Domsic Robyn Therese (Orcid ID: 0000-0002-2765-0922)
Frech Tracy (Orcid ID: 0000-0002-5472-3840)
Herrick Ariane L (Orcid ID: 0000-0003-4941-7926)
Denton Christopher P (Orcid ID: 0000-0003-3975-8938)
Khanna Dinesh (Orcid ID: 0000-0003-1412-4453)
Pauling John D (Orcid ID: 0000-0002-2793-2364)

The Assessment of Systemic sclerosis-associated RAynaud's Phenomenon (ASRAP) questionnaire: Item Bank and Short Form Development

Lan Yu PhD[1], Robyn T Domsic MD MPH [2] Lesley-Ann Saketkoo MD MPH[3], Jane Withey[4], Tracy M. Frech MD MS[5], Ariane L Herrick MD FRCP [6], Laura K Hummers MD MSc [7], Ami A Shah MD MHS [7], Christopher P Denton PhD FRCP [8], Dinesh Khanna MD MSc [9,10], John D Pauling BMedSci BMBS PhD FRCP [4,11,12]

[1] Department of Medicine, University of Pittsburgh, PA, USA

[2] Division of Rheumatology and Clinical Immunology, University of Pittsburgh Pittsburgh, PA, USA

[3] University of Tulane Medical Center, New Orleans, LA, USA

[4] Royal National Hospital for Rheumatic Diseases (part of the Royal United Hospitals NHS Foundation Trust), Bath, UK

[5] Vanderbilt University, Nashville, TN USA

[6] The University of Manchester, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, and NIHR Manchester Biomedical Research Centre, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre, UK.

The University of Manchester and Salford Royal NHS Foundation Trust, UK

[7] Johns Hopkins University School of Medicine, Baltimore, MD, USA

[8] Centre for Rheumatology and Connective Tissue Diseases, Royal Free Hospital, London, UK

[9] University of Michigan Scleroderma Program, Ann Arbor, MI, USA

[10] Division of Rheumatology University of Michigan Ann Arbor, MI, USA

[11]Department of Rheumatology, North Bristol NHS Trust, Bristol, UK

[12] Musculoskeletal Research Unit, Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acr.25038

Corresponding Author:

Dr John D Pauling BMedSci PhD FRCP

Consultant Rheumatologist & Honorary Senior Lecturer

Department of Rheumatology, North Bristol NHS Trust, Bristol, UK

Tel: (0044) 01179505050 JohnPauling@nhs.net

Key Indexing Terms: Systemic sclerosis; Raynaud's phenomenon, Patient-reported outcome instrument, clinical trial, validation

Short running title: Development and refinement of the ASRAP questionnaire

Sources of support: This work was supported by a grant from the US Department of Defense, Award Number W81XWH-18-1-0602and the Scleroderma Clinical Trials Consortium.

Word Count: 4088

Abstract

Objectives: To develop, refine and score a novel patient-reported outcome (PRO) 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 multi-center 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 438 SSc patients from UK and US scleroderma centers over two consecutive winters. Factor analysis with item response theory (IRT) 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 (GRM). A fixed 10-item short-form ASRAP 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.

Word count: 250

Significance & Innovation

- Development of a novel patient-reported outcome instrument (ASRAP questionnaire) grounded in the lived patient experience of Raynaud's phenomenon in systemic sclerosis
- Extensive target patient population involvement in development of ASRAP questionnaire item bank
- Data driven approach to remove redundant and/or poorly fitting items from ASRAP questionnaire
- Development, calibration and scoring of both ASRAP questionnaire and 10-item short form ASRAP questionnaire

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 ischaemic 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 discolouration (2, 3). People living with SSc adapt to avoid and ameliorate symptoms of RP, although the need to avoid cold exposure and stressful interactions does itself impact on social and work participation; often necessitating the support of others (2, 3).

The episodic and uniquely personalised experience of SSc-RP has led to a reliance upon patient-reported outcome (PRO) 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 utilised 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 or 100mm visual analogue scale (VAS) 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 endpoints has been challenging and there are currently no FDA-approved medications for SSc-RP (8). Meta-analyses have indicated that the net benefit of treatments such as calcium channel blockers (CCBs) and phosphodiesterase V inhibitors (PDEVi) on existing clinical trial endpoints are either absent, or modest at best (9, 10).

SSc patients were not involved in the development of the three 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 monitoring (5-7, 11). Concerns have been raised by SSc patients and experts about the RCS diary, whose focus on RP attack frequency/duration 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 emotional impact of SSc-RP are not fully captured by RCS diary parameters (3). Against this backdrop, a multi-center collaborative effort has been undertaken to develop a new PRO 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 PRO instrument for capturing the impact and severity of SSc-RP.

Methods

Conceptual framework and development of 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 PRO 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 study steering committee with lived experience of SSc (JW), to capture the themes and sub-themes comprising the lived experience of SSc-RP identified in an earlier international multicentre qualitative research study of SSc-RP and 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 Outcome Information System (PROMIS) initiative with response levels best suited for item response theory (IRT) modelling (15-17). A recall time frame of the past 7 days was agreed amongst SSc experts and patients to optimise feasibility and utility in clinical trial setting.

Qualitative item review

Development and refinement of the ASRAP questionnaire

Iterative modification of the ASRAP instructions and preliminary 37 items was undertaken with input from 4 SSc experts (JDP, RTD, LAS and DK) to ensure item wording, recall period and response options were simple, understandable, relevant to specific domain concepts and 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 emotional impact of SSc-RP limiting usual activity) resulting in a 39-item preliminary ASRAP questionnaire (Supplementary Material 1). Linguistic evaluation of the 39-item preliminary ASRAP questionnaire was undertaken using the Simple Measure of Gobbledygook (SMOG) to ensure 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 (Supplementary Material 2).

Cognitive de-briefing 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 approximately 90 minutes with a scheduled comfort break and three subjects began the de-briefing interview on item 20 to ensure interview fatigue did not impact on responses. Cognitive debriefing also sought to ensure 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 (summarised in Table 1). No new items were proposed. An item tracking matrix was devised to document modifications during item development. The ASRAP questionnaire was then deemed ready for field testing.

Cross-sectional calibration study

The international multicentre ASRAP cross-sectional calibration study enrolled English-speaking SSc patients from seven UK (Bath, London and Manchester) and US (Pittsburgh,

Baltimore, Michigan and Salt Lake City) scleroderma centers over seven winter months (February-March 2019 and November 2019-March 2020). To provide an expected adequate sample for robust unidimensional Graded Response Model (GRM) analysis, our target sample size was 500 (18). The study had research ethics committee approval at each UK and US site (available on request) and was conducted in accordance with the principles of the Declaration of Helsinki.

Eligibility and study procedures

All patients were aged ≥18 years, fulfilled the American College of Rheumatology/European League Against Rheumatism classification criteria for SSc (19) and had good comprehension of written/spoken English. Pregnant people and/or subjects whose vasodilator medication had not been stable within the previous 4 weeks were excluded. To ensure 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 demographics and clinical phenotype were captured using a clinician case-report form.

Statistical analysis

A 2-step analytical approach was undertaken.

Factor Analysis

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 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 uni-dimensionality to allow us to proceed with IRT analyses.

Development and refinement of the ASRAP questionnaire

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 two similarly sized sub-samples; one for exploratory factor analysis (unweighted least squares, EFA) and the second for confirmatory factor analysis (CFA). Both EFA and CFA were conducted using Mplus 6.2 with promax rotation (20). In the CFA, the items were treated as categorical variables, and the robust weighted least squares (WLSMV) 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 <0.5 and/or low item discrimination parameter estimates (α<1.0) were considered for removal.

Item Response Theory Analysis

The most commonly used IRT model for polytomous items (i.e., items with three or more ordinal response categories) is the two-parameter graded response model (GRM) (21) . The GRM has a slope parameter and *n* - 1 threshold parameters for each item, where *n* is the number of response categories (five in the present analyses). The slope parameter measures item discrimination, i.e., how well the item differentiates between higher vs lower levels of severity (or Θ in IRT terms). Useful items have large slope parameters. Threshold parameters measure item difficulty, i.e., the ease vs 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 v. 2.1 (22).

Local dependency (LD) marginal Chi square analyses identified redundant items due to high local dependency (residual correlations) with other items and were removed (21,22).

Differential item functioning (DIF) occurs when characteristics such as age, gender, 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 vs. age>60), birth year (odd vs. even birth year) and location (UK vs. 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 significant DIF (p<.01) (23).

Fixed Short Form Development

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 endpoint in SSc clinical trials. After the ASRAP items were calibrated using IRT, we developed fixed ASRAP-SF based on computerized adaptive testing (CAT) simulations. We used four 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 of zero and SD of one, and expected information under a normal distribution with a larger SD (i.e., a mean of zero and SD of 1.5). The CAT simulations were performed using the Firestar program (24) .

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 α.

ASRAP scoring

To aid easier conceptualisation of ASRAP scores, Θ values (theoretically ranging from -3.0 to +3.0 SD around zero), were re-calibrated by multiplying the Θ score by 10 and adding 50, to achieve a scoreable ASRAP range of between 20-80.

Results

Study population

Four hundred and thirty-eight SSc subjects were enrolled at UK (n=238) and US (n=200) sites, with full ASRAP completion (no missing item responses) on 421 subjects (96.1%), indicating strong feasibility within the target patient population. Enrolment 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 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 (79.8% female), ethnicity (86.5% white), disease subset (59.4%, 34.7% and 5.9% for limited, diffuse and *sine* scleroderma respectively), age (mean 58.9 years, SD 12.4), disease duration (15.4 years (SD 12.1) since RP-onset and 11.8 (SD 9.7) since first non-RP symptom of SSc), clinical features and autoantibody specificities; confirming this was a representative cohort of SSc patients. There were only 19 (4.5%) current smokers, with 154 (36.6% past smokers.

Factor Analysis

Prior to factor analysis, we examined the frequencies of each of 39 ASRAP items for sparse response frequency. There were no items with response categories having less than 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=202) and CFA (n=219). The first two eigenvalues were 21.58 and 3.08 with a ratio of 7.0, which is a strong indication of 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 clinical meaningful construct. The majority of items had a factor loading larger than .450 (with 36 of 39 ASRAP items with factor loadings larger than .80). Three items (items 30, 31, and 33) had factor loadings between 0.415 and 0.443 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 (SRMR) of 0.08, and a Confirmatory Factor Index (CFI) of 0.92, except for a Root Mean Square Error of Approximation (RMSEA) of 0.11 (just above the desired range of .08).

IRT calibrations

The 36 items retained from factor analysis were calibrated using IRT graded response model. Item Discrimination parameters were generally satisfactory with α ranging from 1.18 to 3.62 (Figure 2, with α>1.0 considered satisfactory), with the exception of one item (item 34, pertaining to use of hand warming devices) with a marginal item discrimination parameter estimate of 1.06, 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 (>10 indicates redundancy), six individual items (items 2, 5, 10, 25, 29, and 38) were removed (Supplementary Table 1). 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 the differential item functioning. Similarly, items 10 and 38 had been noted to exhibit suboptimal model fit (see later). Items 25 and 29 were removed solely on the basis of redundancy. For both items, 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 'home family life' (Items 25 vs. 26). We also considered RP symptoms aggravated by 'visits to grocery store' as superior to 'household chores/washing vegetables' given the broader

inclusivity suggested in our qualitative study (items 28 vs 29). We retained 2 potential redundant item pairs (item pairs 6 & 7, and 36 & 37) as the individual item content captured distinct domains considered important by patients in our qualitative work (colour change vs feeling cold, and 'doing things differently' vs. 'requiring help from others' respectively). It was felt the content validity of ASRAP might be impaired by not capturing these experiences.

Model fit assessment

Four 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 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 endpoint of the last 2 decades). We inspected the individual item information curves for the remaining 28 items which all contributed meaningful information.

Unanticipated differential item functioning

The IRT likelihood ratio test did not flag any age-related DIF items, but found three location-related DIF items (items 2, 5 and 35) which indicated 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 above refining steps, 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 are of the 27 retained long-form ASRAP items is presented in Table 3. Figure 3 displays the test information curve for the 27-item long-form ASRAP questionnaire and its corresponding standard error. A standard error of .30 corresponds approximately to a classical reliability of .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 re-calibrated ([Θ score x 10] +50) ASRAP scores ranging from 32-78). A scoring platform is under development

to connect the raw score with the corresponding IRT Θ score for the ASRAP questionnaire (shall be made available through SCTC website).

Fixed item selection to create the ASRAP-short form (ASRAP-SF)

To develop the fixed short form, we rank ordered all 27 items on four criteria: discrimination parameter estimates, the percentage of items the item being selected in a simulated CAT based on the calibration sample, expected information under the standard normal distribution (mean =0, SD=1), and expected information under a normal distribution with a wider standard deviation (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 top 8 items based on the convergence of the four 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 α values of 0.920 (ASRAP) and 0.902 (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 (shall be made available through SCTC website).

Discussion

We have developed a novel PRO 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 endpoint 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 PRO instrument that would capture the severity and impact of SSc-RP, with a specific focus on the impact of RP on how people with SSc 'feel' and

'function', which are vital pre-requisites for instruments evaluated by regulatory bodies when considering marketing authorisation 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 PRO instruments devoted to assessing SSc-RP, the items were developed with patient insight partner support, and where possible we utilised 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 de-briefing interviews within the target patient population at UK and US sites. This item testing has ensured the ASRAP questionnaire is comprehensible within 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 438 patients with SSc from 7 English-speaking SSc centres 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 modelling. Factor analyses have confirmed a strong single underlying trait and 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 validity to be jeopardised by their removal. Of the 13 potentially problematic items, we were reassured to find that 7/12 of the removed items were identified as problematic across 2 or more domains of our 2-stage analysis. We have confidence that each of the remaining 27-item long form ASRAP questionnaire 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 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 ischaemic symptoms (3, 14, 25). Symptoms of SSc-RP appear to evolve with greater disease duration with some patients identifying with a paradigm of more persistent digital ischaemic 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). Whilst, 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 shall 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 report has focussed on the development and refinement of the ASRAP questionnaire. Analyses reporting the construct validity, reliability and responsiveness of the ASRAP and ASRAP-SF questionnaires shall be reported separately.

Conclusions

The ASRAP questionnaire is a novel PRO 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.

Development and refinement of the ASRAP questionnaire

Acknowledgements

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.

Funding

This work was supported by the US Department of Defense, Award W81XWH1810602, and the Scleroderma Clinical Trials Consortium.

Disclosures

JP has received personal support from Janssen, Astra Zeneca, Permeatus Inc, Boehringher-Ingelheim and Sojournix Pharma.

CD reports grants and personal fees from GlaxoSmithKline and CSL Behring; grants from Horizon, Servier and Arxx Therapeutics; and personal fees from Sanofi, Boehringer Ingelheim, Roche, Corbus, Acceleron and Gesynta.

DK reports consulting fees from Acceleron, Actelion, Amgen, Bayer, Boehringer Ingelheim, Chemomab, CSL Behring, Genentech/Roche Horizon, Paracrine Cell Therapy, Mitsubishi Tanabe Pharma, Prometheus, Theraly; DK is Chief Medical Officer of Eicos Sciences, Inc, a subsidiary of CiviBioPharma and has stock options.

ALH has received consultancy fees from Arena, Boehringer-Ingelheim, Camurus, CSL Behring and Gesynta Pharma, speaker fees from Actelion and Janssen, and research funding from Actelion and Gesynta Pharma.

LKH has received research or consulting support from Boehringer-Ingelheim, Cumberland Pharmaceuticals, CSH Behring, Mitsubishi Tanabe Pharma Corporation, Horizon Pharmaceuticals, Biotest

Data availability statement

The data supporting the ASRAP development can be made available upon reasonable request.

References

Development and refinement of the ASRAP questionnaire

1. Walker UA, Tyndall A, Czirjak L, Denton CP, Farge-Bancel D, Kowal-Bielecka O, 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(6):856-62.
2. Pauling JD, Saketkoo LA, Matucci-Cerinic M, Ingegnoli F, Khanna D. The patient experience of Raynaud's phenomenon in systemic sclerosis. Rheumatology (Oxford). 2019;58(1):18-26.
3. Pauling JD, Domsic RT, Saketkoo LA, Almeida C, Withey J, Jay H, et al. Multinational Qualitative Research Study Exploring the Patient Experience of Raynaud's Phenomenon in Systemic Sclerosis. Arthritis Care Res (Hoboken). 2018;70(9):1373-84.
4. Bassel M, Hudson M, Taillefer SS, Schieir O, Baron M, Thombs BD. Frequency and impact of symptoms experienced by patients with systemic sclerosis: results from a Canadian National Survey. Rheumatology (Oxford). 2011;50(4):762-7.
5. Black CM, Halkier-Sorensen L, Belch JJ, Ullman S, Madhok R, Smit AJ, et al. Oral iloprost in Raynaud's phenomenon secondary to systemic sclerosis: a multicentre, placebo-controlled, dose-comparison study. Br J Rheumatol. 1998;37(9):952-60.
6. Wigley FM, Korn JH, Csuka ME, Medsger TA, Jr., Rothfield NF, Ellman M, 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(4):670-7.
7. Merkel PA, Herlyn K, Martin RW, Anderson JJ, Mayes MD, Bell P, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. Arthritis Rheum. 2002;46(9):2410-20.
8. Pauling JD. The challenge of establishing treatment efficacy for cutaneous vascular manifestations of systemic sclerosis. Expert Rev Clin Immunol. 2018;14(5):431-42.
9. Roustit M, Blaise S, Allanore Y, Carpentier PH, Caglayan E, Cracowski JL. Phosphodiesterase-5 inhibitors for the treatment of secondary Raynaud's phenomenon: systematic review and meta-analysis of randomised trials. Ann Rheum Dis. 2013;72(10):1696-9.
10. Khouri C, Lepelley M, Bailly S, Blaise S, Herrick AL, Matucci-Cerinic M, et al. Comparative efficacy and safety of treatments for secondary Raynaud's phenomenon: a systematic review and network meta-analysis of randomised trials. The Lancet Rheumatology. 2019;1(4):e237-e46.
11. Pauling JD, Frech TM, Domsic RT, Hudson M. Patient participation in patient-reported outcome instrument development in systemic sclerosis. Clin Exp Rheumatol. 2017;35 Suppl 106(4):184-92.
12. Pauling JD, Frech TM, Hughes M, Gordon JK, Domsic RT, Anderson ME, 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(3):249-52.
13. Pauling JD, Reilly E, Smith T, Frech TM. 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(8):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. Arthritis Rheumatol. 2018;70(6):973-4.
15. DeWalt DA, Rothrock N, Yount S, Stone AA, Group PC. Evaluation of item candidates: the PROMIS qualitative item review. Med Care. 2007;45(5 Suppl 1):S12-21.
16. Reise SP, Yu JY. Parameter Recovery in the Graded Response Model Using Multilog. J Educ Meas. 1990;27(2):133-44.
17. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, 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(8):978-88.
18. Reise SP, Yu, J. Parameter recovery in the graded response model using MULTILOG. J Educational Measurement. 1990;27:133-44.

19. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis. 2013;72(11):1747-55.
20. Muthén LM, B. . Mplus User's Guide (Version 6.2). Los Angeles, CA2010.
21. Samejima F. Estimation of Latent Ability Using a Response Pattern of Graded Responses. Psychometric Society, Richmond, VA. 1968.
22. Cai L, Thissen, D., du Toit, S.H.C. IRTPRO for Windows. Lincolnwood, IL, Scientific Software International. 2011.
23. Thissen DS, L.; Wainer, H. Detection of differential item functioning using the parameters of item response models. In: Holland PW, H., , editor. Differential Item Functioning. Hillsdale, NJ:: Lawrence Erlbaum; 1993. p. 67-113.
24. Choi SW. Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. Applied Psychological Measurement. 2009;33(3):644-5.
25. Murphy SL, Lescoat A, Alore M, Hughes M, Pauling JD, Sabbagh M, et al. How do patients define Raynaud's phenomenon? Differences between primary and secondary disease. Clin Rheumatol. 2021;40(4):1611-6.

Tables and Figures

Figure 1. Conceptual map of the major themes and sub-themes comprising the patient experience of Raynaud's phenomenon in systemic sclerosis

Reproduced with permission from (3)

Table 1. Summary of modifications made to ASRAP items following formal cognitive de-briefing interviews

Item number or al 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 / super-market has caused Raynaud's symptoms	Being inside a grocery store / super-market 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

Figure 2. Examples of Item Discrimination Parameter estimates for ASRAP items 1 and 2.

The dashed 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 labelled 0-4 to corresponding the 5 response categories. The discrimination parameter (α) was added at the top of the figure with α values >1.0 considered satisfactory

Table 2. Summary of item removal from preliminary ASRAP questionnaire

Item	Factor Analysis		IRT Analysis				Final appraisal
	Exploratory	Confirmatory	*a* parameter	Local Dependency	Model Fit	DIF-Location	
Item 1							
Item 2				X		X	REMOVE
Item 3							
Item 4							
Item 5				X		X	REMOVE
Item 6-7				X			RETAIN BOTH
Item 8							
Item 9					X		RETAIN
Item 10				X	X		REMOVE
Item 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
Item 36-37				X			RETAIN BOTH
Item 38				X	X		REMOVE
Item 39							

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

-		Final Calibration				
Item	**Original ASRAP item number**	***a***	***b_1***	***b_2***	***b_3***	***b_4***
1	ASRAP_1 *	2.39	-1.3	-0.33	0.36	1.25
2	ASRAP_3	1.78	-1.31	-0.37	0.41	1.47
3	ASRAP_4	2.09	-0.92	-0.2	0.39	1.31
4	ASRAP_6	1.96	-2.5	-1.32	-0.7	0.46
5	ASRAP_7	1.7	-2.37	-1.37	-0.56	0.47
6	ASRAP_8 *	2.92	-1.3	-0.47	0.26	0.84
7	ASRAP_9	1.42	-2.46	-0.09	1.17	2.25
8	ASRAP_11	1.23	-2.7	-1.19	0.11	1.67
9	ASRAP_12	2.74	0.42	0.97	1.41	1.98
10	ASRAP_13 *	3.71	-0.37	0.42	0.92	1.46
11	ASRAP_14 *	3.81	-0.71	0.08	0.55	1.2
12	ASRAP_15 *	3.51	-0.14	0.44	0.98	1.56
13	ASRAP_16 *	3.4	0.44	0.99	1.57	1.98
14	ASRAP_17	1.83	0.25	1.14	1.75	2.21
15	ASRAP_18	2.66	0.16	0.99	1.51	2.14
16	ASRAP_19 *	3.76	-0.5	0.27	0.81	1.36
17	ASRAP_20	2.23	0.53	1.23	1.82	2.77
18	ASRAP_21	2.6	-0.76	0.15	0.67	1.48
19	ASRAP_22 *	3.41	-0.55	0.29	0.93	1.55
20	ASRAP_23 *	3.26	-0.28	0.52	1	1.64
21	ASRAP_24	2.81	-0.11	0.51	0.97	1.58
22	ASRAP_26 *	2.84	-0.11	0.55	1.18	1.71
23	ASRAP_28	1.12	-2.36	-1.53	-0.3	0.78
24	ASRAP_32	1.43	-2.72	-1.17	0.06	2.93
25	ASRAP_36	2.47	-1.33	-0.62	0.4	1.31
26	ASRAP_37	2.63	-0.7	0.03	0.69	1.64
27	ASRAP_39	2.52	-0.05	0.54	1.22	2.04

Note: items with * are the final 10-item fixed short-form ASRAP (ASRAP-SF) questionnaire items

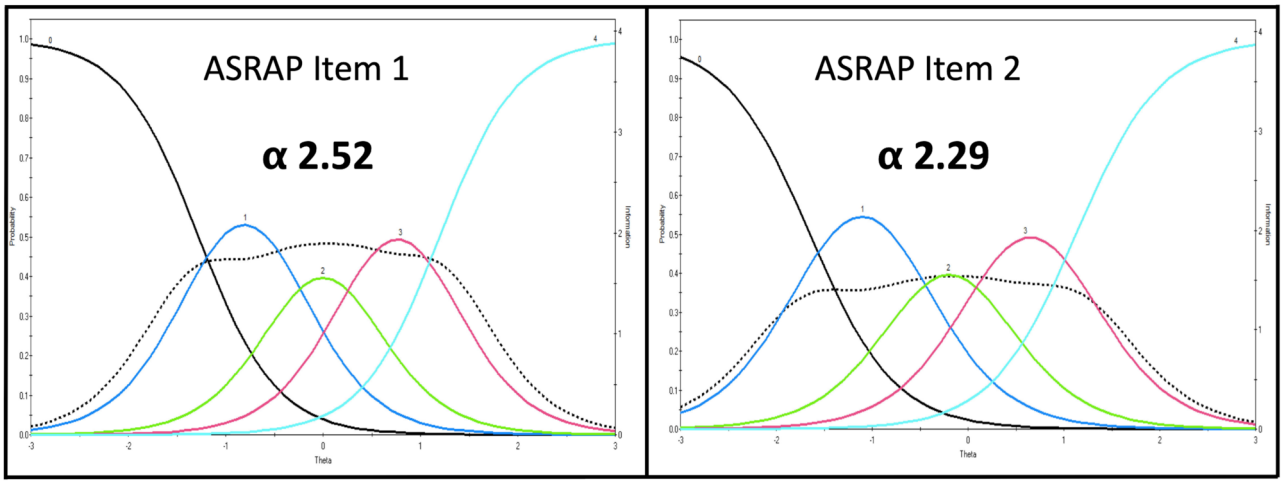
Figure 3. Test information of the 27-item long-form ASRAP questionnaire

A standard error of .30 corresponds approximately to a classical reliability of .90. The effective range of measurement for the final 27-item ASRAP item bank is estimated at -1.8 to +2.8 SDs.

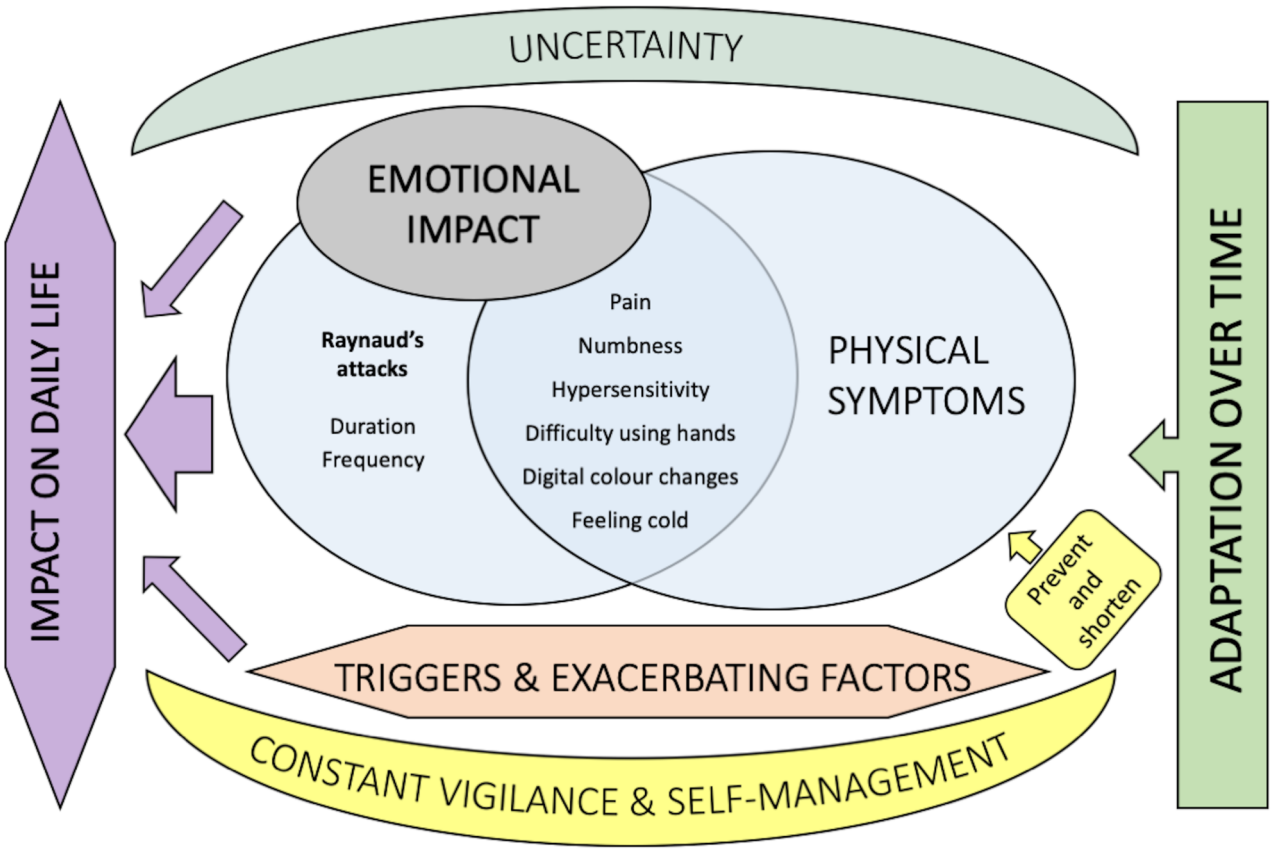
Supplementary Material 1. The provisional 39-item ASRAP questionnaire

Supplementary Material 2. Results of item modification following linguistic testing

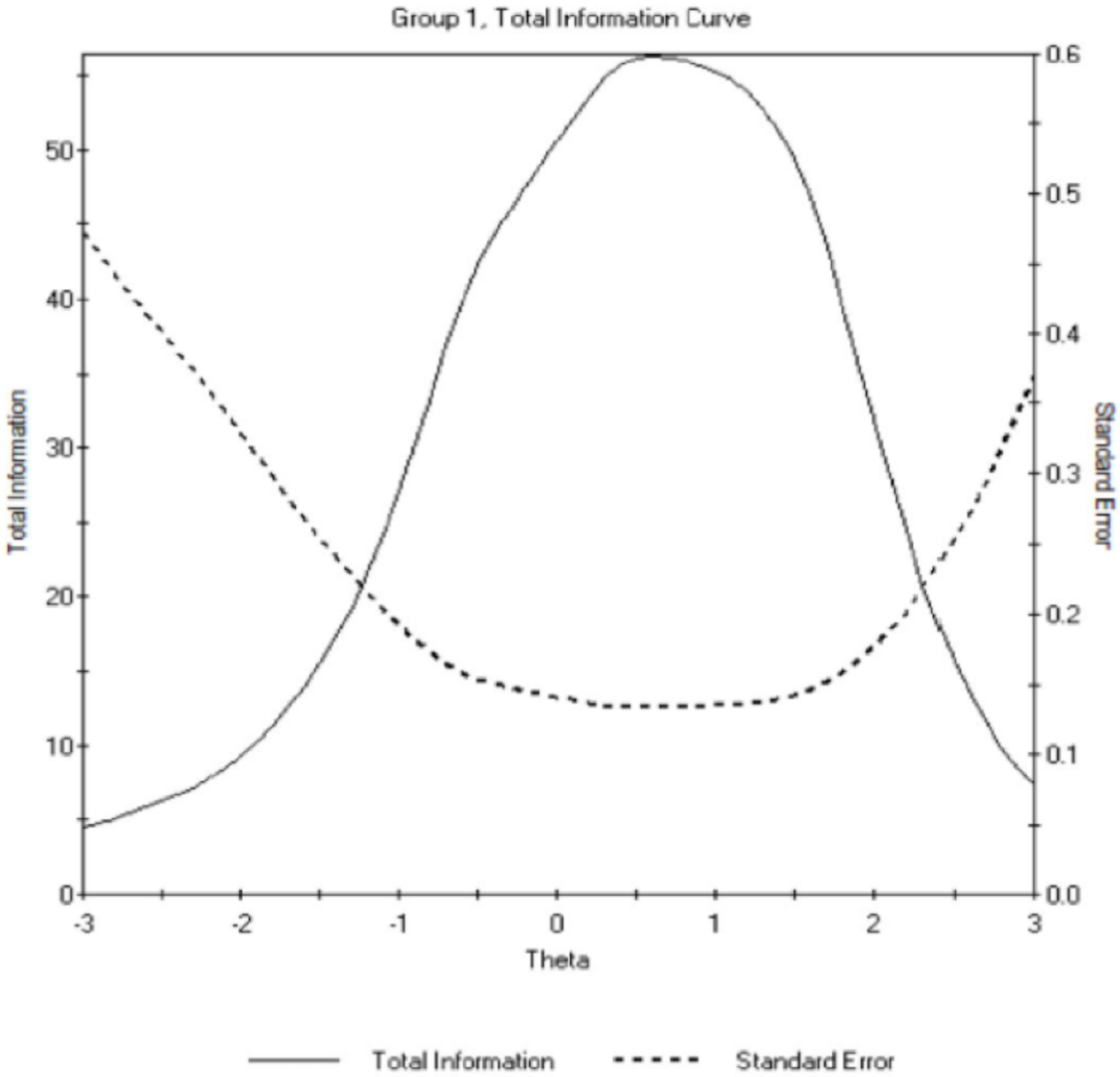
Supplementary Table 1. Local Dependency Marginal Chi Square analysis to remove redundant items



ACR_25038_ACR&R Figure 2.tiff



ACR_25038_FIGURE 1 AC&R ASRAP_cropped.tiff



ACR_25038_FIGURE 3 AC&R ASRAP_cropped.tiff