**ORIGINAL ARTICLE**

WILEY International Journal of Rheumatic Diseases

# Evaluation of Scleroderma Clinical Trials Consortium training recommendations on modified Rodnan skin score assessment in scleroderma

Andrea H. L. Low[1,2] | Sue-Ann Ng[1,2] (iD) | Veronica Berrocal[3] | Benjamin Brennan[3] | Grace Chan[4] | Swee-Cheng Ng[1,2] | Dinesh Khanna[5]

[1]Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore

[2]Duke-National University of Singapore, Singapore, Singapore

[3]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan

[4]Department of Rheumatology, Allergy and Immunology, Tan Tock Seng Hospital, Singapore, Singapore

[5]Scleroderma Program, University of Michigan, Ann Arbor, Michigan

**Correspondence**
Andrea H. L. Low, Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore.
Email: andrea.low.h.l@singhealth.com.sg

**Funding information**
Dr Khanna was funded by the NIH/NIAMS K24 AR06312. Dr Low was funded by the NMRC.

## Abstract

**Aim:** The modified Rodnan skin score (mRSS) is a validated outcome measure for skin thickness in systemic sclerosis (SSc). Training has been shown to reduce variability in the measurement of mRSS. Our objective was to assess the inter- and intra-observer variability of mRSS scoring using the proposed recommendations for training by the Scleroderma Clinical Trials Consortium (SCTC) and World Scleroderma Foundation (WSF).

**Method:** Fifty-two trainees and eight adult SSc patients participated in the SSc skin scoring workshop that was conducted in two sessions by four teachers. Each session, attended by 26 trainees, had a teaching and evaluation phase. The teaching phase comprised of: (a) lecture on mRSS scoring; (b) video demonstration of mRSS scoring; and (c) live demonstration of mRSS on one SSc patient. In the evaluation phase, each trainee independently assessed the mRSS in four SSc patients. For intra-observer reliability, 14 trainees re-assessed the mRSS of two SSc patients whom they had previously examined. We computed the inter- and intra-observer variability using a linear mixed model.

**Results:** For the evaluation phase, 34 (65.4%) trainees were within five units of the established teachers' score in 3 out of 4 patients. Overall, the whole group had acceptable inter-observer variability (intra-class correlation coefficient [ICC] = 0.71, mean = 8.64 and within-patient standard deviation [SD] = 4.25). The intra-observer ICC was 0.85 and within-patient SD was 2.73.

**Conclusion:** There was good inter-observer and excellent intra-observer reliability. This is the first study examining the training of assessors using the SCTC/WSF recommendations and our results support the importance of standardized training for skin scoring.

**KEYWORDS**
modified Rodnan skin score, standardization, systemic sclerosis, training

# 1 | INTRODUCTION

Systemic sclerosis (SSc) is characterized by skin tightening and thickening. Based on the extent of skin involvement, SSc is sub-classified into limited cutaneous SSc (lcSSc) with skin thickening distal to the elbows and knees with or without face involvement, or diffuse cutaneous SSc (dcSSc) with skin thickening proximal as well as distal to the elbows and knees, and can occur with or without involvement of the face.[1] In particular, early in the disease, more extensive skin involvement is associated with more severe internal organ manifestations and poorer prognosis.[2,3] The modified Rodnan skin score (mRSS)[4] is a validated outcome measure for skin thickness in SSc clinical trials. It is recommended that the same assessor examine the patient for the duration of the trial as each outcome measure inherently has measurement variability.[5] Training assessors has been shown to reduce variability in mRSS assessment.[5]

The aim of this study was to assess the effectiveness of a systematic training workshop on standardization of mRSS in SSc as per the recommendations proposed by the Scleroderma Clinical Trials Consortium (SCTC)/World Scleroderma Foundation (WSF) (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5431585/pdf/nihms849080.pdf).

# 2 | METHODS

We assessed the inter- and intra-observer variability of mRSS in a group of trainees who participated in the skin score workshop conducted as part of the 18th Singapore Society of Rheumatology-Malaysian Society of Rheumatology pre-congress workshops in Singapore.

## 2.1 | Participants

Two SSc experts (DK, AL) and two facilitators (GC, NSC), 52 trainees and eight adult SSc patients of Asian descent fulfilling the 2013 American College of Rheumatology/European League Against Rheumatism criteria participated in a SSc skin scoring workshop. DK is an expert in SSc who co-authored the manuscript on the standardization of mRSS[5] and is a member of the SCTC. AL is a local SSc expert with more than 500 SSc patient visits per year, a member of the SCTC, and has been trained and facilitated in two prior SSc workshops in Singapore by Dr D. Furst (2009) and Dr C. Denton (2014). The two facilitators are rheumatologists who have participated in the Singapore Scleroderma Research Workgroup since 2009 and they assisted with the conduct of the workshop. The 52 trainees included physicians and research coordinators. Of the eight SSc patients, four had lcSSc and four had dcSSc (see Table 1 for their mRSS scores).

## 2.2 | Skin scoring

The mRSS is calculated by summation of measurements of skin thickness in 17 different body sites including the face, upper arms, forearms, dorsum of hands, fingers, chest, abdomen, thighs, legs and feet.

**TABLE 1** Modified Rodnan skin scores (mRSS) of 8 systemic sclerosis patients

| | mRSS | Mean (SD) mRSS in each group |
|---|---|---|
| Group A | | |
| Patient 1 | 14 | |
| Patient 2 | 0 | 6 (5.9) |
| Patient 3 | 4 | |
| Patient 4 | 6 | |
| Group B | | |
| Patient 5 | 3 | |
| Patient 6 | 6 | 5 (4.7) |
| Patient 7 | 11 | |
| Patient 8 | 0 | |

The maximum total score is 51 and each area is graded as follows: 0 = normal skin, 1 = mild skin thickening, 2 = moderate skin thickening with difficulty in making skin folds and no wrinkles, 3 = severe skin thickness with inability to make skin folds between two examining fingers.[5] Trainees were taught that the three commonly used techniques for mRSS skin scoring are that of global average, maximum score and representative area. For purposes of standardization in this workshop, the global average method was used.

## 2.3 | Conduct of the skin scoring workshop

Similar to how other workshops are conducted,[6] the experts first evaluated the eight patients together with the facilitators to establish the standard teachers' mRSS score. The workshop was conducted in two sessions, with each session comprising of a teaching phase and an evaluation phase (Figure 1). The duration of each session was approximately 2 hours and included 26 trainees per session.

The teaching phase comprised of a: (a) lecture on mRSS skin scoring by an SSc expert (DK); (b) video demonstration of mRSS skin scoring by Dr D. Furst, an SSc expert, examining a patient exhibiting different aspects of skin thickness that corresponds to the expected values of the mRSS using the global average method; and (c) live demonstration of mRSS scoring on one of the eight SSc patients by DK or AL.

For the evaluation phase, the trainees were divided into two groups (Groups A and B), with each group evaluating four patients. Trainees were given a mRSS sheet as per the SCTC/WSF recommendations and approximately 5 minutes to evaluate one patient. Discussion was not allowed during the evaluation. Feedback was provided to the trainees at the end of the session. The mRSS scoring for each trainee was then compared to the established teachers' mRSS score: a score difference of ≤5 in 3 out of 4 subjects scored by the trainee was considered acceptable inter-observer variability.[5] Trainees who achieved this passed the evaluation phase of the workshop and obtained mRSS scoring certification for the conduct of clinical trials.

To assess intra-observer variability, 14 randomly selected trainees and two experts re-assessed the mRSS of the same SSc patients
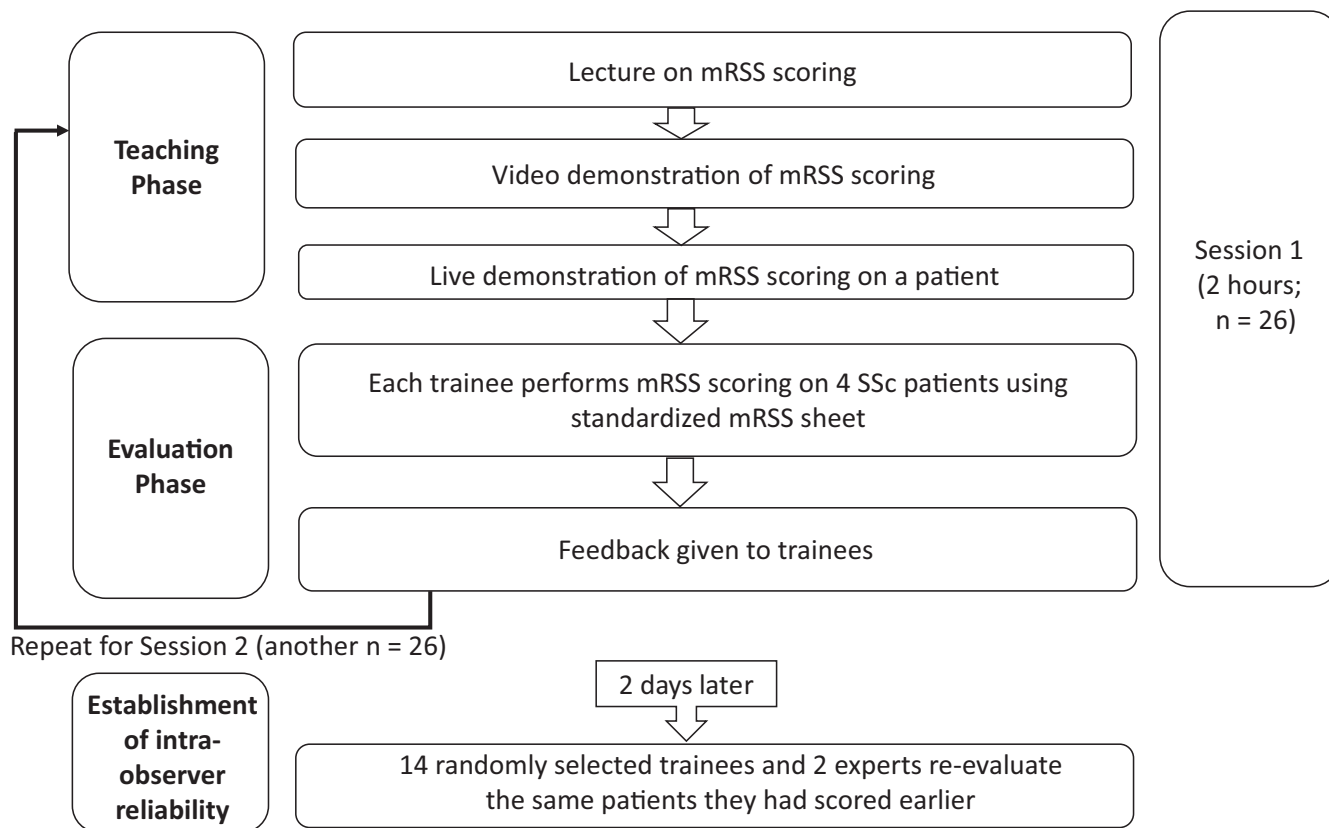
**FIGURE 1** Flow diagram summarizing the conduct of skin score workshop. mRSS, modified Rodnan skin score; SSc, systemic sclerosis

(two each for trainees and four each for experts) whom they had examined 2 days ago. The repeat mRSS score for each trainee was compared to their original scores. A score of ≤3 was considered acceptable intra-observer variability, as supported by previous studies with an intra-observer variability of 2.5-2.9.[6]

## 2.4 | Statistical analysis

We computed the inter- and intra-observer variability using a linear mixed model with an intercept term, a random effect for the patient, a random effect for the scorer, and a random effect for the interaction of patient and scorer, as previously described.[7] Inter-observer variability was calculated for the 52 trainees in the evaluation phase and intra-observer variability was calculated for the 14 trainees and two experts who returned 2 days later for repeat scoring. Agreement among scorers was quantified via the intra-class correlation coefficient (ICC). Usual interpretations of the ICC are as follows: values of 0.4-0.6 are considered moderate; 0.6-0.8 are deemed good and 0.8-1.0 are considered excellent agreement. Summary statistics (mean and within-patient standard deviation, SD) were calculated for the skin scores along with the coefficient of variation (SD/mean). Statistical analysis was performed using the R statistical software, version 3.4.2 (R Foundation, Vienna, Austria). This study was approved by the Singhealth Centralised Institutional Review Board, with waiver of consent obtained.

## 3 | RESULTS

Participants of the workshop came from five different Asian countries (Singapore, Malaysia, Indonesia, Myanmar, Qatar), and were rheumatologists, rheumatology trainees, research coordinators and dermatologists.

For the evaluation phase, 34 of 52 trainees (65.4%) achieved acceptable inter-observer variability and were within five units of the established teachers' score (median [range] mRSS 5 [0-14] by teachers vs 8.6 [0-34] by trainees). The inter-observer variability ICC was 0.71 and the within-patient SD was 4.25 units. The two experts (DK/AL) were within two units in mRSS scores for each of the eight SSc patients.

The intra-observer variability for the 14 trainees and two experts 2 days later showed an acceptable within-patient SD of 2.56 (see Table 2). The test-retest score was within one unit for the experts. The coefficient of variation was 39%.

## 4 | DISCUSSION

The mRSS is a feasible, reliable, valid, and responsive measure and is used as a primary or secondary outcome measure in clinical trials.[8] In a Phase 2 trial,[9] mRSS was able to differentiate tocilizumab from placebo. Mycophenolate mofetil and cyclophosphamide were both shown to be superior to placebo in post-hoc analysis of the

**TABLE 2** Inter- and intra-observer variability for skin score workshop

| Evaluation phase | No. of subjects | Inter-observer variability | | | |
| | | Mean | Within-patient SD | Coefficient of variation % | Intra-class correlation coefficient |
| --- | --- | --- | --- | --- | --- |
| Trainees (n = 52) | 4 | 8.64 | 4.25 | 49 | 0.71 |
| **Establishment of intra-observer reliability** | **No. of subjects** | **Intra-observer variability** | | | |
| | | **Mean** | **Within-patient SD** | **Coefficient of variation %** | **Intra-class correlation coefficient** |
| Trainees (n = 14) and experts (n = 2) | 2 | 6.56 | 2.56 | 39 | 0.86 |
| Trainees only (n = 14) | 2 | 6.92 | 2.73 | 39 | 0.85 |

Scleroderma Lung Study I and II[10] using mRSS as the primary outcome measure. It is important that in the modern era of SSc clinical trials, outcome measures should be valid, reliable and responsive to change.[11] With new insights on the pathogenesis of SSc, there are several targeted novel therapies that are being assessed for skin and interstitial lung disease. Due to the orphan nature of the disease, the trials are being conducted in different countries to recruit appropriate patients. Multi-center recruitment across different regions provides more robust and generalizable results as it accounts for variability in different regions, such as the well-recognized differences in autoantibodies and severity of skin thickness by regions. However, this may result in higher inter-observer variability in the primary outcome measure, the mRSS. One way to reduce this variability is to protocolize the way patients are evaluated to reduce variability between assessors due to subtle differences in assessing the mRSS and also different methods utilized in scoring mRSS (global average vs maximum score vs representative area). The global average and representative area techniques are recommended as they are likely more sensitive to change than the maximum score technique.[5] The intra-observer variability has been shown consistently by various studies to be lower than inter-observer variability.[6,7,12] It is therefore recommended that the same assessor evaluates the patient throughout a trial to reduce measurement variability. Training of practitioners in mRSS evaluation will have the effect of reducing the variability and provides a platform to standardize mRSS for a trial.[5] Although the teaching and standardization has been performed in different trials (Dr D.K. Khanna, personal communication), to our knowledge, this is the first study to prospectively incorporate the SCTC/WSF recommendations in a group of general rheumatologists and research coordinators. We determined that training workshops are feasible, and we found good inter-observer reliability and excellent intra-observer reliability for mRSS scoring in a group of trainees who underwent a structured and standardized mRSS workshop. This study is an important contribution to the field as it provides evidence of intra-rater and inter-rater reliability among both physicians and research staff.

The inter-observer SD was noticed to be 4.3 units and was consistent with previous studies[6] (within-patient SD 3.8-8.5), and was well within the recommended SCTC/WSF proposed upper limit of five units. Likewise, the intra-observer within-patient SD was consistent with previously reported figures of 2.5-2.9.

Few studies have been conducted to investigate intra-observer variability of mRSS scoring. We report intra-observer variability on the same group of patients examined 2 days after the initial training session. This would have reduced recall bias compared to the study by Gordon et al, where patients were re-examined on the same day with an intra-observer variability of 0.94.[7] In the studies by Clements et al[12] and Czirjak et al[6] patients were re-examined 2-8 weeks later to quantitate intra-observer variability and the intra-observer within-patient SD was found to be 2.5-2.9.

The coefficient of variation (CV) was on the higher end of the range in our study (39%-56%) compared to other studies (11.8%-54%)[6,7,12] despite acceptable within-patient SD for both inter- and intra-observer variability. This could be due to the lower median skin score of five in our patients compared to mean scores of 8.6-20.7 reported in other studies, resulting in a larger CV for minimal variation in SD (since CV is derived from SD divided by the mean).

The lower ICC when experts were included in the analysis for intra-observer variability likely reflected the tendency for most trainees to score higher than the experts, an observation similarly noted in other studies,[6] where skin tethering may erroneously be scored as skin thickness.

Our data shows that 65.4% of the trainees were provided certificates, based on the predefined variability of up to five units. We do not have published data from other studies on proportion of those who passed the evaluation phase but this is likely lower due to the mix of the trainees. In the opinion of the senior author (DK), the proportion is higher in those who perform skin scoring training on a regular basis in clinics or those who have participated in previous clinical trials. To address over-scoring of skin tethering, training workshops can include patients with higher mRSS scores (ie reflective of how patients are enriched for clinical trials) as well as patients with atrophic and tethered skin so as to emphasize their difference, to further improve training.

We believe our results are especially encouraging as the trainees were of a heterogeneous background ranging from experienced rheumatologists to rheumatology trainees to research coordinators with no medical training, many of whom are not scleroderma experts. Further, the findings from our study may be generalized to

patients of Asian descent. Previous studies demonstrated no significant differences in initial or peak mRSS between SSc patients of Asian descent and their European counterparts.[13,14]

Our study has several limitations. Ideally the repeat scoring to evaluate intra-observer variability could have been done 2-4 weeks later to further minimize recall bias which may occur within a 2-day interval. This has to be balanced against a longer interval (eg more than 3 months) when any change in skin scores may be due to disease improvement or worsening. The mean skin scores of patients who participated in the workshop were lower than other studies with a range of mRSS from 0 to 14. This limitation could be addressed in future training workshops conducted in Asian patients by including patients with higher mRSS scores. Pre-workshop assessment of the trainees could also have been conducted so as to compare trainees' performances pre- and post-workshop to further support the SCTC/WSF training recommendations.

In summary, this is the first study examining the training of assessors using the SCTC/WSF training recommendations demonstrating good and excellent inter- and intra-observer variability for skin scoring. Standardized training of skin scoring is strongly encouraged to enable reliable mRSS scoring, a key outcome measure in SSc clinical trials.

## CONFLICT OF INTEREST

The authors declare no conflicts of interest related to this study.

## ORCID

*Sue-Ann Ng* 🆔 https://orcid.org/0000-0003-3732-5725

## REFERENCES

1. LeRoy EC, Medsger TA Jr. Criteria for the classification of early systemic sclerosis. *J Rheumatol*. 2001;28(7):1573-1576.
2. Clements PJ, Hurwitz EL, Wong WK, 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(11):2445-2454.
3. Shand L, Lunt M, Nihtyanova S, et al. Relationship between change in skin score and disease outcome in diffuse cutaneous systemic sclerosis: application of a latent linear trajectory model. *Arthritis Rheum*. 2007;56(7):2422-2431.
4. Clements PJ. Measuring disease activity and severity in scleroderma. *Curr Opin Rheumatol*. 1995;7(6):517-521.
5. Khanna D, Furst DE, Clements PJ, et al. Standardization of the modified Rodnan skin score for use in clinical trials of systemic sclerosis. *J Scleroderma Relat Disord*. 2017;2(1):11-18.
6. Czirjak L, Nagy Z, Aringer M, Riemekasten G, Matucci-Cerinic M, Furst DE. The EUSTAR model for teaching and implementing the modified Rodnan skin score in systemic sclerosis. *Ann Rheum Dis*. 2007;66(7):966-969.
7. Gordon JK, Girish G, Berrocal VJ, et al. Reliability and validity of the tender and swollen joint counts and the modified Rodnan skin score in early diffuse cutaneous systemic sclerosis: analysis from the prospective registry of early systemic sclerosis cohort. *J Rheumatol*. 2017;44(6):791-794.
8. Khanna D, Merkel PA. Outcome measures in systemic sclerosis: an update on instruments and current research. *Curr Rheumatol Rep*. 2007;9(2):151-157.
9. Khanna D, Denton CP, Jahreis A, et al. Safety and efficacy of subcutaneous tocilizumab in adults with systemic sclerosis (faSScinate): a phase 2, randomised, controlled trial. *Lancet*. 2016;387(10038):2630-2640.
10. Namas R, Tashkin DP, Furst DE, et al. Efficacy of mycophenolate mofetil and oral cyclophosphamide on skin thickness: post hoc analyses from two randomized placebo-controlled trials. *Arthritis Care Res (Hoboken)*. 2018;70(3):439–444.
11. Johnson SR, Khanna D, Allanore Y, Matucci-Cerinic M, Furst DE. Systemic sclerosis trial design moving forward. *J Scleroderma Relat Disord*. 2016;1(2):177-180.
12. Clements P, Lachenbruch P, Siebold J, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol*. 1995;22(7):1281-1285.
13. Low AH, Johnson SR, Lee P. Ethnic influence on disease manifestations and autoantibodies in Chinese-descent patients with systemic sclerosis. *J Rheumatol*. 2009;36(4):787-793.
14. Proudman SM, Huq M, Stevens W, et al. What have multicentre registries across the world taught us about the disease features of systemic sclerosis? *JSRD*. 2017;2(3):169-182.