SSc comparison patients included in this study. The utility of a few candidate criteria will require added scrutiny in the next phase of criteria development. Renal crisis and digital pulp loss or acroosteolysis occur uncommonly in SSc. However, they had the strongest discriminating ability because they never occurred in any of the non-SSc comparison patients in any of the cohorts. These criteria are very good at discriminating patients with SSc from patients with other diseases. Criteria that are rare but unique to SSc may be very specific for ruling in the disease, but do not assist in the objective of being more inclusive of those with the disease. Pulmonary arterial hypertension, reduced DLCO, and reduced FVC had ORs <2, indicating a weak discriminating ability. The value of retaining these items will need to be evaluated.

The candidate criteria also had good construct validity. There was good agreement between the empirical-based ranking and the expert-based ranking of the importance of the candidate items. Both methods highlight those criteria that should be considered very important and those that can be considered less important in criteria set development. In this case, the empirical data complement and verify the expert-based data, indicating that criteria development is evolving in the right direction.

This study has a number of strengths. First, our validation study has used large numbers of patients (for an uncommon disease). It has been recommended that a sample size of at least 50 patients be used to evaluate the frequency of an item (26). Other criteria sets have been criticized for using inadequate numbers of patients and controls (21,23). The comparator groups reflect other CTDs and comparisons with nonrheumatic diseases and nonrheumatology settings (21,23). Patients included in this study were recruited from multiple sites in North America and Europe. Previous SSc criteria development did not

Table 6. Pooled ORs and ranking of candidate criteria*

Criterion	Pooled mean OR (95% CrI)	Empirical ranking	Expert-based ranking
Renal crisis	NE	1†	9
Digital pulp loss or acroosteolysis	NE	1†	13‡
Scleroderma skin changes	426.7 (256.5–691.2)	2	1
Telangiectasias	91.4 (57.6–154.5)	3	11
Anti–RNA polymerase III antibody	75.4 (13.2–312.6)	4	6
Puffy fingers	34.9 (24.0–49.2)	5	12
Finger flexion contractures	29.0 (17.8–46.2)	6	19
Tendon or bursal friction rubs	26.81 (2.4–91.9)	7	10
Anti–topoisomerase I antibody	24.9 (12.7–48.0)	8	2
Raynaud's phenomenon	24.1 (15.3–37.5)	9	7
Fingertip ulcers or pitting scars	19.3 (12.7–28.8)	10	5
Anticentromere antibody	13.8 (9.0–21.0)	11	3
Abnormal nailfold capillary pattern	10.4 (6.9–15.1)	12	4
Gastroesophageal reflux disease	7.6 (5.9–9.7)	13	17
Antinuclear antibody	6.06 (4.1–8.8)	14	13‡
Calcinosis	6.05 (3.4–10.5)	15	18
Dysphagia	5.7 (4.2–7.7)	16	19
Esophageal dilation	5.6 (2.9–10.2)	17	14
ILD or pulmonary fibrosis	4.5 (3.4–5.8)	18	8
Anti–PM-Scl antibody	2.4 (1.9–7.1)	19	20‡
Pulmonary arterial hypertension	1.9 (1.4–2.4)	20	15
Reduced DLco	1.5 (1.1–2.0)	21	16
Reduced FVC	0.9 (0.6–1.3)	22	20‡

* OR = odds ratio; 95% CrI = 95% credible interval; NE = not estimated; ILD = interstitial lung disease; DLco = carbon monoxide diffusing capacity; FVC = forced vital capacity.
† Tied rankings. Renal crisis, digital pulp loss, and acroosteolysis did not occur in any mimicker patients; therefore, ORs were not estimated. An infinitely small numerical adjustment would result in an infinitely large OR.
‡ Tied rankings.

have such broad geographic representation (14). Experts involved in generating the candidate criteria were different from those supplying patients (with the exception of 1 expert [TAM]), thereby reducing potential bias from circularity of reasoning (21,23).

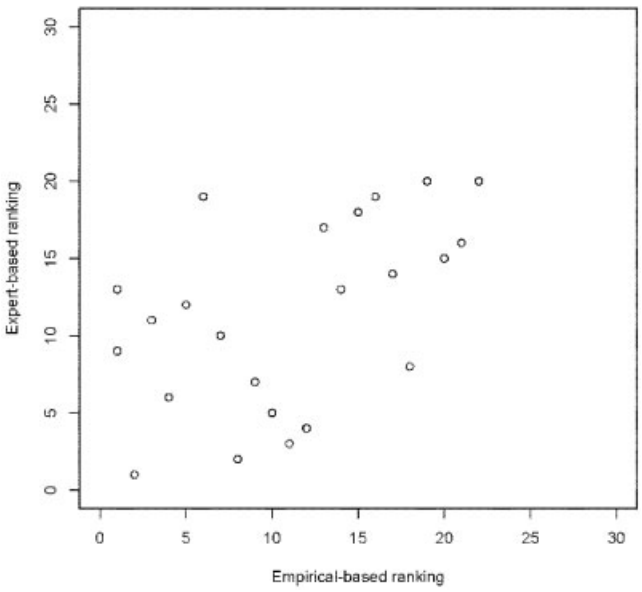

Figure 1. Scatter plot of the empirical-based ranking and the expert-based ranking of the candidate items. The correlation between the 2 rankings was a Spearman's rho of 0.53 ($P = 0.01$).

There are limitations to consider in the interpretation of this study. One limitation to consider is missing data. This is partially related to the fact that data were not collected specifically for this study, but rather had been previously collected for other purposes. As a result, not all sites collect the same variables. To overcome this challenge, we included multiple sites so that there were sufficient data to evaluate each candidate criterion. Not all sites categorized the variable in the same manner in which the candidate criteria have been proposed. This has introduced some variability in comparisons across sites. However, the same definition for each criterion was applied for within-site comparisons. Furthermore, despite the variability in definitions of items across sites, we were able to demonstrate a moderate correlation between the empirical and expert rankings. Many sites were academic medical centers; however, the CSRG and 1000 Faces of Lupus databases enroll patients from both academic and community sites, so the generalizability to other nonacademic practices is likely present (especially due to the fact that the ORs were similar among the various databases used for this study).

The ethnic background of the patients was not evaluated in this study. There may be an overrepresentation of white patients. Given variations in the frequency of specific

criteria (e.g., autoantibodies, lung disease) across ethnic groups, this may affect the external validity of the developed criteria (31). In this study, there is some ethnic variation across the databases we used. The Pittsburgh cohort includes African American patients (32); the 1000 Faces of Lupus database includes Asian, First Nations, and African American patients (33); and the Toronto cohort includes African American and Asian (East Asian and Southeast Asian) patients (30). Subsequent phases of criteria development will need to consider the performance of classification criteria in different ethnic groups.

A potential limitation is that the investigators ascertaining the criteria knew the diagnoses. Criteria were evaluated based on local research protocols or the local standard of care, and this may introduce verification bias. Verification bias occurs when disease status is not determined in all subjects who are evaluated for criteria and when the probability of verification depends on the criteria result and/or other clinical variables. When verification of disease status is more likely among patients with positive criteria, a bias is introduced that can increase the sensitivity of the criteria and reduce its specificity (34). In our study, the majority of SSc patients underwent evaluation of all criteria (e.g., echocardiogram and pulmonary function tests). However, in the case of SSc comparator patients, evaluation of many of the criteria is not routinely done in asymptomatic patients and even in symptomatic patients; performing invasive tests such as right heart catheterization or high-resolution computed tomography thorax scans may not be done on mimickers as often as SSc patients. Subsequent phases of criteria development may need to consider design or analytic techniques to account for verification bias (34). It would not be likely that within a database the investigators did not have a working definition of the disease(s) studied, but the criteria used to make the diagnosis may have been formal criteria or expert opinion. Future prospective data collection that compares patients with SSc and mimickers may reduce this bias when cases are then reanalyzed by experts blinded to the diagnosis.

Our study results provide sufficient fidelity to justify proceeding with the next phase of criteria development, which is prospective case and control ascertainment. During the next phase, the same definitions of items will be applied to all patients and multiple sites will test each item. Given the high discriminating ability of the items using the non-SSc comparisons in this study (e.g., SLE), the next phase of development should include non-SSc comparisons that more closely resemble SSc, such as eosinophilic fasciitis, generalized morphea, and nephrogenic systemic fibrosis. During the next phase, the scaling of the criteria will need to be considered. The criteria could be additive (e.g., SLE classification criteria [35,36]), hierarchical (e.g., 1980 SSc classification criteria [2,3]), or weighted (e.g., rheumatoid arthritis classification criteria [37]).

In conclusion, our study has demonstrated that the candidate SSc items have good face, discriminant, and construct validity. These items should be tested in the next phases of SSc classification development.

ACKNOWLEDGMENTS

The authors thank Dr. Hector Arbillaga, Dr. Mike Oliver Becker, Dr. Sasha Bernatsky, Mr. Ashley Bonner, Dr. Gaelle Chedéville, Dr. Ann Clarke, Dr. Peter Docherty, Dr. Paul R. Fortin, Dr. Marvin Fritzler, Dr. Dafna Gladman, Dr. John Granton, Dr. Tamara Grodzicky, Dr. Carol A. Hitchon, Dr. Adam Huber, Dr. Marie Hudson, Dr. H. Niall Jones, Dr. Elzbieta Kaminska, Dr. Nadir Khalidi, Dr. Peter Lee, Dr. Sophie Ligier, Dr. Janet Markland, Dr. Ariel Masetto, Dr. Jean-Pierre Mathieu, Dr. Ross Petty, Dr. Christian Pineau, Dr. Suzanne Ramsey, Dr. David Robinson, Dr. Earl Silverman, Dr. C. Douglas Smith, Dr. Virginia Steen, Dr. Evelyn Sutton, Dr. J. Carter Thorne, Dr. Lori Tucker, Dr. Murray Urowitz, and Dr. Michel Zummer for their assistance with this study.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Johnson had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Johnson, Khanna, Baron, van den Hoogen, Riemekasten, Tyndall, Matucci-Cerinic, Pope.

Acquisition of data. Johnson, Khanna, Baron, van den Hoogen, Medsger, Peschken, Carreira, Riemekasten, Tyndall, Pope.

Analysis and interpretation of data. Johnson, Fransen, Khanna, Baron, van den Hoogen, Medsger, Riemekasten, Matucci-Cerinic, Pope.

REFERENCES

1. Masi AT, Medsger TA Jr, Rodnan GP, Fries JF, Altman RD, Brown BW, et al. Methods and preliminary results of the Scleroderma Criteria Cooperative Study of the American Rheumatism Association. Clin Rheum Dis 1979;5:27–79.
2. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.
3. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Bull Rheum Dis 1981;31:1–6.
4. Nadashkevich O, Davis P, Fritzler MJ. A proposal of criteria for the classification of systemic sclerosis. Med Sci Monit 2004;10:CR615–21.
5. Walker JG, Pope J, Baron M, Leclercq S, Hudson M, Taillefer S, et al. The development of systemic sclerosis classification criteria. Clin Rheumatol 2007;26:1401–9.
6. Hudson M, Taillefer S, Steele R, Dunne J, Johnson SR, Jones N, et al. Improving the sensitivity of the American College of Rheumatology classification criteria for systemic sclerosis. Clin Exp Rheumatol 2007;25:754–7.
7. Johnson SR, Feldman BM, Hawker GA. Classification criteria for systemic sclerosis subsets. J Rheumatol 2007;34:1855–63.
8. Lonzetti LS, Joyal F, Raynauld JP, Roussin A, Goulet JR, Rich E, et al. Updating the American College of Rheumatology preliminary classification criteria for systemic sclerosis: addition of severe nailfold capillaroscopy abnormalities markedly increases the sensitivity for limited scleroderma [letter]. Arthritis Rheum 2001;44:735–6.
9. Hachulla E, Launay D. Diagnosis and classification of systemic sclerosis. Clin Rev Allergy Immunol 2011;40:78–83.
10. Avouac J, Fransen J, Walker UA, Riccieri V, Smith V, Muller C, et al. Preliminary criteria for the very early diagnosis of systemic sclerosis: results of a Delphi consensus study from EULAR Scleroderma Trials and Research Group. Ann Rheum Dis 2011;70:476–81.

11. Winterbauer RH. Multiple telangiectasia, Raynaud's phenomenon, sclerodactyly, and subcutaneous calcinosis: a syndrome mimicking hereditary hemorrhagic telangiectasia. Bull Johns Hopkins Hosp 1964;114:361–83.
12. Nadashkevich O, Davis P, Fritzler MJ. Revising the classification criteria for systemic sclerosis [letter]. Arthritis Rheum 2006;55:992–3.
13. Goetz R, Berne M. The pathophysiology of progressive systemic sclerosis (generalised scleroderma) with special reference to changes in the viscere. Clin Proc 1945;4:337–92.
14. Wollheim FA. Classification of systemic sclerosis: visions and reality. Rheumatology (Oxford) 2005;44:1212–6.
15. Maricq HR, Valter I. A working classification of scleroderma spectrum disorders: a proposal and the results of testing on a sample of patients. Clin Exp Rheumatol 2004;22 Suppl:S5–13.
16. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202–5.
17. Tuffanelli DL, Winkelmann RK. Diffuse systemic scleroderma: a comparison with acrosclerosis. Ann Intern Med 1962;57:198–203.
18. Ferri C, Valentini G, Cozzi F, Sebastiani M, Michelassi C, La Montagna G, et al. Systemic sclerosis: demographic, clinical, and serologic features and survival in 1,012 Italian patients. Medicine (Baltimore) 2002;81:139–53.
19. Felson DT, Anderson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. Baillieres Clin Rheumatol 1995;9:253–66.
20. Classification and Response Criteria Subcommittee of the American College of Rheumatology Committee on Quality Measures. Development of classification and response criteria for rheumatic diseases [editorial]. Arthritis Rheum 2006;55:348–52.
21. Johnson SR, Goek ON, Singh-Grewal D, Vlad SC, Feldman BM, Felson DT, et al. Classification criteria in rheumatic diseases: a review of methodologic properties. Arthritis Rheum 2007;57:1119–33.
22. Dougados M, Betteridge N, Burmester GR, Euller-Ziegler L, Guillemin F, Hirvonen J, et al. EULAR standardised operating procedures for the elaboration, evaluation, dissemination, and implementation of recommendations endorsed by the EULAR standing committees. Ann Rheum Dis 2004;63:1172–6.
23. Dougados M, Gossec L. Classification criteria for rheumatic diseases: why and how? [editorial]. Arthritis Rheum 2007;57:1112–5.
24. Fransen J, Johnson SR, van den Hoogen F, Baron M, Allanore Y, Carreira PE, et al. Items for developing revised classification criteria in systemic sclerosis: results of a consensus exercise. Arthritis Care Res (Hoboken) 2012;64:351–7.
25. Johnson SR, Hawker GA, Davis AM. The Health Assessment Questionnaire disability index and Scleroderma Health Assessment Questionnaire in scleroderma trials: an evaluation of their measurement properties. Arthritis Rheum 2005;53:256–62.
26. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.
27. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004;159:882–90.
28. Fletcher RH, Fletcher SW. Clinical epidemiology: the essentials. 4th ed. Baltimore: Lippincott Williams and Wilkins; 2005.
29. Spiegelhalter D, Abrams K, Myles J. Bayesian approaches to clinical trials and health-care evaluation. Chichester: John Wiley & Sons; 2004.
30. Low AH, Johnson SR, Lee P. Ethnic influence on disease manifestations and autoantibodies in Chinese-descent patients with systemic sclerosis. J Rheumatol 2009;36:787–93.
31. Kuwana M, Kaburaki J, Arnett FC, Howard RF, Medsger TA Jr, Wright TM. Influence of ethnic background on clinical and serologic features in patients with systemic sclerosis and anti–DNA topoisomerase I antibody. Arthritis Rheum 1999;42:465–74.
32. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972-2002. Ann Rheum Dis 2007;66:940–4.
33. Peschken CA, Katz SJ, Silverman E, Pope JE, Fortin PR, Pineau C, et al. The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort. J Rheumatol 2009;36:1200–8.
34. Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM. Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. N Engl J Med 2003;349:335–42.
35. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
36. Hochberg MC, for the Diagnostic and Therapeutic Criteria Committee of the American College of Rheumatology. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
37. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569–81.
38. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). N Engl J Med 1975;292:344–7.
39. Bohan A, Peter JB. Polymyositis and dermatomyositis (second of two parts). N Engl J Med 1975;292:403–7.