Classification Criteria for Systemic Sclerosis Subsets

SINDHU R. JOHNSON, BRIAN M. FELDMAN, and GILLIAN A. HAWKER

ABSTRACT. ***Objective.*** To evaluate the measurement properties of criteria for systemic sclerosis (SSc) subsets for classification of patients in SSc trials, and to determine if any one criteria set confers measurement advantage over others.

Methods. A systematic review of articles describing classification criteria for SSc subsets was performed. Evidence supporting the sensibility (statement of purpose for which the criteria will be used, population, setting, face and content validity, and feasibility), validity, and reliability of the criteria was evaluated.

Results. Fourteen sets of criteria for SSc subsets were identified. There is variability in the intended purpose and setting for which criteria sets are to be applied. Although face validity improves with the addition of less commonly encountered subsets or disease manifestations as criteria, the feasibility of implementing such criteria is conversely limited. Content validity for most criteria sets has not been evaluated due to lack of an explicitly stated conceptual framework for SSc. The criteria with 3 or more subsets do not provide incremental predictive validity over the 2-subset criteria. Our ability to compare subset criteria on divergent validity and reliability is limited by a lack of data.

Conclusion. The 2-subset criteria of LeRoy, *et al* have good feasibility, acceptable face validity, and good predictive validity. Further research is needed to compare the content validity, divergent validity, and reliability of these with other subset criteria for use in SSc trials. (First Release August 1 2007; J Rheumatol 2007;34:1855–63)

Key Indexing Terms:
SYSTEMIC SCLEROSIS SCLERODERMA CRITERIA VALIDITY RELIABILITY

Systemic sclerosis (SSc) is a heterogeneous disease characterized by fibrosis, vasculopathy, and immune activation. It affects a variety of organ systems and is a condition that lacks a pathognomonic diagnostic test. Classification criteria are an essential component of SSc research as they ensure recruitment of patients with similar features into studies and they allow for comparison of results across studies. There has been increasing recognition of subsets within the spectrum of SSc with a belief that subsets of patients have variable disease expression[1,2], response to therapy[3], morbidity[1], and prognosis[2,4]. Thus, the accurate identification of disease subsets may improve the ability to prognosticate organ involvement and survival[5,6], develop appropriate surveillance programs[5,6], and guide tailored treatment recommendations.

The standards of measurement science have evolved over the time period that descriptions of clinical SSc subsets have been proposed. The Committee on Classification and Response Criteria, a subcommittee of the American College of Rheumatology (ACR) Quality Measures Committee, has been charged with encouraging development and validation of new and improved classification criteria for various rheumatic diseases. Recommendations for development and validation of criteria sets have been developed based on the current standards of measurement[7]. Evaluation of classification criteria designed to differentiate rheumatic diseases from one another has been completed[8]. To supplement the work of the Classification and Response Criteria subcommittee, the objectives of our study were (1) to identify classification criteria for subsets within SSc, and (2) to comparatively evaluate the criteria against current standards of measurement.

MATERIALS AND METHODS

Data sources. Eligible articles were identified using Medline (1966-2005) and Embase (1966-2005). The search strategy was limited to human studies but not limited to English language. The bibliography of eligible studies was searched.

Search terms. The following keywords were used in the search: (systemic sclerosis OR scleroderma) AND (criteria OR criteria development OR classification OR classification criteria OR classification tree OR diagnostic criteria OR diagnostic assessment OR diagnostic index OR disease criteria OR

From the Division of Rheumatology, University Health Network; The Hospital for Sick Children; Women's College Hospital; Departments of Health Policy, Management and Evaluation, Public Health Sciences, and Pediatrics, University of Toronto, Toronto, Ontario, Canada.

Dr. Johnson has been awarded an Abbott Scholar Award in Rheumatology Research and Canadian Arthritis Network Fellowship. Dr. Hawker is a Senior Distinguished Rheumatologist Investigator of The Arthritis Society and is the F. M. Hill Chair in Academic Women's Medicine, University of Toronto. Dr. Feldman holds the Canada Research Chair in Childhood Arthritis.

S.R. Johnson, MD, FRCPC, Division of Rheumatology, University Health Network, The Hospital for Sick Children, and Department of Health Policy, Management and Evaluation; B.M. Feldman, MD, MSc, FRCPC, Division of Rheumatology, The Hospital for Sick Children, and Departments of Health Policy, Management and Evaluation, Public Health Sciences, and Pediatrics; G.A. Hawker, MD, MSc, FRCPC, Division of Rheumatology, Women's College Hospital, and Department of Health Policy, Management and Evaluation, University of Toronto.

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

Accepted for publication May 23, 2007.

disease measures OR disease assessment OR disease index OR validity OR face validity OR content validity OR construct validity).

Screening for relevance. Titles and abstracts were screened to identify articles that addressed criteria or classification of SSc and measurement properties (sensibility, validity, reliability). Articles describing classification of localized scleroderma were excluded[9].

Citation index. Web of Science (v3.0) was used to search the Science Citation Index Expanded (1945-April 2, 2006), Social Sciences Citation Index (1956-April 2, 2006), and Arts and Humanities Index (1975-April 2, 2006) to identify the number of times each article was cited. The citation number was used as a measure of the frequency the criteria set was cited in the published literature. It was used as a proxy to assess the degree classification criteria were used in research.

Evaluation of measurement properties. The criteria sets were reviewed to ascertain if the following properties have been examined.

1. Item generation and reduction. Item generation is the process used to identify potential items for the criteria set[10]. Item reduction is the process used to eliminate inappropriate items for the final criteria set[11]. Articles were evaluated for specification of methods of item generation and reduction.

2. Sensibility. Sensibility evaluates the usefulness of the criteria[12]. Principles used to evaluate sensibility include a statement of purpose, setting, face and content validity, and feasibility. Face validity evaluates if the criteria reflect the attributes of the disease, and if there is biological coherence of the items[12]. Content validity evaluates if the criteria set reflects all the domains in the conceptual framework of the disease[7,13]. Criteria sets were evaluated to determine if the criteria and subset classification reflected our current understanding of SSc. Feasibility refers to the ease of usage of the criteria set[12]. Determinants of feasibility include time required to use the criteria, access to laboratory testing, or specialized clinical skills needed to apply the criteria.

3. Validity. Due to the lack of a gold standard diagnostic test for SSc or SSc subsets, divergent construct validity evaluates the ability of a criteria set to correctly distinguish subsets of patients[14]. Sensitivity and specificity are used as measures of construct validity[15]. Predictive validity assesses the relationship between classification at baseline and a measure administered some time later[15]. In SSc, the ability to predict organ involvement and survival are considered important outcomes.

4. Reliability. Reliability refers to the reproducibility of the measure. Test-retest reliability is evaluated when the criteria are applied to the same group of patients on 2 different occasions[16].

RESULTS

Literature review. The literature search identified 530 citations. Five hundred twelve citations were excluded as they did not describe classification criteria, were review articles, or discussed classification criteria in non-SSc rheumatic diseases. Nineteen articles describing classification criteria in SSc were identified. Five articles described classification criteria used to differentiate SSc from other rheumatic diseases[17-21]. Since the objective of our study was to identify criteria for subsets within SSc, these 5 articles were excluded. This included the ACR criteria for SSc, which have been evaluated elsewhere[8]. In total, 14 criteria sets classifying SSc subsets were identified (Table 1).

Measurement properties of criteria sets. Table 2 summarizes the measurement properties of the criteria sets.

1. Goetz and Berne criteria. These 2-subset criteria, based on the clinical judgment of the authors, was developed to classify a case series of patients in South Africa[22]. The criteria set has limited face validity as it does not reflect current knowledge of SSc. It is easy to apply by all clinicians. The authors were among the first to identify gastrointestinal involvement as a disease manifestation, but this was not incorporated in the criteria. Divergent validity, predictive validity, and reliability have not been tested.

2. Winterbauer criteria. Patients with CRST (calcinosis, Raynaud's phenomenon (RP), sclerodactyly, telangiectasia) were described as a benign subset of SSc[23]. Subsequent investigators have added "E" (esophageal dysmotility). This description was intended for clinical practice and is easy to use. Face validity is limited as the criteria only describe one subset. These criteria have limited content validity as they do not address other important domains. These criteria poorly discriminate between patients with mild, moderate, and extensive skin involvement as all subtypes may have > 4 CREST manifestations[24]. The reliability of the criteria was not specified.

3. LeRoy criteria. This 2-subset criteria set was proposed to improve the nomenclature of SSc, identify patients at risk of visceral complications, and classify homogeneous groups of patients for clinical research[25]. Criteria development was based on the judgment of an expert panel. These criteria have good face validity; they recognize important attributes of the disease. The authors did not specify the construct of SSc on which the criteria were based, thus content validity could not be evaluated. The convergent and divergent validity of the criteria have been demonstrated in several studies. Diffuse SSc is frequently associated with tendon friction rubs, anti-topoisomerase I, and poor prognosis, whereas limited SSc is frequently associated with calcinosis, telangiectasia, anti-kinetochore antibody, and pulmonary hypertension[26]. Human leukocyte antigen (HLA) DR1, DR5, DR6, and Bw35 occur more commonly in patients with limited SSc; however, only HLA-DR1 had statistical significance[27]. Patients classified as having diffuse SSc have Scl-70 antibodies present and show nucleolar pattern on antinuclear antibody (ANA) staining, and have anticentromere antibodies (ACA) less frequently than patients classified as having limited SSc[28]. More recent evaluation of autoantibody profiles suggest that the LeRoy criteria have divergent validity in relation to ACA, as they occur in 89% of patients classified as having limited SSc and 7% of patients classified as having diffuse SSc[29]. Scl-70 antibodies occur in 36% of patients classified as having limited SSc and 60% of patients classified as having diffuse SSc[29]. Although the feasibility of the criteria is limited to clinicians experienced in skin examination, auscultation of friction rubs, and capillaroscopy, it has been successfully applied in multinational, tertiary-care settings[29].

The criteria have good predictive validity for survival, but poorly predict the development of restrictive lung disease. Four studies[26,28,30,31] demonstrated strong predictive validity for survival where patients with limited SSc have better survival than patients with diffuse SSc. However, these criteria poorly predict lung involvement as restrictive lung disease

Table 1. Classification of systemic sclerosis subsets.

Study	Classification Scheme	Number of Citations
Barnett[36]	3 subsets: limited, moderate, extensive, based on skin involvement of the fingers only, limbs and face, and involvement of the trunk, respectively	66
Ferri[30]	4 subsets: sine scleroderma SSc: absence of cutaneous involvement with visceral involvement, NC changes and autoantibodies; limited cutaneous: skin involvement of fingers with or without involvement of neck, face, and axillae; intermediate cutaneous: skin involvement of upper and lower limbs, neck and face without truncal involvement, diffuse cutaneous: distal and truncal skin involvement	52
Giordano[28]	6 subsets: I: sclerodactyly only; II: sclerodactyly and skin involvement of neck, lower eyelid, or axillae; III: skin involvement of hands and forearms ± legs ± face; IV: group III and arm and/or thigh skin involvement; V: group III and thorax; VI: group III and/or IV and/or V plus the abdomen	121
	3 subsets: limited: skin involvement of fingers, face, neck, axillae; intermediate: skin involvement proximal to fingers; diffuse: truncal skin involvement	121
Goetz[22]	2 subsets: acrosclerosis and diffuse: based on skin thickening limited to extremities or includes trunk	227
Holzmann[53]	5 subsets (Types I–IV) based on presence/absence of RP, sclerosis, extracutaneous manifestations, ANA	10
LeRoy[25]	2 subsets: diffuse cutaneous SSc: onset of RP within 1 year; truncal and acral skin involvement; tendon friction rubs; early incidence of ILD, renal failure, diffuse GI disease, myocardial involvement; absence of ACA, abnormal ND; limited cutaneous SSc: RP for years, skin involvement limited to hands, face, feet, forearms or absent; late incidence of PAH, trigeminal neuralgia, calcinosis, telangiectasia; high incidence of ACA, abnormal NC	877
LeRoy and Medsger[41]	4 subsets: limited SSc (LSSc) consists of (1) objective RP plus any one of NC changes or SSc selective autoantibodies OR (2) subjective RP plus both NC changes and SSc selective autoantibodies; limited cutaneous SSc (lcSSc): criteria for LSSc plus distal cutaneous changes; diffuse cutaneous (dcSSc): criteria for lcSSc plus proximal cutaneous changes; diffuse fasciitis with eosinophilia: proximal cutaneous changes without criteria for lSSc or lcSSc	46
Maricq[6]	6 subsets: diffuse, intermediate, digital, scleroderma sine scleroderma, undifferentiated connective tissue disease with scleroderma, CREST syndrome	3
Masi[43]	3 subsets: digital: skin involvement of fingers or toes but not proximal extremity or trunk; proximal extremity: proximal extremities or face but not trunk; truncal: thorax or abdomen	42
Rodnan[2]	3 subsets: classical disease involving skin of the trunk, face and proximal extremities, and early involvement of esophagus, intestine, heart, lung and kidney; CREST syndrome; and overlap syndromes including sclerodermatomyositis and mixed connective tissue disease	79
Scussel-Lonzetti[39]	4 subsets: normal skin, limited: skin involvement restricted to fingers, with RP, calcinosis, esophageal involvement and telangiectasia; intermediate: skin involvement of arms proximal to metacarpophalangeal but not trunk; diffuse: skin involvement of the trunk	1
Tuffanelli and Winkelmann[35]	2 subsets: acrosclerosis: RP, acral skin involvement; diffuse SSc: no RP, skin involvement beginning centrally	42
Winterbauer[23]	CREST syndrome: calcinosis, RP, sclerodactyly, telangiectasia	176

RP: Raynaud's phenomenon; NC: nailfold capillary; ILD: interstitial lung diseases; GI: gastrointestinal; ACA: anticentromere antibodies; PAH: pulmonary arterial hypertension; LSSc: limited SSc.

occurs in 30% of patients with limited SSc and 50% of patients with diffuse SSc ($p = 0.16$)[32].

4. Giordano 3-subset criteria. These 3-subset criteria, based on degree of skin involvement, were proposed for the classification of patients with SSc in the hospital setting[28]. These criteria have face validity, content validity, and feasibility comparable to other 3-subset criteria[25,33]. The criteria lack divergent validity with regard to antibodies[28]. The predictive validity of the criteria is good; patients with limited disease have better survival than those with diffuse disease[28]. Reliability is not specified.

5. Giordano 6-subset criteria. These 6-subset criteria[28], based on degree of skin involvement, were proposed based on a previous iteration of criteria development[34]. The criteria were developed for research in a hospital setting. Face validity is good and comparable to their predecessors. Feasibility is limited to those competent in SSc skin examination. The construct of SSc is not specified, thus content validity cannot be evaluated. The criteria have poor divergent validity with regard to serology (ACA, ANA, anti-Scl70 antibody) and poor predictive validity with regard to survival[28].

6. Tuffanelli and Winkelmann criteria. These 2-subset criteria classify patients as acrosclerosis and diffuse based on extent of skin involvement and presence of RP[35]. The purpose of the criteria was to classify patients in a retrospective cohort study at the Mayo Clinic from 1935 to 1958. Although face, content, and construct validity of these criteria is improved compared to predecessors as they include RP as a differentiating manifestation (reflecting the belief at that time), this differentiating point is no longer valid. Although discriminant validity is not reported, the authors report good predictive validity as patients with diffuse disease have decreased survival[35].

Table 2. Summary of measurement properties in systemic sclerosis subset criteria.

Study	Item Generation and Reduction Methods	Measurement Property: Sensibility: Purpose and Setting	Sensibility: Face Validity	Sensibility: Content Validity	Validity: Feasibility	Validity: Divergent	Validity: Predictive	Validity: Reliability
Barnett[36]	NS	Yes	Yes	NS	Yes	Yes[37]	Yes[33,37,38]	NS
Ferri[30]	NS	Yes	Yes	NS	Yes	No[30]	Yes[30]	NS
Giordano[28]	Yes							
6-subset		Yes	Yes	NS	Yes	No[28]	No[28]	NS
3-subset	Yes	Yes	Yes	NS	Yes	No[28]	Yes[28]	NS
Goetz[22]	NS	Yes	No	No	Yes	No	No	NS
Holzmann[5]	NS	Yes	Yes	NS	Yes	NS	NS	NS
LeRoy[25]	Yes	Yes	Yes	Yes[26]	Yes	NS	Yes[26,28,30,31], No[32]	NS
LeRoy[41]	NS	Yes	Yes	NS	No	NS	NS	NS
Maricq[6]	NS	Yes	Yes	Yes	No[44]	NS	NS	NS
Masi[43]	NS	Yes	Yes	NS	Yes	NS	NS	NS
Rodnan[2]	Yes	NS	Yes	NS	Yes	NS	NS	NS
Scussel-Lonzetti[39]	NS	Yes	Yes	NS	Yes	NS	Yes[39]	NS
Tuffanelli[35]	NS	Yes	No	No	No	NS	Yes	NS
Winterbauer[23]	NS	Yes	Yes	No	Yes	No[24]	NS	NS

NS: not specified.

7. Barnett criteria. This 3-subset criteria set[36], based on degree of skin involvement, was developed to describe subsets of SSc in a clinical setting. The use of extent of skin involvement for thresholds gives it face validity comparable to other criteria sets. However, as the construct of SSc is not specified, one is unable to evaluate content validity. Its feasibility is limited to clinicians competent in SSc skin assessment.

The Barnett criteria have good divergent validity when compared to serology. Zero (0%), 7 (31.8%), and 5 (55.5%) individuals with types I to III, respectively, were Scl-70 antibody-positive. Conversely, among individuals with types I to III, respectively, 1 (10%), 7 (31.8%), and 0 (0%) had antibodies to single-stranded DNA[37]. This criteria set has good predictive validity; patients with type 1 disease have the longest survival, type 2 have intermediate survival, and type 3 have the shortest survival[37]. These criteria have good predictive validity; the 10-year survival is 71% in type 1, 58% in type 2, and 21% in type 3[38].

8. Rodnan criteria. These 3-subset criteria[2] were developed based on the clinical judgment of the authors. Methods of item generation, item reduction, and determinants of sensibility (purpose, setting) are not specified. The criteria have good face validity as they represent clinical knowledge at that time. The criteria have good feasibility as they are easy to apply in the clinical setting and do not require specialized testing or personnel. However, the subsets are not mutually exclusive[26]. Thus, there may be confusion in their application, which may result in misclassification error. Validity (construct, divergent, and predictive) and reliability are not specified.

9. Ferri criteria. This 4-subset criteria set classifies patients based on extent of skin involvement[30]. The purpose of the criteria is to identify subsets for an Italian descriptive and prognostic study. Face validity is good and comparable to its predecessors. Its feasibility is limited to those competent in SSc skin examination. The construct of SSc is not specified, thus content validity cannot be evaluated. The criteria have good predictive validity, with 10-year survival rates in the limited, intermediate, and diffuse subsets of 78.3%, 65.5%, and 52.2%, respectively[30]. Statistically significant differences between limited cutaneous SSc (lcSSc) versus intermediate cutaneous SSc and lcSSc versus diffuse cutaneous SSc (dcSSc) in the frequency of hypermelanosis, sicca syndrome, esophageal involvement, and lung involvement have been demonstrated. However, no significant statistical differences are found in these disease manifestations when the lcSSc subset is compared to the dcSSc subset[30]. The reliability of these criteria is not specified.

10. Scussel-Lonzetti criteria. This 4-subset variation[39] on the criteria of Barnett[38], Giordano[28], and Ferri[40] classifies patients based on extent of skin involvement. Item generation and item reduction is not specified. The purpose of the criteria is subgroup classification for a study of prognosis in a tertiary care academic center. Face validity is good and comparable to its predecessors. Its feasibility is limited to those competent in SSc skin examination. The construct of SSc is not specified, thus content validity cannot be evaluated. The criteria have good predictive validity; the cumulative survival rates in the 4 subsets from normal to diffuse are 90.6%, 79%, 75.9%, and 62.4% at 10 years, respectively[39]. Divergent validity and reliability of the criteria are not specified and require further evaluation.

11. LeRoy and Medsger criteria. These 4-subset criteria were

intended for early diagnosis and classification[41]. Item generation and item reduction are not specified. These criteria have improved face validity over their predecessors as they include advances in knowledge gained over time, i.e., the recognition of RP as a frequent manifestation, abnormalities on capillaroscopy, and SSc-specific serology. The authors do not specify the construct of SSc, thus it is difficult to evaluate content validity. However, the inclusion of "diffuse fasciitis with eosinophilia" as a subset is an improvement in content validity over preceding subset criteria. Feasibility is limited to clinicians with access to objective testing of vascular response to cold (e.g., Nielsen test), capillaroscopy, and laboratories capable of testing for SSc-selective autoantibodies (e.g., anti-fibrillarin, anti-fibrillin). Divergent validity, predictive validity, and reliability of these criteria are not specified.

12. Holzmann criteria. These 5-subset criteria classify patients based on skin involvement, presence of RP, internal organ involvement, and presence of ANA[42]. This set of criteria is intended to be more comprehensive as it includes subsets of patients without skin involvement, localized skin involvement, and/or immune activation, thereby improving face validity. The construct of SSc is not described, thereby limiting the evaluation of content validity. Feasibility is limited to clinicians experienced in SSc skin examination and access to ANA testing. Divergent validity, predictive testing, and reliability are not specified.

13. Masi criteria. This 3-subset criteria set classifies patients based on skin involvement[43]. The intent of this classification is to clarify terminology due to confusion between the 3-subset criteria of Barnett[38] and Giordano[28]. Since the intent was not new criteria development, item generation, item reduction, discriminant validity, predictive validity, and reliability are not specified.

14. Maricq criteria. The most comprehensive criteria set classifies patients into 6 mutually exclusive subsets. Item generation and reduction are not specified. The purpose of the criteria is to develop a comprehensive classification for all scleroderma spectrum disorders for use in clinical research. These criteria have incrementally improved face validity over other criteria sets as they incorporate subsets within the spectrum of SSc (scleroderma sine SSc, undifferentiated connective tissue disease with scleroderma) that have previously been excluded. The feasibility of applying these criteria is a concern; these criteria have been criticized for being "too complicated"[44] and thus feasible for use only by clinicians competent in SSc skin examination and capillaroscopy. Divergent validity, predictive validity, and reliability are not specified.

Comparison of 2-subset criteria to criteria with ≥ 3 subsets. One study comparing 2-subset criteria[25] to 3-subset criteria[30] demonstrated that the 3-subset criteria have incremental predictive validity with regard to survival[30]. However, this finding may have been confounded by significantly different disease duration across the 3 subsets[44]. Three studies failed to demonstrate that the 3-subset criteria have incremental predictive validity over the 2-subset criteria with regard to survival. Jacobsen, *et al*[45] compared survival between 2-subset criteria[25] and 3-subset criteria[38] and found the standardized mortality ratio (SMR) was similar between subsets with digital involvement (SMR = 2.1) and extremity involvement (SMR = 2.3), and compared to patients with limited skin involvement (SMR = 2.3). Patients with truncal involvement (SMR = 4.6) were comparable to patients classified as having diffuse disease (SMR = 4.5)[45]. Giordano, *et al* demonstrated no significant difference in survival between patients classified as having intermediate disease compared to both limited and diffuse disease[28]. Scussel-Lonzetti, *et al* demonstrated no significant difference in survival between limited and intermediate SSc (log-rank test, $p = 0.2$), but did demonstrate a significant difference in survival between patients with limited and diffuse SSc (log-rank test, $p = 0.0005$)[39].

DISCUSSION

We found that the criteria sets developed for SSc subsets do not meet current standards for measurement properties, and thus further research must be done to validate current sets or to develop new, valid criteria. Our study summarizes the measurement properties of classification criteria for subsets of SSc for the purpose of assisting researchers in evaluating their appropriateness for use in SSc trials and highlighting areas in need of further inquiry. Second, our study comparatively evaluates criteria sets to ascertain if criteria with more subsets provide incremental value over criteria with fewer subsets[46].

There is variability in sensibility (purpose, setting, face and content validity, and feasibility) across criteria sets. Few criteria sets were developed for wide-scale application for classifying patients in clinical research[6,25]. Many criteria were developed for use in the clinic or for the study at hand[35,39,40]. The setting in which the criteria were applied included both outpatient clinics and hospital wards, both largely in academic settings. Use of subset criteria as diagnostic criteria requires further research to establish their diagnostic utility. Criteria for diagnostic purposes require high specificity with good sensitivity, whereas criteria for use in epidemiologic studies may only require a balance of sensitivity and specificity. The variability in the thresholds for sensitivity and specificity relates to their intended use[8]. When a patient has been given a diagnosis using criteria with a high specificity, the clinician can be sure the patient has the disease. In epidemiologic studies evaluating incidence and prevalence, overly specific criteria would result in underestimation of the true prevalence, and overly sensitive criteria would result in overestimation of the true prevalence. In this situation, researchers would prefer a balance of sensitivity and specificity[8]. Until these characteristics of criteria have been evaluated, clinicians should be cautious when applying the criteria for diagnostic purposes.

The lack of data on divergent validity — the ability to distinguish mutually exclusive subsets — is a threat to the utility

of all available criteria sets. Previous iterations of criteria validation used healthy participants or patients with other rheumatic diseases [systemic lupus erythematosus (SLE), rheumatoid arthritis] as the control populations[17]. One might argue that after the early stages of disease, most rheumatologists do not have difficulty discriminating between SSc and SLE[5]. Rather, there is greater difficulty discriminating between SSc subsets and other diseases characterized by fibrosis and immune activation. Thus additional research regarding the divergent validity (sensitivity, specificity) of subset criteria is needed, particularly using carefully selected control populations that have SSc-like features.

A critical but poorly documented domain of sensibility is content validity. Content validity evaluates if all the subsets that reflect the relevant domains in the conceptual framework of the disease and the means to identify those subsets have been included. In the case of SSc subset criteria, few investigators have explicitly outlined the construct of SSc on which the subset criteria have been based. Thus, important domains (i.e., disease manifestations) may not have been included. Researchers have increasingly relied on the use of conceptual frameworks to guide their thinking[47]. We propose a conceptual framework for the construct of SSc based on clinical observations and supporting evidence from the literature[48-51] (Figure 1). The framework suggests that SSc comprises 3 domains that may overlap in varying degrees: fibrosis, vasculopathy, and immune activation/inflammation. Through this conceptual framework, SSc subsets can be identified based on the degree of overlap across domains. We do not propose this as a static framework; rather, we present this to stimulate debate and modification as further insights are gained into the immunopathophysiology of the disease.

The addition of subsets to a 2-subset classification improves content validity by reflecting a greater SSc spectrum of disease, but does not improve predictive or divergent validity. The addition of intermediate skin involvement[29,39,40,43], scleroderma sine SSc (ssSSc)[6,28,43], overlap/undifferentiated syndromes[2,6], and eosinophilic fasciitis[41] as SSc subsets has been proposed. Based on our construct of SSc, this is an improvement in content validity. The addition of ssSSc may have important utility in clinical practice. Some patients, early in their disease, may present to clinic without skin involvement but a diagnosis is made based on other clinical and serologic findings. However, the addition of ssSSc and intermediate skin involvement have not been shown to improve divergent or predictive validity[52]. Indeed, the weight of evidence to date does not demonstrate incremental predictive validity of 3-subset criteria over 2-subset criteria with regard to survival[28,45].

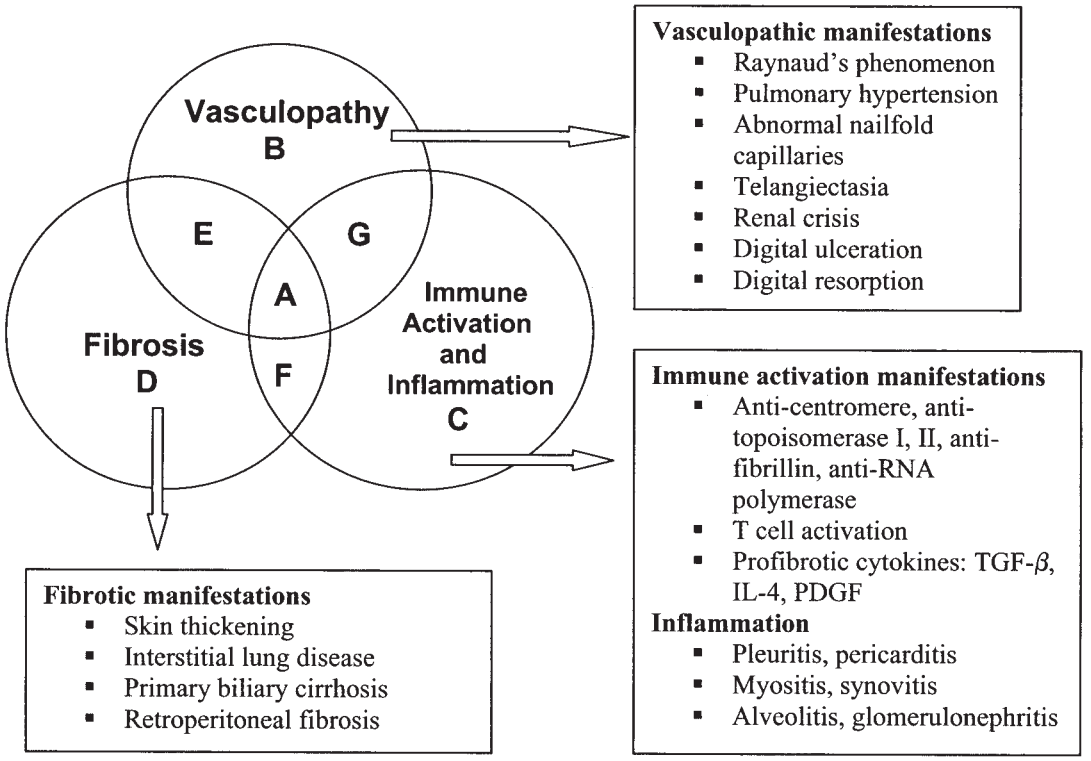
Further, the tradeoff between feasibility and content validity affects the incremental value of one criteria set over another. The feasibility of all criteria is limited to clinicians competent in SSc skin examination. The addition of capillaroscopy, antibodies, and vascular testing as criteria further limits the feasibility of the criteria in general practice, but this is not a hindrance for use in a specialized clinic. Further investigation is needed to ascertain if the tradeoff between feasibility and content validity is worthwhile.

Reliability is an essential quality of classification criteria as it represents the degree of consistency with repeated use[13], and for which there are insufficient data. At the level of the criteria set, inadequate reliability may result in misclassification of patients within and between studies, thereby threatening both the internal and external validity of the study results. The strength of a criteria set is threatened by a weak criterion. For example, poor inter- or intrarater reliability in skin assessment or capillaroscopy may result in misclassification of subsets. Similarly, poor within-laboratory or between-laboratory testing of SSc-selective antibodies may lead to misclassification of subsets. Reliability testing of the criteria sets and each criterion is therefore needed.

The next generation of classification criteria will need to consider some unresolved issues. First, over the time period of a longitudinal study, some patients may change subsets (the "transitional" form)[5]. Although clinical experts suggest this is uncommon, the prevalence of this shift and implications for classification criteria should be considered. Second, the relationship between localized and systemic scleroderma needs to be elucidated as some patients with SSc develop plaques of morphea or vice versa. Third, dependence on "extent of skin involvement" as the main criterion is being challenged. Data from the European League Against Rheumatism (EULAR) Scleroderma Trials and Research (EUSTAR) group suggests that autoantibody status is more closely associated with clinical manifestations than subset of disease[29]. However, the presence of anti-Scl-70 or ACA is not exclusively associated with particular disease manifestations. Thus research is needed to ascertain if autoantibody profiling confers incremental predictive validity over the subset criteria of LeRoy (or others). Alternatively, research is necessary to identify a combination of clinical and laboratory definitions for SSc subset classification criteria that confer improved (incremental) validity and reliability.

We have used the citation number as a measure of the frequency a criteria set was cited in the published literature. Our study did not evaluate why the criteria set was cited or how the criteria set was used. Criteria sets that have been in the public domain for a longer period of time, that were published in the English language, that had "buy-in" from multiple sites during their creation, or that are easier to use may all contribute to a higher citation number.

The 2-subset criteria of LeRoy, *et al*[25] have good feasibility, acceptable face validity, and good predictive validity. Although face and content validity improve with the addition of subsets and disease manifestations as criteria, the feasibility of the criteria is conversely reduced. Content validity of most criteria sets has not been evaluated due to the lack of an explicitly stated conceptual framework for SSc. Criteria with



Domain	Features	Clinical Example
A	Vasculopathy and fibrosis and immune activation	Systemic sclerosis[48,50,51], idiopathic pulmonary fibrosis
B	Vasculopathy	RP, idiopathic pulmonary arterial hypertension, hereditary hemorrhagic telangiectasia[23]
C	Immune activation	LE, antiphospholipid antibody syndrome
D	Fibrosis	Idiopathic pulmonary fibrosis, retroperitoneal fibrosis, nephrogenic fibrosis
E	Vasculopathy and fibrosis	Idiopathic pulmonary fibrosis
F	Fibrosis and immune activation	Eosinophilic fasciitis[54], idiopathic pulmonary fibrosis, primary biliary cirrhosis[55]
G	Vasculopathy and immune activation	Pulmonary arterial hypertension, RP[48,51], LE, dermatomyositis, polymyositis

RP: Raynaud's phenomenon, LE: lupus erythematosus, TGF-β: transforming growth factor-β, IL-4: interleukin 4, PDGF: platelet-derived growth factor

Figure 1. A conceptual framework for the construct of SSc based on clinical observations and supporting evidence from the literature.

3 or more subsets do not provide incremental predictive validity over the 2-subset criteria for survival. Research is needed to compare the content validity, divergent validity, and reliability of subset criteria for use in SSc trials.

REFERENCES

1. Johnson SR, Gladman DD, Schentag CT, Lee P. Quality of life and functional status in systemic sclerosis compared to other rheumatic diseases. J Rheumatol 2006;33:1117-22.
2. Rodnan GP, Jablonska S, Medsger TA. Classification and nomenclature of progressive systemic sclerosis. Clin Rheum Dis 1979;5:5-13.

3. Medsger TA Jr. Progressive systemic sclerosis. Clin Rheum Dis 1983;9:655-70.
4. Simeon CP, Armadans L, Fonollosa V, et al. Survival prognostic factors and markers of morbidity in Spanish patients with systemic sclerosis. Ann Rheum Dis 1997;56:723-8.
5. Valentini G. Classification of systemic sclerosis. Clin Dermatol 1994;12:217-23.
6. 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:S5-13.
7. Singh JA, Solomon DH, Dougados M, et al. Development of classification and response criteria for rheumatic diseases. Arthritis Rheum 2006;55:348-52.
8. Johnson SR, Goek ON, Singh-Grewal D, et al. Classification criteria in rheumatic diseases. A review of methodologic properties. Arthritis Care Res 2007; (in press).
9. Laxer RM, Zulian F. Localized scleroderma. Curr Opin Rheumatol 2006;18:606-13.
10. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985;38:27-36.
11. Wright JG, McLeod RS, Lossing A, Walters BC, Hu X. Measurement in surgical clinical research. Surgery 1996;119:241-4.
12. Feinstein AR. The theory and evaluation of sensibility. In: Feinstein AR, editor. Clinimetrics. New Haven: Yale University Press; 1987:141-65.
13. 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.
14. Felson DT, Anderson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. Baillieres Clin Rheumatol 1995;9:253-66.
15. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press; 1995.
16. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293-307.
17. Masi AT, Medsger TA Jr, Rodnan G, et al. Methods and preliminary results of the Scleroderma Criteria Cooperative Study of the American Rheumatology Association. Clin Rheum Dis 1979; 5:27-48.
18. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 1980;23:581-90.
19. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Bull Rheum Dis 1981;31:1-6.
20. Janssens X, Herman L, Mielants H, Verbruggen G, Veys EM. Disease manifestations of progressive systemic sclerosis: sensitivity and specificity. Clin Rheumatol 1987;6:532-8.
21. Nadashkevich O, Davis P, Fritzler MJ. A proposal of criteria for the classification of systemic sclerosis. Med Sci Monit 2004;10:CR615-CR621.
22. Goetz RH, Berne MB. The pathophysiology of progressive systemic sclerosis (generalised scleroderma) with special reference to changes in the viscera. Clin Proc 1945;4:337-92.
23. 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.
24. Barnett AJ, Miller M, Littlejohn GO. The diagnosis and classification of scleroderma (systemic sclerosis). Postgrad Med J 1988;64:121-5.
25. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202-5.
26. Silman AJ. Scleroderma and survival. Ann Rheum Dis 1991;50:267-9.
27. Black CM, Welsh KI, Maddison PJ, Jayson MI, Bernstein RM. HLA antigens, autoantibodies and clinical subsets in scleroderma. Br J Rheumatol 1984;23:267-71.
28. Giordano M, Valentini G, Migliaresi S, Picillo U, Vatti M. Different antibody patterns and different prognoses in patients with scleroderma with various extent of skin sclerosis. J Rheumatol 1986;13:911-6.
29. Walker UA, Tyndall A, Czirjak L, et al. Clinical risk assessment of organ manifestations in systemic sclerosis — a report from the EULAR Scleroderma Trials And Research (EUSTAR) group data base. Ann Rheum Dis 2007;66:754-63.
30. Ferri C, Valentini G, Cozzi F, et al. Systemic sclerosis: demographic, clinical, and serologic features and survival in 1,012 Italian patients. Medicine Baltimore 2002;81:139-53.
31. Hesselstrand R, Scheja A, Akesson A. Mortality and causes of death in a Swedish series of systemic sclerosis patients. Ann Rheum Dis 1998;57:682-6.
32. Kane GC, Varga J, Conant EF, Spirn PW, Jimenez S, Fish JE. Lung involvement in systemic sclerosis (scleroderma): relation to classification based on extent of skin involvement or autoantibody status. Respir Med 1996;90:223-30.
33. Barnett AJ. Scleroderma (progressive systemic sclerosis): progress and course based on a personal series of 118 cases. Med J Aust 1978;2:129-34.
34. Giordano M, Ara M, Capelli L, Tirri G, Vatti M. Variability of the clinical picture and the classification of progressive systemic scleroderma [German]. Z Rheumatol 1976;35:286-300.
35. Tuffanelli D, Winkelmann R. Diffuse systemic scleroderma. A comparison with acrosclerosis. Ann Intern Med 1962;57:198-203.
36. Barnett AJ, Coventry DA. Scleroderma. 1. Clinical features, course of illness and response to treatment in 61 cases. Med J Aust 1969;1:992-1001.
37. Burgos-Vargas R, Martinez-Cordero E, Reyes-Lopez PA, Herrera-Esparza R. Antibody pattern and other criteria for diagnosis and classification of PSS. J Rheumatol 1988;15:153-4.
38. Barnett AJ, Miller MH, Littlejohn GO. A survival study of patients with scleroderma diagnosed over 30 years (1953-1983): the value of a simple cutaneous classification in the early stages of the disease. J Rheumatol 1988;15:276-83.
39. Scussel-Lonzetti L, Joyal F, Raynauld JP, et al. Predicting mortality in systemic sclerosis: analysis of a cohort of 309 French Canadian patients with emphasis on features at diagnosis as predictive factors for survival. Medicine Baltimore 2002;81:154-67.
40. Ferri C, Bernini L, Cecchetti R, et al. Cutaneous and serologic subsets of systemic sclerosis. J Rheumatol 1991;18:1826-32.
41. LeRoy EC, Medsger TA Jr. Criteria for the classification of early systemic sclerosis. J Rheumatol 2001;28:1573-6.
42. Holzmann H, Sollberg S, Altmeyer P. Classification of progressive systemic scleroderma [German]. Hautarzt 1987;38:253-7.
43. Masi AT. Classification of systemic sclerosis (scleroderma): Relationship of cutaneous subgroups in early disease to outcome and serologic reactivity. J Rheumatol 1988;15:894-8.
44. Wollheim FA. Classification of systemic sclerosis. Visions and reality. Rheumatology Oxford 2005;44:1212-6.
45. Jacobsen S, Halberg P, Ullman S. Mortality and causes of death of 344 Danish patients with systemic sclerosis (scleroderma). Br J Rheumatol 1998;37:750-5.
46. Haynes SN, Lench HC. Incremental validity of new clinical assessment measures. Psychol Assess 2003;15:456-66.
47. Hawker GA, Gignac MA. How meaningful is our evaluation of meaningful change in osteoarthritis? J Rheumatol 2006;33:639-41.

48. Weiner ES, Hildebrandt S, Senecal JL, et al. Prognostic significance of anticentromere antibodies and anti-topoisomerase I antibodies in Raynaud's disease. A prospective study. Arthritis Rheum 1991;34:68-77.
49. Morse JH, Antohi S, Kasturi K, et al. Fine specificity of anti-fibrillin-1 autoantibodies in primary pulmonary hypertension syndrome. Scand J Immunol 2000;51:607-11.
50. Ulanet DB, Wigley FM, Gelber AC, Rosen A. Autoantibodies against B23, a nucleolar phosphoprotein, occur in scleroderma and are associated with pulmonary hypertension. Arthritis Rheum 2003;49:85-92.
51. Kallenberg CG, Wouda AA, Hoet MH, van Venrooij WJ. Development of connective tissue disease in patients presenting with Raynaud's phenomenon: a six year follow up with emphasis on the predictive value of antinuclear antibodies as detected by immunoblotting. Ann Rheum Dis 1988;47:634-41.
52. Poormoghim H, Lucas M, Fertig N, Medsger TA Jr. Systemic sclerosis sine scleroderma: demographic, clinical, and serologic features and survival in forty-eight patients. Arthritis Rheum 2000;43:444-51.
53. Holzmann H, Sollberg S, Altmeyer P. Classification of progressive systemic scleroderma. Hautarzt 1987;38:253-7.
54. Rodnan GP, DiBartolomeo A, Medsger TA Jr. Proceedings: Eosinophilic fasciitis. Report of six cases of a newly recognized scleroderma-like syndrome. Arthritis Rheum 1975;18:525.
55. Mayo MJ, Jenkins RN, Combes B, Lipsky PE. Association of clonally expanded T cells with the syndrome of primary biliary cirrhosis and limited scleroderma. Hepatology 1999;29:1635-42.