Arthritis Care & Research
Vol. 75, No. 8, August 2023, pp 1725–1734
DOI 10.1002/acr.25038

American College of Rheumatology
Empowering Rheumatology Professionals

Assessment of the Systemic Sclerosis–Associated Raynaud's Phenomenon Questionnaire: Item Bank and Short-Form Development

Lan Yu,[1] Robyn T. Domsic,[1] Lesley-Ann Saketkoo,[2] Jane Withey,[3] Tracy M. Frech,[4] Ariane L. Herrick,[5] Laura K. Hummers,[6] Ami A. Shah,[6] Christopher P. Denton,[7] Dinesh Khanna,[8] and John D. Pauling[9]

Objective. To develop, refine, and score a novel patient-reported outcome instrument to assess the severity and impact of Raynaud's phenomenon (RP) in systemic sclerosis (SSc).

Methods. The Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon (ASRAP) questionnaire items were developed with patient insight partner support and grounded in the lived patient experience of SSc-RP. ASRAP items underwent formal qualitative assessment and linguistic testing. An international multicenter study was undertaken to field test the preliminary ASRAP questionnaire.

Results. A preliminary 37-item ASRAP questionnaire was supplemented with 2 additional items following expert review to enhance content coverage before undergoing formal linguistic testing to optimize readability. Patient cognitive debriefing interviews were undertaken to enhance comprehension, ambiguity, cognitive difficulty, relevance, and content coverage of both the ASRAP items and instructions. We enrolled 420 SSc patients from scleroderma centers in the UK and US over 2 consecutive winters. Factor analysis with item response theory was undertaken to remove redundant and poorly fitting items. The retained 27-item long-form ASRAP questionnaire was calibrated and scored using the graded response model. A fixed 10-item short-form ASRAP questionnaire was developed using computerized adaptive testing simulations.

Conclusion. The ASRAP questionnaire has been developed with extensive SSc patient input, with items grounded in the lived experience of SSc-RP to ensure strong content validity, with a focus on how patients feel and function. An advanced psychometric approach with expert input has removed redundant and/or poorly fitting items without eroding content validity. Long- and short-form ASRAP questionnaires have been calibrated and scored to permit formal validation.

INTRODUCTION

Raynaud's phenomenon (RP) is the commonest disease manifestation of systemic sclerosis (SSc) (1). SSc-RP is a major cause of disease-related morbidity and ranked highly by patients in terms of severity and impact (2–4). SSc-RP symptoms include ischemic pain, sensory impairment (numbness/tingling/burning), and impaired finger function. The unpleasant physical symptoms of SSc-RP lead to emotional distress, often aggravated by feelings of body image dissatisfaction and embarrassment related to marked digital discoloration (2,3). Individuals living with SSc adapt to avoid and ameliorate symptoms of RP, although the need to avoid cold exposure and stressful interactions itself impacts social

Supported by the US Department of Defense (grant W81XWH-18-1-0602) and the Scleroderma Clinical Trials Consortium (grant SCTC RFA 2018).

[1]Lan Yu, PhD, Robyn T. Domsic, MD, MPH: University of Pittsburgh, Pittsburgh, Pennsylvania; [2]Lesley-Ann Saketkoo, MD, MPH: University of Tulane Medical Center, New Orleans, Louisiana; [3]Jane Withey: Royal National Hospital for Rheumatic Diseases, Royal United Hospitals NHS Foundation Trust, Bath, UK; [4]Tracy M. Frech, MD, MS: Vanderbilt University, Nashville, Tennessee; [5]Ariane L. Herrick, MD, FRCP: The University of Manchester, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Science Centre, and NIHR Manchester Biomedical Research Centre, Central Manchester NHS Foundation Trust, Manchester, UK, and Salford Royal NHS Foundation Trust, Salford, UK; [6]Laura K. Hummers, MD, MSc, Ami A. Shah, MD, MHS: Johns Hopkins University School of Medicine, Baltimore, Maryland; [7]Christopher P. Denton, PhD, FRCP: Royal Free Hospital, London, UK; [8]Dinesh Khanna, MD, MSc: University of Michigan, Ann Arbor; [9]John D. Pauling, BMedSci, BMBS, PhD, FRCP: Royal National Hospital for Rheumatic Diseases, Royal United Hospitals NHS Foundation Trust, Bath, UK, and North Bristol NHS Trust and University of Bristol, Bristol, UK.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25038&file=acr25038-sup-0001-Disclosureform.pdf.

Address correspondence via email to John D. Pauling, BMedSci, BMBS, PhD, FRCP, at JohnPauling@nhs.net.

Submitted for publication June 24, 2022; accepted in revised form October 4, 2022.

SIGNIFICANCE & INNOVATIONS

- We developed a novel patient-reported outcome instrument, the Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon (ASRAP) questionnaire, grounded in the lived patient experience of Raynaud's phenomenon in systemic sclerosis.
- An extensive target patient population was involved in the development of the ASRAP questionnaire item bank.
- We used a data-driven approach to remove redundant and/or poorly fitting items from the ASRAP questionnaire.
- This study involved the development, calibration, and scoring of both the ASRAP questionnaire and the 10-item short-form ASRAP questionnaire.

and work participation, often necessitating the support of others (2,3).

The episodic and uniquely personalized experience of SSc-RP has led to a reliance upon patient-reported outcome instruments to capture how patients feel and function with respect to SSc-RP. For over 20 years, the majority of clinical trials in SSc-RP have utilized diary-based approaches to capture the frequency and aggregate duration of SSc-RP attacks, alongside a daily assessment of the Raynaud's Condition Score (RCS), an 11-point numeric rating scale or 100-mm visual analog scale assessing the overall severity and impact of RP symptoms (5–7). A large number of therapeutic interventions have been tested in clinical trials of SSc-RP, but establishing treatment efficacy using existing clinical trial end points has been challenging, and there are currently no medications approved by the Food and Drug Administration for SSc-RP (8). Meta-analyses have indicated that the net benefit of treatments such as calcium-channel blockers and phosphodiesterase 5 inhibitors on existing clinical trial end points are either absent or modest at best (9,10).

SSc patients were not involved in the development of the 3 parameters of the RCS diary comprising mean daily frequency of SSc-RP attacks, mean daily aggregate duration of SSc-RP attacks, and the mean daily RCS over 7–14 days of monitoring (5–7,11). Concerns have been raised by SSc patients and experts about the RCS diary, the focus of which is RP attack frequency/duration and does not take into consideration the significant efforts adopted by patients to avoid and ameliorate attacks or the evolution of RP symptoms that may accompany progression of the obliterative microangiopathy of SSc (11–14). A number of important patient experiences of SSc-RP, such as the emotional impact of SSc-RP, are not fully captured by RCS diary parameters (3). Against this backdrop, a multicenter collaborative effort has been undertaken to develop a new patient-reported outcome instrument derived from patient experience. Here we report the development, refinement, and scoring of the Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon (ASRAP) questionnaire, a novel patient-reported outcome instrument for capturing the impact and severity of SSc-RP.

MATERIALS AND METHODS

Conceptual framework and development of the item pool. The ASRAP questionnaire was developed with oversight from the Scleroderma Clinical Trials Consortium (SCTC) Vascular Working Group. The agreed conceptual framework was to devise a novel patient-reported outcome instrument that captured the severity and impact of SSc-RP grounded in the patient experience of SSc-RP (Figure 1).

A preliminary item bank of 37 candidate items was devised with support from a patient insight partner member of the steering committee who has lived experience of SSc (JW) to capture the themes and subthemes comprising the lived experience of SSc-RP, identified in an earlier international multicenter qualitative research study of SSc-RP and a comprehensive scoping review (Figure 1) (2,3). The ASRAP items were devised to capture domains comprising physical symptoms (n = 10), emotional distress (n = 8), impact on daily life (n = 6), exacerbating factors (n = 5), self-management (n = 4), adaptation (n = 2), and uncertainty (n = 2) (2,3). Where possible, the language used in quotations from patients in the earlier qualitative work was used in item wording. Item response options were designed to conform with standards developed for the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative with response levels best suited for item response theory (IRT) modeling (15–17). A recall time frame of the past 7 days was agreed among SSc experts and patients to optimize feasibility and utility in a clinical trial setting.

Qualitative item review. Iterative modification of the ASRAP instructions and preliminary 37 items was undertaken with input from 4 SSc experts (RTD, LAS, DK, and JDP) to ensure that item wording, recall period, and response options were simple, understandable, and relevant to specific domain concepts, and that they conformed to internationally agreed standards (15–17). Two additional items were proposed following expert review to enhance content coverage (one concerning average duration of a typical RP attack, and another concerning the emotional impact of SSc-RP in limiting usual activity), resulting in a 39-item preliminary ASRAP questionnaire (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25038). Linguistic evaluation of the 39-item preliminary ASRAP questionnaire was undertaken using the Simple Measure of Gobbledygook (SMOG) to ensure that the ASRAP items were concise and simply worded (to achieve a readability age of <14 years). SMOG assessment led to item modification of 11 items (see Supplementary Appendix B, available on the

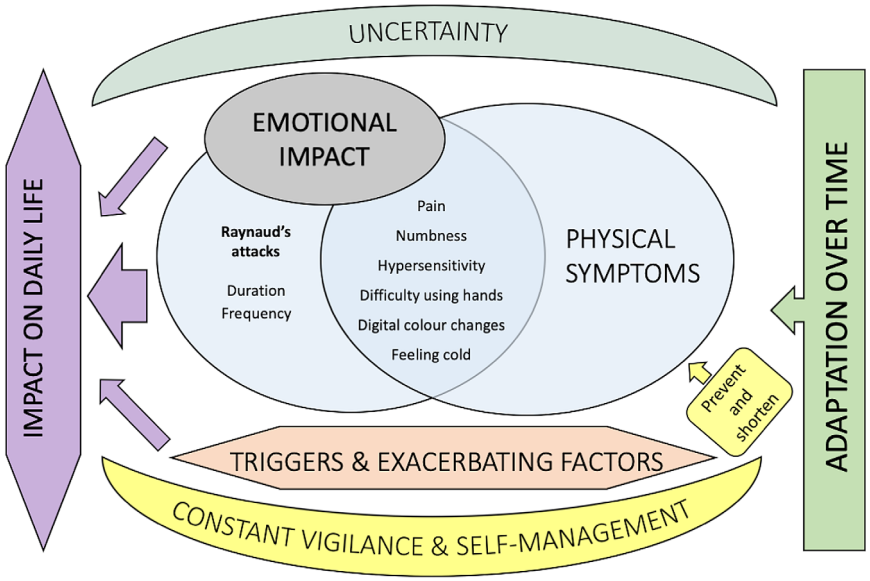

Figure 1. Conceptual map of the major themes and subthemes comprising the patient experience of Raynaud's phenomenon in systemic sclerosis. Reproduced from ref. 3.

Arthritis Care & Research website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25038). Data supporting the development of the ASRAP questionnaire can be made available upon reasonable request.

Cognitive debriefing interviews were held with English-speaking SSc patients (n = 7) in both the US and UK to evaluate patient perceptions regarding the language, comprehensibility, ambiguity, cognitive difficulty, relevance, and content coverage of the items (17). The interviews lasted ~90 minutes with a scheduled comfort break, and 3 subjects began the debriefing interview on item 20 to ensure that interview fatigue did not impact responses. Cognitive debriefing also sought to ensure that the ASRAP items met accepted criteria for optimal translatability into non-English languages. A scripted interview, incorporating standard probes, was devised to elicit patient feedback on comprehension, memory retrieval process, item structure (stem, recall period, and response options), response processing, and overall ASRAP format (content coverage and length of questionnaire), with an opportunity to propose new items if necessary. Interviews were audio recorded and transcribed for future verification. Cognitive debriefing interviews led to modification of 15 items and changes to format and structure of the ASRAP questionnaire instructions (Table 1). No new items were proposed. An item tracking matrix was devised to document modifications during item development (see Supplementary Appendix B, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25038). The ASRAP questionnaire was then deemed ready for field testing.

Cross-sectional calibration study. The international multicenter ASRAP cross-sectional calibration study enrolled English-speaking SSc patients from 7 UK (Bath, London, and Manchester) and US (Pittsburgh, Baltimore, Ann Arbor, and Salt Lake City) scleroderma centers over 7 winter months (February to March 2019 and November 2019 to March 2020). To provide an expected adequate sample for robust unidimensional graded response model (GRM) analysis, our target sample size was 500 (16). The study had research ethics committee approval at each UK and US site (information available on request) and was conducted in accordance with the principles of the Declaration of Helsinki.

Eligibility and study procedures. All patients were age ≥18 years, fulfilled the American College of Rheumatology/EULAR classification criteria for SSc (18), and had good comprehension of written/spoken English. Pregnant women and/or subjects whose vasodilator medication had not been stable within the previous 4 weeks were excluded. To ensure that our cohort was reflective of real-life practice, we permitted background use of vasodilators (whether for SSc-RP or other cardiovascular disease) to be continued at a stable dose throughout the study. All patients provided informed written consent. All participants completed the provisional 39-item ASRAP questionnaire. Relevant patient demographic information and clinical phenotypes were captured using a clinician case-report form.

Statistical analysis. *Factor analysis.* A 2-step analytical approach was undertaken. The initial analyses involved descriptive statistics of each ASRAP item. Although our ASRAP questionnaire items were designed to capture 7 domains deemed relevant to patients' SSc-RP experience, we made no assumptions regarding the most appropriate factor structure for the

Table 1. Summary of modifications made to ASRAP items following formal cognitive debriefing interviews*

Item number or general change	Original item wording/formatting	New item wording/format change
Strengthen focus on 7-day recall period by changing wording of instructions at the start of each section	Considering your Raynaud's attacks over the last 7 days (presented in lower case bold)	In the past 7 days (shortened with "past 7 days" presented in capital letters, in bold, and underlined using largest font possible)
9	How often have you experienced attacks of Raynaud's symptoms?	On average, how often have you experienced attacks of Raynaud's symptoms?
10	On average, how much total time have you spent each day experiencing attacks of Raynaud's symptoms?	On average, how much total time per day have you experienced attacks of Raynaud's symptoms?
11	Response option 4: "10–20 minutes"; response option 5: "over 30 minutes"	Change response option 4 to "10–25 minutes"; change response option 5 to "over 25 minutes"
12	Raynaud's symptoms have made me upset/tearful	Raynaud's symptoms have made me tearful
14	Raynaud's symptoms have made me annoyed/frustrated	Raynaud's symptoms have made me frustrated
18	Raynaud's symptoms have caused low mood/made me depressed	Raynaud's symptoms have made me sad/depressed
23	Raynaud's symptoms have made social events/doing sport difficult	Raynaud's symptoms have made social events/doing exercise difficult
27	I have been able to reduce (control) the intensity my Raynaud's symptoms?	I have been able to reduce (control) the intensity of my Raynaud's symptoms?
28-32	"This activity not undertaken" currently on far left of response options	Move "This activity not undertaken" response option column to the far right
28	Going inside a grocery store/supermarket has caused Raynaud's symptoms	Being inside a grocery store/supermarket has caused Raynaud's symptoms
31 and 32	Inadvertent consideration of being indoors by one subject	Switch positions of items 31 and 32
34	I have used hand warmers/put my hands in warm water to control/manage Raynaud's symptoms	I have used techniques (e.g., hand warmers/putting hands in warm water/sitting on hands) to control/manage Raynaud's symptoms
39	Changes to my normal routine have caused me to worry about possible worsening of Raynaud's symptoms	A change in my normal routine has caused me to worry about possible worsening of my Raynaud's symptoms

* ASRAP = Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon.

39 ASRAP items. It was not assumed the 39 ASRAP items would form a single underlying dimension that covered the broad-ranging patient experience of SSc-RP. Therefore, our goal was to identify the best performing items representing the robust underlying traits and to document sufficient unidimensionality to allow us to proceed with IRT analyses.

First, we inspected frequency distributions of individual items for sparse response frequency. We then investigated dimensionality by using the statistical software to divide the sample randomly into 2 similarly sized subsamples: one for exploratory factor analysis (EFA; unweighted least squares), and the second for confirmatory factor analysis (CFA). Both EFA and CFA were conducted using Mplus 6.2 with promax rotation (19). In the CFA, the items were treated as categorical variables, and the robust weighted least squares estimator was used. Scree plots, eigenvalues, and factor loadings were examined. We examined the ratio of eigenvalues in EFA and the relative proportion of variance accounted for by the factors extracted. We also carefully assessed the size of factor loadings in both EFA and CFA and the information values for individual items from the IRT models. Items with low factor loadings of <0.5 and/or low item discrimination parameter estimates ($\alpha < 1.0$) were considered for removal.

IRT analysis. The most commonly used IRT model for polytomous items (i.e., items with ≥3 ordinal response categories) is the 2-parameter GRM (20). The GRM has a slope parameter and n – 1 threshold parameters for each item, where n is the number of response categories (5 in the present analyses). The slope parameter measures item discrimination, i.e., how well the item differentiates between higher versus lower levels of severity (or Θ in IRT terms). Useful items have large slope parameters. Threshold parameters measure item difficulty, i.e., the ease versus difficulty of endorsing different response options for an item. For example, the first threshold parameter for an item tells us where along the Θ scale of impact and severity of Raynaud's symptoms a respondent is more likely to endorse one response option over another, e.g., "a little bit" rather than "not at all." Items were calibrated using IRTPRO, version 2.1 (21). Local dependency (LD) marginal chi-square analyses identified redundant items due to high LD (residual correlations) with other items, and the items were removed (20,21).

Differential item functioning (DIF) occurs when characteristics such as age, sex, or ethnicity, which may seem extraneous to the assessment of the construct under consideration, actually do affect the measurement of the construct. An item functions

differentially if the item is more (or less) discriminating or more (or less) difficult to endorse in one group compared with a reference group (e.g., women versus men) when the different subgroups have been matched on the latent variable under investigation. We conducted DIF analyses for both uniform (difficulty) and nonuniform (discrimination) DIF on the basis of age (median split, age ≤60 years versus age >60 years), birth year (odd versus even birth year), and location (UK versus US). We focused on these variables because the relevant comparison groups were adequately represented. The DIF procedure, the IRT likelihood ratio method, was used, and items were considered for removal if they showed poor DIF ($P < 0.01$) (22).

Fixed short-form development and ASRAP scoring. When computerized adaptive testing (CAT) is not available, a static short form is a useful tool. A short-form version of the ASRAP (ASRAP-SF) questionnaire was also desirable for use in clinical practice or as a secondary end point in SSc clinical trials. After the ASRAP items were calibrated using IRT, we developed fixed ASRAP-SF items based on CAT simulations. We used 4 criteria to rank order ASRAP items: discrimination parameters; the percentage of times the item would have been selected in a simulated CAT using the calibration sample; expected information under the standard normal distribution with a mean ± SD of 0 ± 1; and expected information under a normal distribution with a larger SD (i.e., a mean ± SD of 0 ± 1.5). The CAT simulations were performed using the Firestar program (23).

Once the final ASRAP items were calibrated, each participant was scored (i.e., Θ scores derived from IRT calibration) for both the full bank and the fixed ASRAP-SF. Lookup tables have been created to build the connection between the raw scores and the corresponding Θ scores for both the ASRAP and ASRAP-SF questionnaires. Internal consistency between the ASRAP and ASRAP-SF questionnaires was evaluated using Pearson's correlation coefficient and Cronbach's alpha.

To aid easier conceptualization of ASRAP scores, Θ values (theoretically ranging from –3.0 to +3.0 SD from 0) were recalibrated by multiplying the Θ score by 10 and adding 50 to achieve a scoreable ASRAP range of 20–80.

RESULTS

Study population. A total of 420 SSc subjects were enrolled at UK (n = 222) and US (n = 198) sites and completed at least 1 ASRAP questionnaire. Full ASRAP questionnaire data were available for 404 subjects (96.0%), indicating strong feasibility within the target patient population. Enrollment was suspended in early March 2020 before achieving our planned sample size due to the impact of COVID-19 on clinical services and the perceived potential impact of lockdown restrictions of movement on cold exposure and SSc-RP severity. Our subsequent analyses confirmed that the sample size was sufficient to undertake our planned objectives.

The clinical phenotype of the enrolled study population conformed to expected distributions with respect to sex (79.7% female), ethnicity (86.4% White), disease subset (58.9%, 34.9%, and 6.2% for limited, diffuse, and sine scleroderma, respectively), age (mean ± SD 58.9 ± 12.4 years), disease duration (mean ± SD 15.1 ± 12.1 years since RP onset and 11.5 ± 9.6 since first non-RP symptom of SSc), clinical features, and autoantibody specificities; confirming this was a representative cohort of SSc patients. There were only 18 (4.5%) current smokers, with 150 (37.1%) past smokers.

Factor analysis. Prior to factor analysis, we examined the frequencies of each of the 39 ASRAP items for sparse response frequency. There were no items with response categories having <5 observations, enabling all 39 items to be included in the factor analysis (and suggesting no significant floor or ceiling effects for individual items). The sample was split randomly for EFA (n = 191) and CFA (n = 213). The first 2 eigenvalues were 21.07 and 3.23, with a ratio of 6.52, which is a strong indication of a single underlying trait. The 2-factor solution appeared to distinguish domains of SSc-RP with respect to physical symptoms and emotional distress. For EFA, we compared the 1-factor and 2-factor solutions, and the 1-factor solution provided a better fitting and a clinically meaningful construct. The majority of items had a factor loading of >0.450. Three items (items 30, 31, and 33) had factor loadings between 0.399 and 0.418 and were removed from further analysis. We then used the CFA sample to confirm the 36-item 1-factor structure identified in EFA. The factor loadings on all 36 items were above 0.54, and the fit indices were all excellent or acceptable, with a standardized root mean square residual of 0.08 and a confirmatory factor index of 0.92, except for a root mean square error of approximation of 0.11 (just above the desired range of 0.08).

IRT calibrations. The 36 items retained from factor analysis were calibrated using an IRT GRM. Item discrimination parameters were generally satisfactory, with α ranging from 1.21 to 3.66 (Figure 2, with $\alpha > 1.0$ considered satisfactory), with the exception of 1 item (item 34, pertaining to the use of hand-warming devices), with a marginal item discrimination parameter estimate of 1.05, which was removed at this stage.

Redundant items. Only 14 potentially redundant pairs showed high local dependency (residual correlations) with other items. Using LD marginal chi-square testing (>10 indicates redundancy), 6 individual items (items 2, 5, 10, 25, 29, and 38) were removed (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25038). When choosing the removal of redundant items, consideration was also given to whether items had been identified as problematic in other aspects of the analysis. For example, items 2 and 5 were identified as problematic in

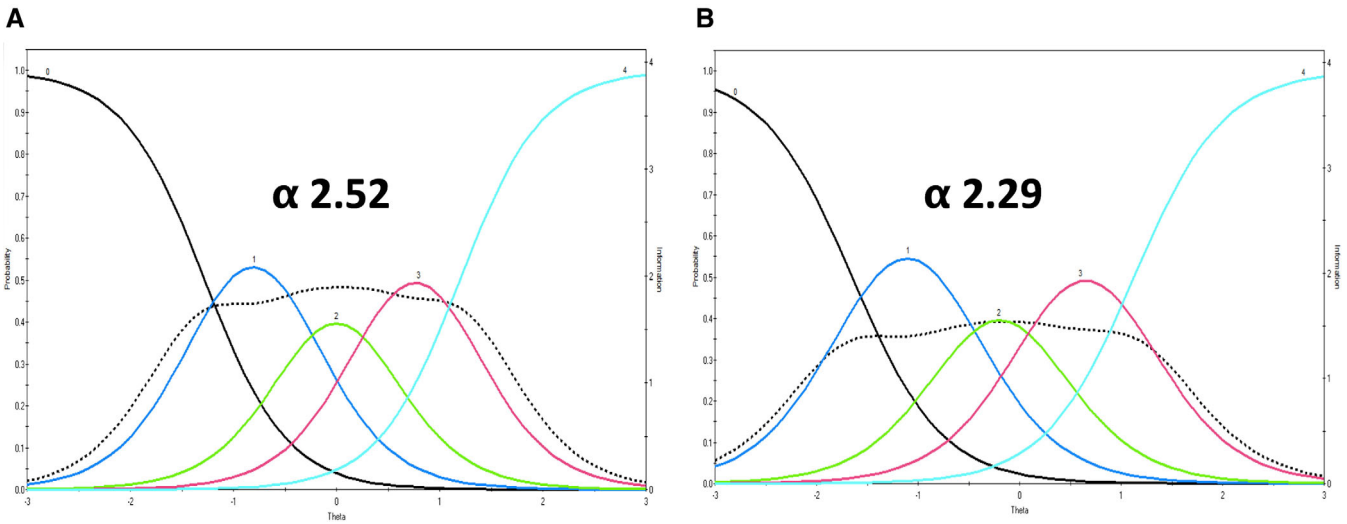

Figure 2. Examples of item discrimination parameter estimates for Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon (ASRAP) items 1 (**A**) and 2 (**B**) (these items are used as an example, as both were retained in the final ASRAP questionnaire). The broken curve is the item information curve, using the right-hand side information axis, to demonstrate how much information the individual item can contribute along the underlying Θ continuum. The solid curves are item characteristic curves, using the left-hand side probability axis, to demonstrate the probabilities of each response category being endorsed on the same Θ continuum, with each labeled 0–4 to correspond to the 5 response categories. The discrimination parameter (α) was added at the top of the figure, with α values >1.0 considered satisfactory. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25038/abstract.

the DIF. Similarly, items 10 and 38 had been noted to exhibit suboptimal model fit. Items 25 and 29 were removed solely on the basis of redundancy. For both item pairs, we applied expert consensus to retain the items that were felt to have the stronger content validity and inclusivity based on earlier qualitative work. For example, we considered RP impact on personal/private life more inclusive than on home family life (items 25 versus 26). We also considered RP symptoms aggravated by visits to the grocery store as superior to household chores/washing vegetables given the broader inclusivity suggested in our qualitative study (items 28 versus 29). We retained 2 potential redundant item pairs (item pairs 6 and 7, and 36 and 37), as the individual item content captured distinct domains considered important by patients in our qualitative work (color change versus feeling cold, and doing things differently versus requiring help from others, respectively) (see Supplementary Table 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25038). It was felt that the content validity of the ASRAP questionnaire might be impaired by not capturing these experiences.

Model fit assessment. Four items (items 9, 10, 27, and 38) exhibited suboptimal model fit; 2 of which had been flagged for potential removal due to item redundancy. For the other 2 items, one was removed (item 27, pertaining to ability to control the intensity of SSc-RP) and one retained due to perceived important content validity (item 9, frequency of SSc-RP attacks, which has formed the basis of the most widely used SSc-RP clinical trial end point of the last 2 decades). We inspected the individual item information curves for the remaining 28 items, which all contributed meaningful information.

Unanticipated DIF. The IRT likelihood ratio test did not flag any age-related DIF items but found 3 location-related DIF items (items 2, 5, and 35), which indicated that responses between US and UK respondents differed without explanation. One item (item 35, concerning avoidance of activities that could aggravate SSc-RPs symptoms) was subsequently removed (the others having already been identified for removal during earlier steps).

Summary of retained ASRAP items. After the refining steps described above, the final ASRAP item bank retained 27 items. A summary of the steps leading to the removal of redundant and/or poorly fitting items is presented in Table 2. The parameter estimates of the 27 retained long-form ASRAP items are presented in Table 3. Figure 3 displays the test information curve for the 27-item long-form ASRAP questionnaire and its corresponding SE. An SE of 0.30 corresponds approximately to a classical reliability of 0.90. The effective range of measurement for this final 27-item ASRAP item bank was estimated at –1.8 to +2.8 SDs (equating to recalibrated [(Θ score × 10) + 50] ASRAP scores ranging from 32 to 78). A scoring platform is under development to connect the raw score with the corresponding IRT Θ score for the ASRAP questionnaire (to be made available through the SCTC website).

Fixed item selection to create the ASRAP-SF. To develop the fixed short form, we rank ordered all 27 items on 4 criteria: discrimination parameter estimates; the percentage of items being selected in a simulated CAT based on the calibration sample; expected information under the standard normal

Table 2. Summary of item removal from preliminary ASRAP questionnaire*

Item	Factor analysis		IRT analysis				
	Exploratory	Confirmatory	*a* parameter	Local dependency	Model fit	DIF location	Final appraisal
Item 1							
Item 2				X		X	Remove
Item 3							
Item 4							
Item 5				X		X	Remove
Items 6–7				X			Retain both
Item 8							
Item 9					X		Retain
Item 10				X	X		Remove
Items 11–24							
Item 25				X			Remove
Item 26							
Item 27					X		Remove
Item 28							
Item 29				X			Remove
Item 30	X						Remove
Item 31	X						Remove
Item 32							
Item 33	X						Remove
Item 34			X				Remove
Item 35						X	Remove
Items 36–37				X			Retain both
Item 38				X	X		Remove
Item 39							

* Each X indicates the poor performing parameter. ASRAP = Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon; DIF = differential item functioning; IRT = item response theory.

distribution (mean ± SD 0 ± 1); and expected information under a normal distribution with a wider SD (SD 1.5). For CAT simulations, we set the minimum number of items to 8 and the maximum number of items to the full bank. We selected the top 8 items based on the convergence of the 4 psychometric criteria. In addition, we added 2 additional items based on content importance to build the 10-item fixed short form (pain during attacks and effect on personal/private life). The Pearson's correlation coefficient of the 10-item short form Θ scores and the full bank Θ scores was 0.976 ($P < 0.001$). The 2 ASRAP questionnaires had similar internal consistency, with Cronbach's alpha values of 0.918 (ASRAP) and 0.899 (ASRAP-SF). To facilitate the usage of the ASRAP-SF, a scoring platform is under development to connect the raw score with the corresponding IRT Θ score (to be made available through the SCTC website).

DISCUSSION

We have developed a novel patient-reported outcome instrument for assessing the severity and impact of SSc-RP. The ASRAP questionnaire has been devised and tested with direct input from SSc patients throughout the process, combined with an international consortium of SSc experts, to achieve the goal of fully capturing the patient lived experience of SSc-RP. The ASRAP questionnaire could play an important role as a clinical trial end point in future therapeutic trials of SSc-RP, pending reporting of construct validity, reliability, and responsiveness testing. From the outset, we aimed to develop a novel patient-reported outcome instrument that would capture the severity and impact of SSc-RP, with a specific focus on the impact of RP on how patients with SSc feel and function, which are vital prerequisites for instruments evaluated by regulatory bodies when considering marketing authorization of therapeutic interventions.

In this preliminary study, we have demonstrated strong content validity and feasibility of the novel ASRAP questionnaire. The items of the ASRAP questionnaire benefit from being grounded in themes and subthemes identified in our preparatory work to understand the patient experience of SSc-RP across diverse UK and US patient populations (2,3). Unlike existing patient-reported outcome instruments devoted to assessing SSc-RP, the items were developed with patient insight partner support, and where possible, we utilized the wording of quotations obtained from our underpinning qualitative research (3,11). To further strengthen content validity and comprehension, we undertook formal linguistic testing and cognitive debriefing interviews within the target patient population at UK and US sites. This item testing has ensured that the ASRAP questionnaire is comprehensible in different countries, minimizes ambiguity, and meets accepted criteria for optimal translatability into non-English languages. Feasibility has been confirmed by the high completion rate of the ASRAP questionnaire (>96%). The items have performed well with no discernible floor or ceiling effects.

We have refined and tested the ASRAP questionnaire in the largest study of SSc-RP undertaken to date, having enrolled

Table 3. Parameter estimates for the 27-item long-form ASRAP questionnaire*

Item	Original ASRAP item number	Final calibration				
		a	b_1	b_2	b_3	b_4
1	ASRAP 1†	2.37	−1.33	−0.34	0.35	1.25
2	ASRAP 3	1.76	−1.31	−0.36	0.39	1.47
3	ASRAP 4	2.04	−0.95	−0.21	0.39	1.33
4	ASRAP 6	1.97	−2.53	−1.33	−0.71	0.44
5	ASRAP 7	1.72	−2.45	−1.43	−0.60	0.44
6	ASRAP 8†	2.87	−1.34	−0.48	0.24	0.84
7	ASRAP 9	1.43	−2.49	−0.09	1.20	2.26
8	ASRAP 11	1.21	−2.78	−1.19	0.12	1.72
9	ASRAP 12	2.82	0.43	0.96	1.40	1.96
10	ASRAP 13†	3.66	−0.37	0.42	0.92	1.45
11	ASRAP 14†	3.84	−0.71	0.08	0.54	1.19
12	ASRAP 15†	3.54	−0.14	0.42	0.98	1.56
13	ASRAP 16†	3.38	0.44	0.99	1.56	1.99
14	ASRAP 17	1.93	0.25	1.09	1.68	2.13
15	ASRAP 18	2.75	0.16	0.95	1.47	2.10
16	ASRAP 19†	3.79	−0.51	0.26	0.80	1.34
17	ASRAP 20	2.28	0.53	1.23	1.79	2.71
18	ASRAP 21	2.55	−0.77	0.15	0.68	1.47
19	ASRAP 22†	3.34	−0.56	0.29	0.92	1.54
20	ASRAP 23†	3.25	−0.27	0.54	1.02	1.67
21	ASRAP 24	2.73	−0.10	0.53	0.99	1.60
22	ASRAP 26†	2.81	−0.10	0.57	1.18	1.75
23	ASRAP 28	1.13	−2.36	−1.52	−0.32	0.73
24	ASRAP 32	1.41	−2.74	−1.18	0.06	2.97
25	ASRAP 36	2.40	−1.35	−0.63	0.4	1.33
26	ASRAP 37	2.57	−0.71	0.03	0.69	1.66
27	ASRAP 39	2.50	−0.05	0.55	1.22	2.01

* ASRAP = Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon.
† Final 10-item fixed ASRAP short-form (ASRAP-SF) questionnaire items.

420 patients with SSc from 7 English-speaking SSc centers in the UK and US. We have taken a data-driven approach to refining and scoring the ASRAP questionnaire based on factor analysis and IRT graded response modeling. Factor analyses have confirmed a strong single underlying trait and a clinically meaningful construct of SSc-RP. We have undertaken a data-driven approach to removing redundant or poorly fitting items, only retaining items (n = 3) when SSc experts considered the content

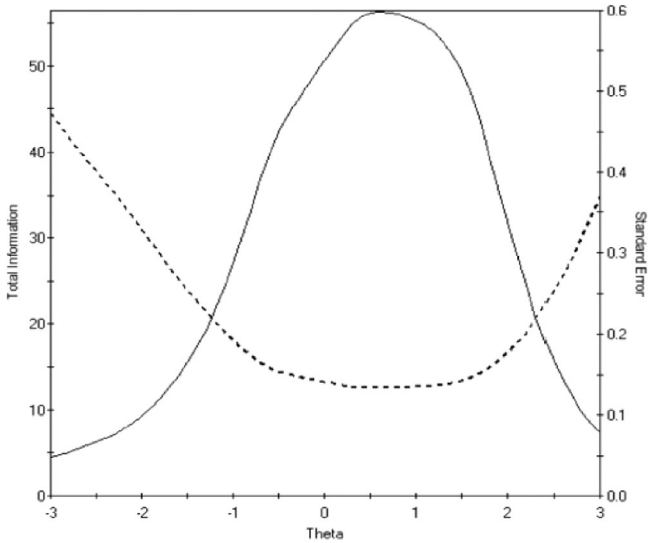

Figure 3. Test information of the 27-item long-form Assessment of Systemic Sclerosis–Associated Raynaud's Phenomenon (ASRAP) questionnaire for group 1. An SE of 0.30 corresponds approximately to a classical reliability of 0.90. The effective range of measurement for the final 27-item ASRAP item bank is estimated at −1.8 to +2.8 SDs. The broken curve shows the SE; the solid curve shows total information.

validity to be jeopardized by their removal. Of the 13 potentially problematic items, we were reassured to find that 7 of 12 of the removed items were identified as problematic across ≥2 domains of our 2-stage analysis. We have confidence that each of the remaining 27-item long-form ASRAP questionnaire items capture the broad-ranging lived experiences of SSc-RP in a concise but comprehensive manner. The development of a 10-item short-form questionnaire, meanwhile, will provide a much needed tool to capture the severity and impact of SSc-RP (and therapeutic response) in routine clinical practice. The 2-week prospective RCS diary has not easily permitted the capture of practice-based evidence on SSc-RP severity and treatment response, but the retrospective (1-week recall period) nature of the ASRAP questionnaire could provide a valuable opportunity to do so in the future.

The ASRAP questionnaire also captures important experiences of SSc-RP that are not assessed using RCS diary parameters, such as emotional distress, exacerbating factors, self-management, and adaptation. The ASRAP captures information on RP attack frequency that could be used alongside traditional diary-based approaches for more fully assessing SSc-RP attack characteristics if necessary. The paradigm of RP attacks is relevant to the patient experience of SSc-RP, although some patients report difficulty distinguishing RP attacks from background digital ischemic symptoms (3,14,24). Symptoms of SSc-RP appear to evolve with greater disease duration, with some patients identifying with a paradigm of more persistent digital ischemic symptoms (13). Assessment of SSc-RP attack frequency and duration also fails to consider the severity of individual attacks and/or the considerable efforts taken by patients with SSc to avoid and ameliorate RP symptoms when they occur (2,3,12). While single-item global RP severity instruments such as the RCS attempt to capture the multifaceted nature of RP, they provide little insight into the factors influencing their score. The ASRAP questionnaire will allow a more nuanced assessment of RP severity and impact, and future modifications may allow post hoc analyses to examine the determinants and impact of therapeutic intervention on specific aspects of the lived experience of SSc-RP. The present study has focused on the development and refinement of the ASRAP questionnaire. Analyses reporting the construct validity, reliability, and responsiveness of the ASRAP and ASRAP-SF questionnaires will be reported separately.

In conclusion, the ASRAP questionnaire is a novel patient-reported outcome instrument for assessing the severity and impact of SSc-RP for use in both clinical practice and research/trial protocols. Significant patient participation in the development and refinement of the ASRAP questionnaire has ensured strong content validity and feasibility. Further work is underway to examine the construct validity, reliability, and responsiveness of the ASRAP and ASRAP-SF questionnaires to enable these instruments to be applied in both clinical trials and routine clinical practice for assessing SSc-RP severity.

ACKNOWLEDGMENTS

This work would not have been possible without the continued support of the SCTC committee and SCTC members who have supported the creation and ongoing efforts of the Vascular Working Group.

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. Pauling 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. Yu, Domsic, Saketkoo, Withey, Frech, Herrick, Hummers, Shah, Denton, Khanna, Pauling.

Acquisition of data. Yu, Domsic, Saketkoo, Frech, Herrick, Hummers, Shah, Denton, Khanna, Pauling.

Analysis and interpretation of data. Yu, Domsic, Saketkoo, Withey, Frech, Herrick, Hummers, Shah, Denton, Khanna, Pauling.

REFERENCES

1. Walker UA, Tyndall A, Czirjak L, et al. Geographical variation of disease manifestations in systemic sclerosis: a report from the EULAR Scleroderma Trials and Research (EUSTAR) group database. Ann Rheum Dis 2009;68:856–62.
2. Pauling JD, Saketkoo LA, Matucci-Cerinic M, et al. The patient experience of Raynaud's phenomenon in systemic sclerosis. Rheumatology (Oxford) 2019;58:18–26.
3. Pauling JD, Domsic RT, Saketkoo LA, et al. Multinational qualitative research study exploring the patient experience of Raynaud's phenomenon in systemic sclerosis. Arthritis Care Res (Hoboken) 2018; 70:1373–84.
4. Bassel M, Hudson M, Taillefer SS, et al. Frequency and impact of symptoms experienced by patients with systemic sclerosis: results from a Canadian National survey. Rheumatology (Oxford) 2011;50: 762–7.
5. Black CM, Halkier-Sorensen L, Belch JJ, et al. Oral iloprost in Raynaud's phenomenon secondary to systemic sclerosis: a multicentre, placebo-controlled, dose-comparison study. Br J Rheumatol 1998;37:952–60.
6. Wigley FM, Korn JH, Csuka ME, et al. Oral iloprost treatment in patients with Raynaud's phenomenon secondary to systemic sclerosis: a multicenter, placebo-controlled, double-blind study. Arthritis Rheum 1998;41:670–7.
7. Merkel PA, Herlyn K, Martin RW, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. Arthritis Rheum 2002;46:2410–20.
8. Pauling JD. The challenge of establishing treatment efficacy for cutaneous vascular manifestations of systemic sclerosis. Expert Rev Clin Immunol 2018;14:431–42.
9. Roustit M, Blaise S, Allanore Y, et al. Phosphodiesterase-5 inhibitors for the treatment of secondary Raynaud's phenomenon: systematic review and meta-analysis of randomised trials. Ann Rheum Dis 2013; 72:1696–9.
10. Khouri C, Lepelley M, Bailly S, et al. Comparative efficacy and safety of treatments for secondary Raynaud's phenomenon: a systematic review and network meta-analysis of randomised trials. Lancet Rheumatol 2019;1:e237–46.
11. Pauling JD, Frech TM, Domsic RT, et al. Patient participation in patient-reported outcome instrument development in systemic sclerosis. Clin Exp Rheumatol 2017;35 Suppl 106:184–92.
12. Pauling JD, Frech TM, Hughes M, et al. Patient-reported outcome instruments for assessing Raynaud's phenomenon in systemic

sclerosis: a SCTC Vascular Working Group report. J Scleroderma Relat Disord 2018;3:249–52.

13. Pauling JD, Reilly E, Smith T, et al. Evolving symptom characteristics of Raynaud's phenomenon in systemic sclerosis and their association with physician and patient-reported assessments of disease severity. Arthritis Care Res (Hoboken) 2019;71:1119–26.
14. Pauling JD, Saketkoo LA, Domsic RT. Patient perceptions of the Raynaud's condition score diary provide insight into its performance in clinical trials of Raynaud's phenomenon: comment on the article by Denton et al [letter]. Arthritis Rheumatol 2018;70:973–4.
15. DeWalt DA, Rothrock N, Yount S, et al. Evaluation of item candidates: the PROMIS qualitative item review. Med Care 2007;45 Suppl 1: S12–21.
16. Reise SP, Yu JY. Parameter recovery in the graded response model using MULTILOG. J Educ Meas 1990;27:133–44.
17. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity: establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. Value Health 2011;14:978–88.
18. Van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737–47.
19. Muthén LK, Muthén BO. Mplus: statistical analysis with latent variables. User's guide. 6th ed. 2010. URL: https://www.statmodel.com/download/usersguide/Mplus%20Users%20Guide%20v6.pdf.
20. Samejima F. Estimation of latent ability using a response pattern of graded responses. Psychometrika 1969;34 Suppl 1:1–97.
21. Cai L, Thissen D, du Toit SH. IRTPRO for Windows. Lincolnwood (IL): Scientific Software International; 2011.
22. Holland P, Wainer H, editors. Differential item functioning. Hillsdale (NJ): Routledge; 1993. p. 67–113.
23. Choi SW. Firestar: computerized adaptive testing simulation program for polytomous item response theory models. Appl Psychol Meas 2009;33:644–5.
24. Murphy SL, Lescoat A, Alore M, et al. How do patients define Raynaud's phenomenon? Differences between primary and secondary disease. Clin Rheumatol 2021;40:1611–6.