Agreement Between Physician Evaluation and the Composite Response Index in Diffuse Cutaneous Systemic Sclerosis

Zheng, B.Y.; Wang, M.B.; McKenna, K.; Shapiro, L.; Silver, R.; Csuka, M.E.; ... ; Canadian Scleroderma Res Grp

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 74, No. 11, November 2022, pp 1806–1812
DOI 10.1002/acr.24638

Agreement Between Physician Evaluation and the Composite Response Index in Diffuse Cutaneous Systemic Sclerosis

Boyang Zheng,[1] Mianbo Wang,[2] Kerry McKenna,[2] Lee Shapiro,[3] Richard Silver,[4] Mary Ellen Csuka,[5] Frank van den Hoogen,[6] David Robinson,[7] John D. Pauling,[8] Laura Hummers,[9] Thomas Krieg,[10] Francesco Del Galdo,[11] Robert Spiera,[12] Niall Jones,[13] Nader Khalidi,[14] Alessandra Vacca,[15] Jeska K. de Vries-Bouwstra,[16] Jessica Gordon,[12] and Murray Baron,[1] for the Canadian Scleroderma Research Group

Objective. Diffuse cutaneous systemic sclerosis (SSc) is a highly heterogeneous disease. A provisionally approved Composite Response Index in diffuse cutaneous SSc (CRISS) was developed as a 1-year outcome measure for clinical trials. Our goal was to further validate the CRISS by examining agreement between CRISS definitions for improved/non-improved with physicians' evaluation of disease.

Methods. Patient profiles from a large observational cohort were created for 50 random diffuse cutaneous SSc patients of <5 years disease duration with improved CRISS scores after 1 year and 50 with non-improved CRISS scores. Profiles described disease features used during the initial CRISS development at baseline and at 1 year. Each profile was independently rated by 3 expert physicians. Majority opinion determined whether a patient was improved or not improved, and kappa agreement with the CRISS cutoff of 0.6 was calculated.

Results. Patients had mean ± SD disease duration of 2.2 ± 1.3 years. There was substantial agreement between the physician majority opinion about each case and the CRISS (κ = 0.76 [95% confidence interval (95% CI) 0.64–0.88]). The agreement between each individual physician opinion and the CRISS was also substantial (κ = 0.70 [95% CI 0.62–0.78]). All CRISS non-improvers were also rated as non-improved by physician majority; however, 12 CRISS improvers were rated as non-improved by physicians.

Conclusion. There was substantial agreement between the dichotomous CRISS rating and physician assessment of diffuse cutaneous SSc patients after 1 year. This supports the use of a CRISS cutoff at 0.6 for improvement versus non-improvement, although the CRISS tended to rate more patients as improved than did physicians.

INTRODUCTION

Systemic sclerosis (SSc) is an autoimmune fibrosing disorder affecting the skin and internal organs. In the diffuse cutaneous subtype, the hallmark is widespread skin thickening that extends beyond the elbows and/or knees. This subtype is associated with a worse prognosis (1) and increased organ involvement including the lungs, skin, heart, gastrointestinal tract, and kidneys. These

The Canadian Scleroderma Research Group was supported by the Canadian Institutes of Health Research, the Scleroderma Society of Canada and its Chapters of Ontario, Saskatchewan, and Quebec, the Cure Scleroderma Foundation, INOVA Diagnostics (San Diego, CA), Dr. Fooke Laboratorien (Neuss, Germany), Euroimmun (Lubeck, Germany), Mikrogen (Neuried, Germany), Fonds de la recherche en santé du Québec, the Canadian Arthritis Network, and the Lady Davis Institute of the Jewish General Hospital, Montreal, Quebec.

[1]Boyang Zheng, MD, Murray Baron, MD: McGill University, Jewish General Hospital, Montreal, Quebec, Canada; [2]Mianbo Wang, MSc, Kerry McKenna: Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada; [3]Lee Shapiro, MD: Albany Medical College, The Center for Rheumatology, Albany, New York; [4]Richard Silver, MD: Medical University of South Carolina, Charleston; [5]Mary Ellen Csuka, MD: Medical College of Wisconsin, Milwaukee; [6]Frank van den Hoogen, MD, PhD: Sint Maartenskliniek, Nijmegen, Gelderland, The Netherlands; [7]David Robinson, MD: University of Manitoba, Winnipeg, Manitoba, Canada; [8]John D. Pauling, BMBS, PhD: Royal National Hospital for Rheumatic Diseases, Bath, UK; [9]Laura Hummers MD, MSc: Johns Hopkins University, Baltimore, Maryland; [10]Thomas Krieg, MD: University of Cologne, Cologne, Germany; [11]Francesco Del Galdo, MD, PhD: University of Leeds, St. James University Hospital, West Yorkshire, UK; [12]Robert Spiera, MD, Jessica Gordon, MD: Hospital for Special Surgery, New York, New York; [13]Niall Jones, MD: University of Alberta, Edmonton, Alberta, Canada; [14]Nader Khalidi, MD: McMaster University, Hamilton, Ontario, Canada; [15]Alessandra Vacca, MD: University Hospital of Cagliari, Cagliari, Italy; [16]Jeska K. de Vries-Bouwstra, MD, PhD: Leiden University Medical Center, Leiden, The Netherlands.

Dr. Silver has received consulting fees from Boehringer Ingelheim (more than $10,000). Dr. Pauling has received consulting fees, speaking fees, and/or honoraria from Actelion, Sojournix Pharma, and Boehringer Ingelheim (less than $10,000 each). Dr. Krieg has received consulting fees, speaking fees, and/or honoraria from Actelion (less than $10,000). Dr. Del Galdo has

SIGNIFICANCE & INNOVATIONS

- Agreement between the Combined Response Index in diffuse cutaneous Systemic Sclerosis (CRISS) and an independent panel of expert physicians was substantial in a large diffuse cutaneous systemic sclerosis (SSc) cohort.
- CRISS specificity for improvement compared to physician opinion was lower than the original derivation cohort, and the CRISS was more useful to identify non-improved subjects.
- Our findings support the validity of a dichotomous CRISS rating to assess disease changes at 1 year in diffuse cutaneous SSc.

manifestations are heterogeneous and range in severity, as each organ might be affected differently (2–4).

The multifaceted aspects of the disease and the lack of adequate disease measures are particularly problematic in diffuse cutaneous SSc clinical trials. In the absence of a validated global response measure, outcomes have often been evaluated on an organ-by-organ basis, with the extent of skin involvement historically emerging as the preferred end point for previous clinical trials of diffuse cutaneous SSc. While skin involvement is weakly correlated with other disease features (5,6), changes in 1 organ system are often insufficient to accurately represent the overall disease state (7).

To address this, the American College of Rheumatology (ACR) has approved a provisional Composite Response Index in diffuse cutaneous SSc (CRISS) (8) with the goal of improving outcome assessment in clinical trials. The CRISS assesses the likelihood of improvement after 1 year, with a cutoff score of ≥0.6 considered as the most sensitive and specific for improved disease and <0.6 for not improved (8). This measure assesses the changes over 1 year in 5 measures: skin involvement based on the modified Rodnan skin thickness score (MRSS), lung involvement based on the forced vital capacity (FVC), patient and physician global assessments of disease activity, and function based on the Health Assessment Questionnaire disability index (HAQ DI). While the use of the CRISS has been adopted in several recent or ongoing clinical trials (9–14), it lacks extensive external validation.

Our main goal was to assess the validity of the CRISS in an observational cohort by examining the agreement between the CRISS score and physicians' opinions of disease evolution in real-life diffuse cutaneous SSc patients. We hypothesized that there would be good agreement between the CRISS and physician assessments to identify an improvement in diffuse cutaneous SSc patients after 1 year. Because physician opinions were utilized as the gold standard to develop the CRISS, we also aimed to determine the sensitivity and specificity of the CRISS for physician-assessed improvement.

PATIENTS AND METHODS

Study population. The Canadian Scleroderma Research Group (CSRG) registry follows patients from 15 centers in Canada and Mexico with a confirmed diagnosis of SSc made by an experienced rheumatologist. Registry patients must be ≥18 years of age, provide informed consent, and be fluent in English, French, or Spanish. One hundred CSRG patients with diffuse cutaneous SSc, defined by skin thickening proximal to the elbows or knees and/or trunk at any time, enrolled between January 2005 and July 2017 were selected. Patients must have <5 years disease duration from the first non–Raynaud's phenomenon symptom and have at least 2 consecutive yearly follow-up visits with all required CRISS features available. Fifty patients were randomly selected among those with a CRISS score of ≥0.6, and a remaining 50 patients were randomly selected from among those with a CRISS score of <0.6 to develop 100 patient profiles. Ethics committee approval for this study was obtained at the Jewish General Hospital, Montreal, Canada and at all participating CSRG study sites. See Appendix A for additional members of the Canadian Scleroderma Research Group and their locations.

CRISS calculation. The CRISS score was calculated based on 2 steps (8). Step 1 identified patients with significant worsening or new end-organ damage and automatically assigned them a score of 0. The criteria for significant worsening or end-organ damage are as follows: new-onset scleroderma renal crisis; new left-sided heart failure with left ventricular ejection fraction of ≤45% on transthoracic echocardiogram requiring treatment; new pulmonary arterial hypertension (confirmed on right-sided heart catheterization) requiring treatment; ≥15% decline in percent of predicted FVC (FVC%) or FVC% <80% predicted in the presence of new interstitial lung disease on lung computed tomography scan. Step 2 of the CRISS estimates a likelihood of improvement after 1 year using the CRISS equation. This uses a complex probability equation derived from the logistic regression model that takes into account the change over 1 year in the following: MRSS, FVC%, patient and physician global assessment of disease severity, and HAQ DI (8). Scores range from 0–1. The CRISS score obtained from this equation was used as a

received consulting fees, speaking fees, and/or honoraria from AstraZeneca, Boehringer Ingelheim, Capella, Chemomab, Actelion, and Mitsubishi (less than $10,000 each). Dr. Spiera has received consulting fees from AbbVie, Roche-Genetech, GlaxoSmithKline, CSL Behring, Sanofi, Janssen, Chemocentryx, and Formation Biologics (less than $10,000 each) and research support from Roche-Genetech, GlaxoSmithKline, Bristol Myers Squibb, Boehringer Ingelheim, Chemocentryx, Corbus, Formation Biologics, Sanofi, and Inflarx. Dr. Khalidi has received consulting fees, speaking fees, and/or honoraria from Roche (less than $10,000). Dr. de Vries-Bouwstra has received consulting fees, speaking fees, and/or honoraria from Boehringer Ingelheim (less than $10,000). No other disclosures relevant to this article were reported.

Address correspondence to Murray Baron, MD, Chief Division of Rheumatology, Jewish General Hospital, Suite A-725, 3755 Cote St. Catherine Rd, Montréal, Quebec H3T 1E2, Canada. Email: mbaron@jgh.mcgill.ca.

Submitted for publication September 27, 2020; accepted in revised form April 27, 2021.

dichotomous variable of improved (score ≥0.6) or not improved (score <0.6).

Patient profiles. Patient profiles were developed based on the 15 core disease features used in the initial development of the CRISS score and were recorded at the first (baseline) and second (1 year) visit. These included patient-rated components from the Scleroderma HAQ (SHAQ) as well as patient and physician global assessment scores. These were rated 0–10 (no disease to very severe disease) on numerical rating scales. Individual SHAQ questions asked patients to "rate in the past week: how much have your 'breathing problems,' 'intestinal problems,' 'Raynaud's,' 'finger ulcer(s)' interfered with your daily activities." The pain question asked "in the past week, how much pain have you had because of your illness?" The patient global assessment question asked "in the past week, how was your overall health?" The physician global severity question asked "How would you rate the patient's overall health for the past week?" Function was assessed using the HAQ DI, which is scored from 0 (no disability) to 3 (severe disability). Fatigue was assessed using the Short Form 36 (SF-36) health survey vitality scale, and overall health-related quality of life using the SF-36 physical component score. Skin involvement was assessed using the MRSS, which ranges from 0 (no involvement) to 3 (severe thickening) in 17 areas (score range 0–51). Tendon friction rubs and history of renal crisis were recorded as present or absent. The number of digital ulcers was noted at each visit. FVC% was extracted from pulmonary function tests. Body mass index was obtained from patient measurements. Age and disease duration at baseline were provided.

Fifteen experienced scleroderma physicians assessed the patient profiles. Each physician self-reported their clinical experience and practice setting. No additional contextual information regarding the cases was provided. Profiles were assigned in such a way that each patient was assessed by 3 randomly assigned physicians. Based on the changes over 1 year, physicians then rated each profile as "improved," "stable," "worsened," or "unable to tell." Ratings of "stable," "worsened," and "unable to tell" were considered as physician-rated "not improved" in the primary analyses. A physician majority rating of "improved" or "not improved" was reached if at least 2 of 3 physicians rated the case in the same way. In sensitivity analyses to examine the impact of the "unable to tell" rating, only "stable" and "worsened" were considered as "not improved" for physician consensus, while physician ratings of "unable to tell" were considered as missing values. Physicians also answered a survey on the number of years in practice, their type of practice (academic hospital, community hospital, outpatient clinic), and the average number of SSc patients they manage per year.

Statistical analysis. Descriptive statistics were used to summarize baseline demographic and clinical characteristics. Cohen's kappa was calculated to assess the level of agreement between the majority physician opinion of patient profiles and

Table 1. Baseline characteristics of study patients*

	CRISS improver (n = 50)	CRISS non-improver (n = 50)	All patients (n = 100)
Disease features at baseline			
Age, mean ± SD years	52.4 ± 11.1	51.1 ± 13.4	51.8 ± 12.3
Disease duration, mean ± SD years	2.3 ± 1.3	2.1 ± 1.4	2.2 ± 1.3
Female	35 (70)	40 (80)	75 (75)
Immunosuppression†	23 (46)	21 (42)	44 (44)
Interstitial lung disease	18 (36)	21 (42.9)	39 (39.4)
Pulmonary hypertension	4 (8.2)	2 (4.3)	6 (6.3)
Inflammatory myositis	11 (22)	7 (14)	18 (18)
Inflammatory arthritis	12 (24.5)	12 (25.5)	24 (25)
Digital ulcers	27 (54)	21 (42)	48 (48)
Prior renal crisis	6 (12)	0	6 (6)
Autoantibodies			
Anticentromere	6 (13.6)	8 (18.2)	14 (15.9)
Anti–topoisomerase I	7 (15.9)	13 (29.6)	20 (22.7)
Anti–RNA polymerase III	19 (43.2)	13 (29.6)	32 (36.4)
CRISS variables, mean ± SD			
MRSS	23.0 ± 9.1	18.1 ± 8.9	20.6 ± 9.3
FVC%	89.1 ± 17.9	90.2 ± 16.3	89.6 ± 17.0
HAQ	1.2 ± 0.7	0.9 ± 0.7	1.1 ± 0.7
Patient global assessment	4.9 ± 2.5	3.4 ± 2.6	4.2 ± 2.6
Physician global assessment	4.7 ± 2.6	4.2 ± 2.1	4.4 ± 2.4

* Values are the number (%) unless indicated otherwise. CRISS = Combined Response Index in diffuse cutaneous Systemic Sclerosis; FVC% = percent of predicted forced vital capacity; HAQ = Health Assessment Questionnaire; MRSS = modified Rodnan skin thickness score.
† Immunosuppression includes the use of methotrexate, mycophenolate, azathioprine, or cyclophosphamide at the baseline visit.

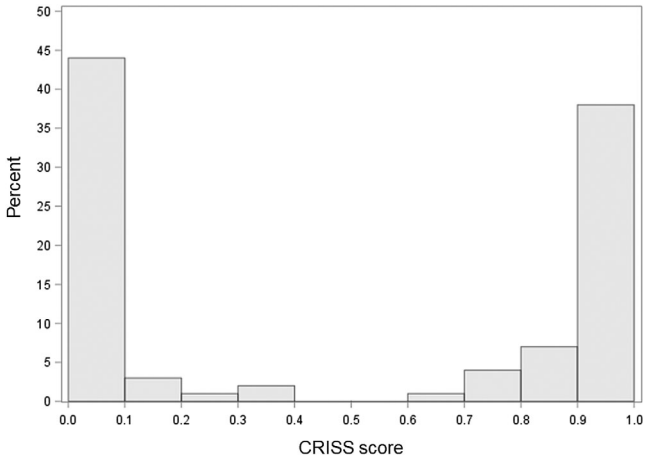

Figure 1. Distribution of Combined Response Index in diffuse cutaneous Systemic Sclerosis (CRISS) scores in the entire cohort of 100 patients with diffuse cutaneous systemic sclerosis.

the CRISS score. Any patient for whom a majority opinion could not be reached because of a missing physician response was excluded. To assess the effect of "unable to tell" ratings, kappa was also calculated by considering physician "unable to tell" ratings as missing values. Kappa values between 0.41 and 0.60 were considered as moderate agreement, 0.61–0.80 as substantial, and 0.81–1 as excellent, almost perfect agreement (15). Sensitivity, specificity, positive predictive value, and negative predictive value of the CRISS for physician majority opinion were subsequently determined.

RESULTS

Baseline characteristics of the 100 randomly selected diffuse cutaneous SSc patients are shown in Table 1. All patients fulfilled the ACR/European Alliance of Associations for Rheumatology 2013 classification criteria, and the mean ± SD disease duration was 2.2 ± 1.3 years. CRISS improvers more frequently had a history of pulmonary hypertension, inflammatory myositis, renal crises, and the presence of anti–RNA polymerase III autoantibodies at baseline (Table 1). The distribution of CRISS scores in the cohort favored the extreme CRISS values near 0 or 1. Of the 50 CRISS non-improvers, 44 (88%) had a CRISS score of <0.1. Of the 50 CRISS improvers, 38 (76%) had a CRISS score of ≥0.9 (Figure 1).

The 15 physician raters represent a multinational group of experts from the US, Canada, Europe, and UK. The majority (86.7%) have >15 years of practice in an academic institution and manage >50 SSc patients every year.

A total of 300 surveys were sent out because each patient profile was rated by 3 different physicians. The distribution of individual physician opinions is shown in Table 2. The kappa agreement between the CRISS and each individual physician response was substantial (κ = 0.70 [95% confidence interval (95% CI) 0.62–0.78]). The kappa agreement between the CRISS and physician majority opinion was also substantial (κ = 0.76 [95% CI 0.64–0.88]). We were able to obtain a majority physician opinion for all 100 patient profiles, despite 3 missing responses, because at least 2 of 3 physicians rated the case the same way (as either improved or non-improved). Based on physician majority opinion, all CRISS non-improvers were also rated as non-improved. There were disagreements concerning 12 CRISS improvers (24% of all CRISS improvers) who were considered non-improved by physician majority opinion (Table 3). Of these 12 cases, 8 were considered stable by the majority of raters, and the 4 remaining patients had split physician opinion. Only a single worsened rating was present in 2 of the cases with split physician opinions (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24638).

The sensitivity of the CRISS cutoff when compared to physician majority opinion as the gold standard was 1.00 (95% CI 0.91–1.00), and the specificity was 0.81 (95% CI 0.69–0.90). The positive predictive value of the CRISS for improved disease was 0.76 (95% CI 0.62–0.87), and the negative predictive value was 1.00 (95% CI 0.93–1.00). The majority (82%) of patients rated as improved by physician majority opinion had CRISS scores of ≥0.9, and the majority (76%) of patients with non-improved physician majority opinion had CRISS scores of ≤0.1.

We also assessed the effect of the "unable to tell" responses. Twenty-one patient profiles were rated by at least 1 physician as "unable to tell" regarding how the disease changed after 1 year. When these ratings were considered as missing, we required both remaining physicians to agree on improved or not improved for those cases. There were 7 profiles for which consensus now

Table 2. Individual physician opinions of 300 patient profiles and agreement with Combined Response Index in diffuse cutaneous Systemic Sclerosis (CRISS) score*

	Improved	Stable†	Worsened†	Unable to tell†	Missing response‡	Total
CRISS improver (CRISS score ≥ 0.6)	105	31	2	10	2	150
CRISS non-improver (CRISS score < 0.6)	2	58	75	14	1	150
Total	107	89	77	24	3	300

* Values are the number. Agreement between CRISS score and all physician responses before consensus κ = 0.70 (95% confidence interval 0.62–0.78).
† A physician's responses of stable, worsened, or unable to tell were all considered as non-improved.
‡ Missing responses were excluded from this kappa calculation.

Table 3. Physician majority opinions of 100 patient profiles and agreement with Combined Response Index in diffuse cutaneous Systemic Sclerosis (CRISS) score*

	Improved	Non-improved	Total
CRISS improver (CRISS score ≥ 0.6)	38	12	50
CRISS non-improver (CRISS score < 0.6)	0	50	50
Total	38	62	100

* Values are the number. Agreement between CRISS score and physician majority opinion κ = 0.76 (95% confidence interval 0.64–0.88).

became impossible. After excluding these profiles from the kappa calculation, the agreement was stronger (κ = 0.81 [95% CI 0.69–0.92]).

DISCUSSION

This is the first study to validate the CRISS in a longitudinal observational cohort. We found a substantial agreement between the dichotomous CRISS rating and physician opinion of clinical changes after 1 year in early diffuse cutaneous SSc patients. This strong agreement was found when we assessed individual physician opinions, majority physician opinions, and when "unable to tell" physician opinions were removed. This provides an external validation of the CRISS score in a manner comparable to what was used in the determination of the CRISS cutoff but with a separate set of physicians assuming the role of evaluators.

This study could be considered as a form of criterion validity, i.e., comparison with a gold standard, as one can hypothesize that any valid assessment of outcome should agree with physician assessment, which in essence is the only gold standard available. This was the underpinning of the CRISS score development, where a cutoff of ≥0.6 had the highest sensitivity and specificity for improvement based on physician opinion. As physician majority opinion was considered the gold standard in our study, we found that this CRISS cutoff for improvement was extremely sensitive and excluded all physician-assessed non-improvers. The specificity of the CRISS was lower in our study than in the original derivation cohort (81% versus 93%). Physicians were less likely than the CRISS to rate a patient as improved, and 24% of patients with a CRISS score of ≥0.6 were considered not improved according to expert opinion. It is difficult to know what impact this degree of error would have in clinical trials, but this finding might influence clinical trial sample sizes in order to increase the precision of the results.

All disagreement stemmed from cases in which the CRISS indicated improvement but physicians felt that these patients were either stable or had split opinions between improved and unable to tell/stable categories. The basis for this disagreement is difficult to assess and may have been related to changes in features in the patient profiles that were not captured by the CRISS. Finally the different weighting that individual physicians apply to different disease variables is a complex issue highlighted by various working groups on SSc outcome measures (16,17). Additional validation of the CRISS against other SSc measures and organ-specific outcomes will be useful, particularly to inform on construct validity as well as content validity.

Several drug trials examining the CRISS score as a secondary outcome have analyzed the results as a continuous variable. Differences were seen in the median CRISS scores between treatment and placebo arms of clinical trials for abatacept (18), tocilizumab (19), lenabasum (20), and belimumab (21) in diffuse cutaneous SSc. While these results support the discriminatory capacity of the CRISS, our study was not designed to assess the validity of the CRISS as a continuous variable. Because our cohort tended to have either very low or very high CRISS scores, we were unable to perform exploratory analyses to examine agreement outcomes at different CRISS cutoff levels. We were also unable to assess what would constitute a clinically meaningful change in the CRISS score.

Our study is not without some limitations. One is that our physician majority opinion does not represent a true consensus across each set of raters. During the original CRISS development process, a higher majority (≥75% of raters and/or steering committee members) was required to reach consensus, and profiles that could not reach consensus were not included as part of the development process (22). However, in our study, the kappa agreement between the CRISS and individual physician assessment was only slightly lower than the agreement with physician majority opinion, and both kappa values remained substantial. The response by physicians to rate profiles as "unable to tell" was regarded conservatively to represent non-improvement, erring on the side of caution. However, if these were regarded as non-responses and the resulting non-consensus profiles excluded, as during the original CRISS development, then the kappa agreement was even stronger. Thus, we found that at worst the κ agreement was 0.70 (95% CI 0.62–0.78) and at best 0.81 (95% CI 0.69–0.92).

The extreme bimodal distribution of CRISS values could have facilitated agreement between physician opinions with the CRISS. Agreement may have been lower if patient CRISS scores were closer to the 0.6 cutoff. This bimodal distribution was found in the original CRISS development cohort (22), and the reasons for this have not been investigated. While patients who developed severe organ involvement with an automatic CRISS score of 0 account for some of the very low scores, the clustering of very high scores is difficult to explain.

Although the CRISS was designed for use in clinical trials, it was created using observational patient cohorts, and there is no compelling reason to suspect that these conclusions should differ in patients with matching inclusion criteria. However, expanded use in everyday clinical practice may be limited by bias introduced from unblinded observations and the difficulty in rapidly calculating the score at the bedside. Furthermore, the validity of the measure beyond the intended SSc population remains unclear, and the clinical meaningfulness of the score in prognosis and mortality requires further elucidation.

Further, although physicians were blinded to the actual CRISS score, the component features of the CRISS are included in the patient profiles. Seeing the change in these variables may allow raters to estimate a CRISS value; however, exact calculations would be extremely difficult. We did not change the way patient profiles were established in order to maintain consistency with the original CRISS development study. Patient profiles included all SSc domains with the exception of cardiac involvement and pulmonary hypertension (23). While the presence of existing pulmonary hypertension was tabulated at baseline, it was not part of the patient profiles. Any new or worsening heart failure and/or pulmonary hypertension would have automatically been rated by the CRISS algorithm as non-improved. However, given that all CRISS non-improvers were also rated as non-improved by physicians, the absence of these 2 items is unlikely to have made an impact on our results.

A final point is that the number of diffuse cutaneous SSc patients with anticentromere antibody (ACA) (16%) is slightly higher in our study than those reported in other cohorts, varying between 6% and 12% (24–26). While this may be a concern for disease misclassification, our study sample is consistent with the overall pattern of diffuse cutaneous SSc previously reported in the CSRG (27). All patients were assessed by expert rheumatologists at baseline and had skin thickening extending proximally to either elbows or knees to be classified as diffuse cutaneous SSc. Furthermore, ACA presence should not have biased the agreement between physician opinion and the CRISS, given that autoantibody status is not disclosed to physicians nor accounted for by the CRISS.

Strengths of this study include the large number of cases assessed in a way that closely mimics the development of the CRISS. None of the expert physician raters were involved in the original development process. As there is no a priori agreement on a standard method for determining physician opinion about individual cases, we used several different methods of assigning physician opinion, and all methods provided substantial agreement with the CRISS. Also, we included patients with <5 years disease duration, and patients had a mean ± SD disease duration of 2.2 ± 1.3 years, which is similar to that of the intended population for which the CRISS will be used, i.e., diffuse cutaneous SSc subjects with short disease duration recruited for clinical trials.

In conclusion, we found that the agreement between the CRISS and expert physician opinion was substantial in evaluating disease improvement in diffuse cutaneous SSc. This finding supports the use of a dichotomous CRISS outcome of either improved or non-improved after 1 year. The specificity of the CRISS was slightly lower than in the original derivation cohort. Further work could help validate the CRISS in other cohorts and assess its association with other SSc outcomes and markers of disease improvement such as survival.

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 submitted for publication. Dr. Baron 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. Zheng, Wang, Baron.

Acquisition of data. McKenna, Shapiro, Silver, Csuka, van den Hoogen, Robinson, Pauling, Hummers, Krieg, Del Galdo, Spiera, Jones, Khalidi, Vacca, de Vries-Bouwstra, Gordon, Baron.

Analysis and interpretation of data. Zheng, Wang, Baron.

REFERENCES

1. 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.
2. Clements PJ, Hurwitz EL, Wong WK, Seibold JR, Mayes M, White B, et al. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. Arthritis Rheum 2000;43:2445–54.
3. Geirsson AJ, Wollheim FA, Akesson A. Disease severity of 100 patients with systemic sclerosis over a period of 14 years: using a modified Medsger scale. Ann Rheum Dis 2001;60:1117–22.
4. Hachulla E, Carpentier P, Gressin V, Diot E, Allanore Y, Sibilia J, et al. Risk factors for death and the 3-year survival of patients with systemic sclerosis: the French ItinerAIR-Sclerodermie study. Rheumatology (Oxford) 2009;48:304–8.
5. Nevskaya T, Zheng B, Baxter CA, Ramey DR, Pope JE, Baron M, et al. Skin improvement is a surrogate for favourable changes in other organ systems in early diffuse cutaneous systemic sclerosis. Rheumatology (Oxford) 2020;59:1715–24.
6. Zheng B, Nevskaya T, Baxter CA, Ramey DR, Pope JE, Baron M, et al. Changes in skin score in early diffuse cutaneous systemic sclerosis are associated with changes in global disease severity. Rheumatology (Oxford) 2020;59:398–406.
7. Melsens K, De Keyser F, Decuman S, Piette Y, Vandecasteele E, Smith V. Disease activity indices in systemic sclerosis: a systematic literature review. Clin Exp Rheumatol 2016;34 Suppl 100:186–92.
8. Khanna D, Berrocal VJ, Giannini EH, Seibold JR, Merkel PA, Mayes MD, et al. The American College of Rheumatology Provisional Composite Response Index for clinical trials in early diffuse cutaneous systemic sclerosis. Arthritis Rheumatol 2016;68:299–311.
9. Distler O, Pope J, Denton C, Allanore Y, Matucci-Cerinic M, de Oliveira Pena J, et al. RISE-SSc: riociguat in diffuse cutaneous systemic sclerosis. Respir Med 2017;122 Suppl 1:S14–S7.
10. Olson AL, Gifford AH, Inase N, Fernandez Perez ER, Suda T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. Eur Respir Rev 2018;27.

11. Song JW, Lee HK, Lee CK, Chae EJ, Jang SJ, Colby TV, et al. Clinical course and outcome of rheumatoid arthritis-related usual interstitial pneumonia. Sarcoidosis, vasculitis, and diffuse lung diseases: official journal of WASOG 2013;30:103–12.
12. Hoffmann-La Roche, sponsor. A study of the efficacy and safety of tocilizumab in participants with systemic sclerosis (SSc). ClinicalTrials.gov identifier: NCT02453256; 2015.
13. Morisset J, Mageto Y, Raghu G. Mortality in interstitial lung disease: do race and skin colour matter? Eur Respir J 2018;51.
14. Kim SK, Park SH, Shin IH, Choe JY. Anti-cyclic citrullinated peptide antibody, smoking, alcohol consumption, and disease duration as risk factors for extraarticular manifestations in Korean patients with rheumatoid arthritis. J Rheumatol 2008;35:995–1001.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
16. Baron M, Kahaleh B, Bernstein EJ, Chung L, Clements PJ, Denton C, et al. An interim report of the Scleroderma Clinical Trials Consortium Working Groups. J Scleroderma Relat Disord 2019;4:17–27.
17. Valentini G, Bencivelli W, Bombardieri S, D'Angelo S, Della Rossa A, Silman AJ, et al. European Scleroderma Study Group to define disease activity criteria for systemic sclerosis. III. Assessment of the construct validity of the preliminary activity criteria. Ann Rheum Dis 2003;62:901–3.
18. Khanna D, Spino C, Johnson S, Chung L, Whitfield ML, Denton CP, et al. Abatacept in early diffuse cutaneous systemic sclerosis: results of a phase II investigator-initiated, multicenter, double-blind, randomized, placebo-controlled trial. Arthritis Rheumatol 2020;72:125–36.
19. Khanna D, Berrocal V, Denton C, Jaheris A, Spotwood H, Lin C, et al. SAT0373 evaluation of American College of Rheumatology provisional composite response (CRISS) index in the Fasscinate trial. Ann Rheum Dis 2017;76 Suppl 2:912.
20. Spiera R, Khanna D, Dgetluck N, Conley B, White B. OP0069 Performance of American College of Rheumatology (ACR) Combined Response Index in diffuse cutaneous Systemic Sclerosis (CRISS) score in phase 2 trial of lenabasum in diffuse cutaneous systemic sclerosis (DCSS). Ann Rheum Dis 2019;78 Suppl 2:107.
21. Gordon JK, Martyanov V, Franks JM, Bernstein EJ, Szymonifka J, Magro C, et al. Belimumab for the treatment of early diffuse systemic sclerosis: results of a randomized, double-blind, placebo-controlled, pilot trial. Arthritis Rheumatol 2018;70:308–16.
22. Khanna D, Distler O, Avouac J, Behrens F, Clements PJ, Denton C, et al. Measures of response in clinical trials of systemic sclerosis: the Combined Response Index for Systemic Sclerosis (CRISS) and Outcome Measures in Pulmonary Arterial Hypertension related to Systemic Sclerosis (EPOSS). J Rheumatol 2009;36:2356–61.
23. Harel D, Hudson M, Iliescu A, Baron M, Canadian Scleroderma Research Group, Steele R. Summed and weighted summary scores for the Medsger disease severity scale compared with the physician's global assessment of disease severity in systemic sclerosis. J Rheumatol 2016;43:1510–8.
24. Walker UA, Tyndall A, Czirjak L, Denton C, Farge-Bancel D, Kowal-Bielecka O, et al. Clinical risk assessment of organ manifestations in systemic sclerosis: a report from the EULAR Scleroderma Trials and Research Group database. Ann Rheum Dis 2007;66:754–63.
25. Morrisroe K, Stevens W, Sahhar J, Rabusa C, Nikpour M, Proudman S, et al. Epidemiology and disease characteristics of systemic sclerosis-related pulmonary arterial hypertension: results from a real-life screening programme. Arthritis Res Ther 2017;19:42.
26. 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.
27. Srivastava N, Hudson M, Tatibouet S, Wang M, Baron M, Fritzler MJ, et al. Thinking outside the box: the associations with cutaneous involvement and autoantibody status in systemic sclerosis are not always what we expect. Semin Arthritis Rheum 2015;45:184–9.

APPENDIX A: THE CANADIAN SCLERODERMA RESEARCH GROUP

Members of the Canadian Scleroderma Research Group, in addition to the authors, are as follows: Janet E. Pope (Western University, St. Joseph's Health Care, London, Ontario, Canada), Marie Hudson (McGill University, Jewish General Hospital, Montreal, Quebec, Canada), Geneviève Gyger (McGill University, Jewish General Hospital, Montreal, Quebec, Canada), Maggie J. Larché (McMaster University, St. Joseph's Healthcare, Hamilton, Ontario, Canada), Ariel Masetto (Université de Sherbrooke, Sherbrooke, Quebec, Canada), Evelyn Sutton (Dalhousie University, Nova Scotia Rehabilitation Centre, Halifax, Nova Scotia, Canada), Tatiana S. Rodriguez-Reyna (Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubiran, Mexico City, Mexico), Nancy Maltez (University of Ottawa, Ottawa, Ontario, Canada), Doug Smith (University of Ottawa, Ottawa, Ontario, Canada), Carter Thorne (Southlake Regional Health Centre, Newmarket, Ontario, Canada), Alena Ikic (Université Laval, CHU de Québec, Quebec City, Quebec, Canada), Paul R. Fortin (Université Laval, CHU de Québec, Quebec City, Quebec, Canada), and Marvin J. Fritzler (University of Calgary, Ottawa, Ontario, Canada).