BMJ Open

Factor structure and convergent validity of the Derriford Appearance Scale-24 using standard scoring versus treating 'not applicable' responses as missing data: a Scleroderma Patient-centered Intervention Network (SPIN) cohort study

Erin L Merz,[1] Linda Kwakkenbos,[2,3,4] Marie-Eve Carrier,[2] Shadi Gholizadeh,[5] Sarah D Mills,[5] Rina S Fox,[5,6] Lisa R Jewett,[2,7] Heidi Williamson,[8] Diana Harcourt,[8] Shervin Assassi,[9] Daniel E Furst,[10] Karen Gottesman,[11] Maureen D Mayes,[9] Tim P Moss,[8] Brett D Thombs,[2,3,7,12,13] Vanessa L Malcarne,[5,14] on behalf of the SPIN Investigators

To cite: Merz EL, Kwakkenbos L, Carrier M-E, *et al*. Factor structure and convergent validity of the Derriford Appearance Scale-24 using standard scoring versus treating 'not applicable' responses as missing data: a Scleroderma Patient-centered Intervention Network (SPIN) cohort study. *BMJ Open* 2018;**8**:e018641. doi:10.1136/bmjopen-2017-018641

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-018641).

Received 11 July 2017
Revised 18 December 2017
Accepted 18 January 2018

For numbered affiliations see end of article.

Correspondence to
Dr Brett D Thombs;
brett.thombs@mcgill.ca

ABSTRACT

Objective Valid measures of appearance concern are needed in systemic sclerosis (SSc), a rare, disfiguring autoimmune disease. The Derriford Appearance Scale-24 (DAS-24) assesses appearance-related distress related to visible differences. There is uncertainty regarding its factor structure, possibly due to its scoring method.

Design Cross-sectional survey.

Setting Participants with SSc were recruited from 27 centres in Canada, the USA and the UK. Participants who self-identified as having visible differences were recruited from community and clinical settings in the UK.

Participants Two samples were analysed (n=950 participants with SSc; n=1265 participants with visible differences).

Primary and secondary outcome measures The DAS-24 factor structure was evaluated using two scoring methods. Convergent validity was evaluated with measures of social interaction anxiety, depression, fear of negative evaluation, social discomfort and dissatisfaction with appearance.

Results When items marked by respondents as 'not applicable' were scored as 0, per standard DAS-24 scoring, a one-factor model fit poorly; when treated as missing data, the one-factor model fit well. Convergent validity analyses revealed strong correlations that were similar across scoring methods.

Conclusions Treating 'not applicable' responses as missing improved the measurement model, but did not substantively influence practical inferences that can be drawn from DAS-24 scores. Indications of item redundancy and poorly performing items suggest that the DAS-24 could be improved and potentially shortened.

Strengths and limitations of this study

- ► This is the first study to evaluate concerns regarding the scoring methodology of items with a 'not applicable' response option in the Derriford Appearance Scale-24 (DAS-24).
- ► Two large samples of individuals with physical disfigurement were drawn from clinical and community settings.
- ► This study used confirmatory factor analysis, a contemporary and theory-driven approach to testing factor structure and goodness of fit, with a weighted least squares means and variance adjusted estimator which performs well with observed variables on an ordinal scale.
- ► The two samples did not include the convergent validity measures used in the original DAS-24 study.
- ► Convergent validity could not be compared between the two samples due to differences in the measures taken for each sample.

INTRODUCTION

Systemic sclerosis (SSc or scleroderma) is an autoimmune disease characterised by skin thickening, internal organ fibrosis, disability, significant appearance changes and disfigurement in highly visible parts of the body.[1] Common physical appearance concerns include hypopigmentation and hyperpigmentation of the skin, changes to skin texture, telangiectasia, hand contractures, sclerodactyly, calcinosis and altered facial features, including a pinched appearance to

the nose and eyes, loss of lip contour with thinning of the lips and a decreased ability to fully open the mouth. Treatments may also lead to appearance changes: certain immunosuppressant drugs can cause hair loss and long-term steroid use may contribute to a Cushingoid facial appearance.[2]

Visible differences in appearance are associated with distress, poorer self-esteem, body image problems and diminished quality of life.[3] In SSc, visible differences are common[4] and associated with greater body image dissatisfaction,[5–7] poorer psychosocial functioning[8] and increased anxiety and depression.[9] Thus, there is a significant need to study physical appearance concerns in SSc and to validate measures related to appearance in SSc.[10 11]

The Derriford Appearance Scale-24 (DAS-24),[12 13] the short form of the DAS-59,[14] measures distress and dysfunction related to physical appearance concerns in populations with visible differences such as SSc. The DAS-24 is hypothesised to yield a single factor representing appearance-related distress and dysfunction.[13] In an adult sample of 535 patients with appearance concerns, the DAS-24 total score demonstrated internal consistency reliability (α=0.92), concurrent validity with the DAS-59 criterion measure (r=0.88) and convergent validity (rs=0.45–0.66 with measures of anxiety, depression, social avoidance and distress, fear of negative evaluation, internalised shame).[12]

However, recent results from 1265 community and clinical participants with visible difference suggested that the DAS-24 may not be unidimensional.[15] Exploratory factor analysis with principal axis factoring and oblique rotation revealed two factors: *general self-consciousness* (18 items) and *sexual and bodily self-consciousness* (six items).[15] The interfactor correlation was not reported; it is presumed to be large given that internal consistency for the total score was high (α=0.93). The authors concluded that the DAS-24 may be conceptualised as either the original unidimensional total score or the two subscale scores.[15]

One possible explanation for the uncertain factor structure may relate to the DAS-24 scoring procedure. To complete the measure, respondents first identify and describe a part of their appearance that is concerning to them (referred to as their 'feature'). The respondent then rates each item on a 1–4 scale, with higher values indicating greater appearance-related distress and dysfunction.[13] Fourteen items include a 'not applicable' (NA) option (scored as 0) to accommodate situations that are not relevant to a respondent. For example, an unemployed respondent might choose NA for the work item. Total scores are calculated by summing the items.[13] Thus, respondents who marked many items as NA could receive a low total score, even if they report significant appearance-related distress and dysfunction on the other items.

At the broadest level, NA responses in surveys are best considered within the context of missing data.[16] Treating these responses as very low scores, as is the procedure for the DAS-24, could lead to factorial solutions that represent an artefact of the NA procedure itself. In the two-factor model reported by Moss *et al*,[15] the second factor (*sexual and bodily self-consciousness*) was composed of six items, which all had the NA option (*distressed at beach, avoid communal changing, avoid undressing with partner, distressed playing sports/games, distressed by clothing limitations, affects sex life*). Among these items,[15] the NA response was chosen an average of 26% of the time (range: 19%–36%). NA was only chosen an average of 13% of the time (for the items with the NA option) that loaded on the first factor of *general self-consciousness* (range: 6%–32%). It is also worth noting that several items without the NA option (eg, *distressed at reflection*) loaded on the first factor (*general self-consciousness*) but conceptually match the second factor (*sexual and bodily self-consciousness*). Thus, it is possible that the two-factor solution may at least partly reflect shared method variance, rather than the conceptual coherence of two factors.

Given these concerns, the aim of this study was to evaluate two different scoring methods to determine whether the standard scoring procedure affects the factor structure and convergent validity of DAS-24 scores. The first aim was to evaluate the one-factor model in a sample of patients with SSc and Moss *et al*'s visible difference sample[15] using the two scoring methods. In the *standard method*, as described in the DAS-24 manual[13] and existing studies, NA responses were scored as 0. In the *missing method*, NA responses were modelled as missing data. It was hypothesised that model fit for the one-factor model scored according to the *missing method* would be superior in both samples. The second aim was to evaluate the convergent validity of DAS-24 scores calculated via both scoring methods in the SSc sample. It was hypothesised that DAS-24 scores would have a strong association with satisfaction with appearance (rs≥0.5), moderate to strong associations with fear of negative evaluation and social interaction anxiety (r≈0.4–0.5) and a moderate association with depressive symptomatology (r≈0.3). It was also hypothesised that correlations would be significantly stronger when the *missing method* was used.

METHODS

Patients and procedure

SSc sample

Patients enrolled in the Scleroderma Patient-centred Intervention Network (SPIN) Cohort completed study questionnaires from April 2014 through April 2016. Research Ethics Committee approval of the Jewish General Hospital and by the Institutional Review Boards of each participating center was obtained. Patients were enrolled at 27 centres from Canada, the USA and the UK. Eligible participants were[1] classified as having SSc according to the 2013 American College of Rheumatology/European League Against Rheumatism classification criteria[17] as confirmed by a SPIN physician,[2] at least 18 years of age,[3] able to give informed consent,[4] fluent in English or French and[5] able to respond to questionnaires

via the Internet. Physicians or supervised nurse coordinators from SPIN recruiting sites invited eligible patients to participate, obtained written informed consent and completed a medical data form to initiate patient registration in the study. After registration, an automated welcome email was sent to patients with instructions to activate their SPIN account and complete the online questionnaires. SPIN participants complete outcome measures online on enrolment and subsequently every 3 months; the present study includes only the baseline assessment.

Visible difference sample

Data were obtained from the study by Moss *et al*.[15] Respondents were community and clinical participants who self-identified as being 'visibly different' recruited through advertisements, general practice physicians and outpatient medical clinics including prosthetics, dermatology, ophthalmology, general plastics, burn, ear/nose/throat clinics (including cleft lip and palate), oncology and laser treatment. Data were downloaded from https://peerj.com/articles/1070/#supplemental-information.

Main study measure (SSc sample and visible difference sample)

DAS-24[12 13] contains 24 items with a response scale from 1 to 4; 10 items have only this scale; 14 items also include a NA option (scored as 0). Total scores range from 10 to 96, with higher scores indicating a greater level of appearance-related distress and dysfunction. Good internal consistency reliability and validity has been reported.[13] In both of the present samples, internal consistency reliability calculated using the *standard method* was: SSc α=0.92; visible difference α=0.96.

Convergent validity measures (SSc sample)

Social Interaction Anxiety Scale-6 (SIAS-6)[18] is the six-item short form of the SIAS[19] and measures anxiety resulting from social interactions. Respondents rate statements (eg, *I feel tense if I am alone with just one person*) on a scale from 0 to 4. Total scores range from 0 to 24, with higher scores indicating greater social anxiety. Internal consistency reliability in the SSc sample was α=0.89.

Patient Health Questionnaire-8 (PHQ-8)[20] is an adapted version of the PHQ-9,[21] a measure of depression that has been validated for use in SSc.[22] Frequency of symptoms (eg, *Little interest or pleasure in doing things*) over the past 2 weeks is rated on a scale from 0 to 3. Total scores range from 0 to 24, with higher scores indicating more symptoms of depression. Internal consistency reliability in the SSc sample was α=0.89.

Brief Fear of Negative Evaluation Scale-Revised (BFNE-II)[23] is a 12-item short form of the BFNE[24] that assesses apprehension and distress of being negatively evaluated. Respondents rate statements (eg, *I worry about what kind of impression I make on people*) on a scale from 1 to 5. Total scores range from 12 to 60, with higher scores indicating greater fear of negative evaluation. Internal consistency reliability in the SSc sample was α=0.98.

Satisfaction with Appearance Scale (SWAP)[25] is a 14-item measure of satisfaction with appearance developed for burn survivors,[25] then adapted for SSc and validated.[6 9] Respondents rate statements about their physical appearance or social discomfort on a scale from 0 to 6. The SWAP yields two subscale scores: *social discomfort* (eg, *My appearance makes other people uncomfortable*) which reflects social discomfort related to body dissatisfaction (six items; range 0–36 with higher scores indicting greater discomfort) and *dissatisfaction with appearance* (eg, *I am satisfied with the appearance of my face*) which reflects dissatisfaction with various body parts (eight items; range 0–48 with higher scores indicating greater dissatisfaction). The SWAP has demonstrated good measurement properties in SSc.[7 9] Internal consistency reliability in the SSc sample was: *social discomfort* α=0.93, *dissatisfaction with appearance* α=0.92.

SSc measures (SSc sample)

Years since the first non-Raynaud's phenomenon symptoms was calculated by taking the deviation between the date that forms were completed online and the date of a patient's first non-Raynaud's symptom. This method is recommended as an approximation of the onset of SSc.[26]

The Modified Rodnan Skin Score (mRSS),[27] a clinician-administered indicator of skin disease severity, is calculated by measuring the extent and severity of skin thickening on 17 body surfaces by palpation on scale ranging from 0 to 3. Total scores range from 0 to 51 with higher scores indicating greater severity.

SSc disease subtype is classified by the pattern and severity of skin fibrosis.[28] The *limited* subtype is characterised by skin involvement distal to the elbows and knees, slower fibrosis and milder internal organ involvement. Limited patients without cutaneous involvement are classified as having *sine SSc*. The *diffuse* subtype is characterised by more extensive skin thickening involving proximal extremities and/or the trunk in addition to proximal involvement, rapidly progressing fibrosis and more significant internal organ involvement.

Appearance-related variables documented by study physicians included telangiectasia (facial or any), skin pigmentation changes (facial or any) and hand contractures. Telangiectasia involves dilated superficial blood vessels that appear as red lines or patches on the skin that collapse on pressure and fill slowly when pressure is released. Skin pigmentation changes were defined as hyperpigmentation or hypopigmentation. Disfigurement due to hand contractures (tightening around the proximal interphalangeal joints, metacarpals and/or wrists that limits range of motion) was defined as the percentage of limitation in range of motion, categorised as none/mild (0%–25%), moderate (25%–50%), or severe (>50%).

Data analysis

Scoring

DAS-24 scores were calculated using two methods. The *standard method* scored the 14 items with the NA option using the 0–4 metric described in the manual.[13] That is, 0 was

used when a person chose NA for an item and 1–4 indicated that a person had chosen that response option. The second method, the *missing method*, scored the 14 items with the NA option such that NA responses were treated as missing data (described below in the section Analytic Strategy) and 1–4 indicated that a person had chosen that response.

Analytic strategy

Confirmatory factor analysis (CFA) was conducted in Mplus V.7.2[29] to examine the one-factor model in the SSc and visible difference samples using both scoring methods. Weighted least squares means and variance adjusted (WLSMV) estimation, which is robust to non-normal and non-independent data, was used to handle the ordinal rating scale of the DAS-24. Missing data were handled using a variation of the full information maximum procedure employed for WLSMV estimation in Mplus which is analogous to pairwise present analysis.[29] Models were evaluated according to statistical and descriptive indices, given that χ^2 tests have significant limitations, including a high degree of dependence on sample size.[30] Following Bentler's[31] recommendations, the root mean square error of approximation (RMSEA)[32] and Comparative Fit Index (CFI)[33] were evaluated. Cut-off thresholds were based on commonly used recommendations.[34] For the RMSEA, values. 0.06–0.08 were interpreted as acceptable model fit, with values <0.06 indicating close fit. For the CFI, values. 0.90–0.95 were interpreted as acceptable model fit, with values >0.95 indicating close fit. Standardised factor loadings were evaluated and compared across the two scoring methods.

Convergent validity analyses of the DAS-24 were conducted using the SSc sample. In order to capture differences between scoring methods, two DAS-24 averages were computed for each respondent. To represent the *standard method*, values for all 24 items were totaled (NA responses were scored as 0) and divided by 24 to yield a *standard method average score* (possible range 0.42–4.00). To represent the *missing method*, values for all items for which a respondent chose a numeric value were totaled (10–24 items) and divided by the total number of items that were answered to yield a *missing method average score* (possible range 1.00–4.00). Bivariate correlations between both average scores (*standard* and *missing*) and each convergent validity measure (SIAS-6, PHQ-8, BFNE-II, SWAP) were evaluated in SPSS V.24.0.[35] Correlation coefficients for each scoring procedure were compared using methods described by Steiger[36] via an online calculator.[37]

RESULTS

Description of SSc and visible difference samples

Characteristics of the SSc sample (n=950) are reported in table 1. The majority of participants were female (87%), married (85%) and white (73%). Age ranged from 18 to 84 years (M=55.5). The average skin thickening (mRSS score) was 8.0 (range 0–48); this differed by disease subtype (diffuse M=13.6 (SD=10.4), limited/sine M=4.2 (SD=4.2)). Nearly three-quarters of the sample reported telangiectasia, one-third reported skin pigmentation changes and one-third reported hand contractures.

Detailed characteristics of the visible difference sample (n=1265; community n=614, clinic n=651) are available elsewhere.[15] The majority of participants were female (69%), married (62%) and white (81%). Age ranged from 18 to 91 years (M=47.3).

Description of DAS-24 scores: SSc sample and visible difference sample

DAS-24 scores were lower for the SSc (M=35.7, SD=13.7) than visible difference (M=41.7, SD=16.3) sample (Hedges' g=0.39). A description of the patterns of NA responses is available in table 2. In the SSc sample, the proportion of NA responses ranged from 11% to 49% per item; in the visible difference sample, the proportion of NA responses ranged from 6% to 36% per item. Item 5 (*self-consciousness affects work;* SSc=44%, visible difference=32%) and item 16 (*distressed playing sports/games*; SSc=49%, visible difference=36%) had the highest percentage of NA responses.

CFA: SSc sample

Overall fit for the one-factor model scored according to the *standard method* was not adequate ($\chi^2(252)=2415.6$, P<0.001; CFI=0.896, RMSEA=0.095; table 3). Standardised factor loadings (table 2) were statistically significant and ranged from small to large (λs=0.18–0.79, P<0.001). When the *missing method* was used, overall fit for the one-factor model was improved and good ($\chi^2(252)=1188.0$, P<0.001; CFI=0.958, RMSEA=0.063; table 3). Standardised factor loadings (table 2) were statistically significant and large (λs=0.62–0.84, P<0.001).

The factor loadings for several items with a high number of NA responses were larger when scored according to the *missing method*. These differences can be seen for item 5 (% missing=44%; *Standard* λ=0.45, *Missing* λ=0.80), item 6 (% missing=30%; *Standard* λ=0.48, *Missing* λ=0.77), item 12 (% missing=40%; *Standard* λ=0.36, *Missing* λ=0.64), item 15 (% missing=26%; *Standard* λ=0.44, *Missing* λ=0.72) and item 16 (% missing=49%; *Standard* λ=0.18, *Missing* λ=0.71).

CFA: visible difference sample

Overall fit for the one-factor model scored according to the *standard method* was not adequate ($\chi^2(252)=4593.3$, P<0.001; CFI=0.870, RMSEA=0.117; table 3). Standardised factor loadings (table 2) were statistically significant and ranged from moderate to large (λs=0.41–0.82, P<0.001). When the *missing method* was used, overall fit for the one-factor model was improved and good ($\chi^2(252)=2061.5$, P<0.001; CFI=0.952, RMSEA=0.075; table 3). Standardised factor loadings (table 2) were statistically significant and large (λs=0.62–0.84, P<0.001).

The factor loadings for several items with high a high number of NA responses were larger when scored according to the *missing method*. These differences can be seen for item 5 (% missing=32%; *Standard* λ=0.58, *Missing*

Table 1 Systemic sclerosis sample description (n=950)

	M±SD or n (%)
Demographic characteristics	
Age in years (n=947)	55.48±12.05
Language	
English	882 (92.8)
French	68 (7.2)
Sex	
Women	827 (87.1)
Men	123 (12.9)
Race/ethnicity	
White	807 (84.9)
Black	54 (5.7)
Other	88 (9.3)
Missing	1 (0.1)
Marital status	
Married/partnered	692 (72.8)
Never married	111 (11.7)
Divorced/separated	110 (11.6)
Widowed	37 (3.9)
Employment status	
Full-time employment	282 (29.7)
Part-time employment	103 (10.8)
Unemployed	123 (12.9)
Homemaker, student, other	439 (46.5)
Missing	1 (0.1)
Disease characteristics	
Years since first non-Raynaud's symptoms (n= 876)	11.75±8.90
Modified Rodnan Skin Score (n=759)	7.98±8.71
Disease subtype	
Diffuse	383 (40.3)
Limited/sine	558 (58.8)
Not available	9 (0.9)
Telangiectasia (any; n=950)	
Yes	677 (71.3)
No	257 (27.1)
Not available	16 (1.7)
Telangiectasia (face; n=677)	
Yes	452 (66.8)
No	75 (11.1)
Not available	150 (22.1)
Pigmentation changes (any; n=950)	
Yes	285 (30.0)
No	600 (63.2)
Not available	65 (6.8)
Pigmentation changes (face; n=285)	
Yes	147 (51.6)
No	94 (33.0)
Not available	44 (15.4)
Hand contractures (n=950)	
No/mild (0%–25%)	674 (70.9)
Moderate (25%–50%)	172 (18.1)
Severe (>50%)	50 (5.3)
Not available	54 (5.7)

λ=0.77), item 15 (% missing=28%; *Standard* λ=0.41, *Missing* λ=0.72) and item 16 (% missing=36%; *Standard* λ=0.43, *Missing* λ=0.74).

Convergent validity: SSc sample

Correlations describing the relationship between average DAS-24 scores (*standard* and *missing*) and scores of social interaction anxiety, depression, fear of negative evaluation, social discomfort and dissatisfaction with appearance are in table 4. Correlations for the *standard method* average score were large and statistically significant (rs=0.44–0.69). Correlations for the *missing method* average score were large and statistically significant (rs=0.47–0.72). In all but one case, the *missing method* correlations were significantly larger than the *standard method* correlations, but the magnitude of differences were very small (0.02–0.07).

DISCUSSION

Confirmatory factor analyses revealed that a one-factor DAS-24 model fit poorly when using the *standard method*, but that the one-factor model fit well when using the *missing method* in both the SSc and visible difference samples. Correlations with measures of convergent validity calculated using the *missing method* were significantly larger than those calculated using the *standard method*, but the implications for inferences drawn from DAS-24 scores are likely negligible given the small magnitude of differences.

The finding that a one-factor model fit well when items were scored in a manner more logically consistent with their intent (*missing method*) supports the original conceptualisation of the DAS-24 as representing a single latent construct of distress and dysfunction related to appearance.[13] This also adds to a growing literature suggesting that model-based missingness methods are the most suitable solution for dealing with NA responses.[16] On one hand, this would suggest that parameter estimates for models based on the *standard method* are misleading and that use of the DAS-24, which includes items that are not relevant to some respondents, should be discontinued. On the other hand,

Table 2 Data missing due to 'not applicable' and standardised factor loadings for the one-factor model of the Derriford Appearance Scale-24 using the SM and MM for the SSc and visible difference samples

	SSc			Visible difference		
Item	Not applicable, n (%)	SM λ	MM λ	Not applicable, n (%)	SM λ	MM λ
1. Feeling confident	–	0.65	0.68	–	0.68	0.71
2. Distressed at reflection	–	0.77	0.77	–	0.77	0.79
3. Irritable at home*	207 (21.8)	0.76	0.76	204 (16.1)	0.75	0.80
4. Feel hurt†	–	0.78	0.79	77 (6.1)	0.78	0.81
5. Self-consciousness affects work*	415 (43.7)	0.45	0.80	401 (31.7)	0.58	0.77
6. Distressed at beach*	289 (30.4)	0.48	0.77	241 (19.1)	0.64	0.77
7. Misjudged due to appearance*	260 (27.4)	0.65	0.69	237 (18.7)	0.60	0.65
8. Feel feminine/masculine	–	0.57	0.62	–	0.57	0.62
9. Self-conscious of appearance*	128 (13.5)	0.73	0.71	79 (6.2)	0.75	0.77
10. Feel irritable	–	0.73	0.73	–	0.72	0.74
11. Adopt concealing gestures	–	0.68	0.65	–	0.68	0.66
12. Avoid communal changing*	377 (39.7)	0.36	0.64	248 (19.6)	0.55	0.67
13. Distressed in supermarkets/department stores*	113 (11.9)	0.65	0.73	101 (8.0)	0.66	0.74
14. Feel rejected	–	0.79	0.79	–	0.82	0.82
15. Avoid undressing with partner*	249 (26.2)	0.44	0.72	351 (27.7)	0.41	0.72
16. Distressed playing sports/games*	462 (48.6)	0.18	0.71	460 (36.4)	0.43	0.74
17. Close into shell	–	0.77	0.79	–	0.74	0.75
18. Distressed by clothing limitations*	285 (30.0)	0.67	0.73	343 (27.1)	0.62	0.74
19. Distressed at social events*	104 (10.9)	0.76	0.84	85 (6.7)	0.75	0.84
20. Feel normal	–	0.73	0.76	–	0.73	0.76
21. Affects sex life*	266 (28.0)	0.50	0.76	321 (25.9)	0.56	0.77
22. Avoid leaving house	–	0.73	0.76	–	0.75	0.78
23. Distressed at others remarks about appearance*	312 (32.8)	0.65	0.71	212 (16.8)	0.67	0.75
24. Avoid pubs/restaurants*	137 (14.4)	0.62	0.78	98 (7.7)	0.59	0.78

*Items with the 'not applicable' option.
†This item contained '0/not applicable' responses in the visible difference data but was scored using the 1–4 scale only in the SSc sample; ps for all factor loadings <0.001.
MM, missing method ('not applicable' scored as missing); SM, standard method ('not applicable' scored as 0); SSc, systemic sclerosis.

convergent validity coefficients were similar across scoring methods, suggesting that use of the existing version of the DAS-24, despite its less optimal structure and scoring method, may not substantively impair its utility. Although factor structure is an important component of construct validity, it should be considered in the context of all validity evidence and, in particular, the extent to which a measure's internal structure affects criterion validity and its practical application.[38] Thus, it appears that scoring NA items as 0 yields

Table 3 Goodness of fit statistics for one-factor models of the DAS-24

Data	Scoring method	χ^2	df	P value	CFI	RMSEA (90% CI)
SSc	Standard	2415.63	252	<0.001	0.896	0.095 (0.092–0. 099)
	Missing	1188.01	252	<0.001	0.958	0.063 (0.059–0. 066)
Visible difference	Standard	4593.25	252	<0.001	0.870	0.117 (0.114–0. 120)
	Missing	2061.49	252	<0.001	0.952	0.075 (0.072–0. 078)

Note. Standard method: 'not applicable' scored as 0; missing method: 'not applicable' scored as missing.
CFI, Comparative Fit Index; DAS-24, Derriford Appearance Scale-24; RMSEA, root mean square error of approximation.

Table 4 Correlations of *standard* and *missing method* average Derriford Appearance Scale-24 (DAS-24) scores with convergent validity measures and evaluation of significant differences between correlations

	r		
Validity measure	Standard	Missing	P value
Social Interaction Anxiety score (total score)	0.51	0.53	0.012
Depressive symptoms (Patient Health Questionnaire-8 total score)	0.65	0.72	<0.001
Fear of negative evaluation (Brief Fear of Negative Evaluation-Revised total score)	0.62	0.62	0.916
Social discomfort (Satisfaction with Appearance Scale (SWAP) subscale score)	0.69	0.71	0.001
Dissatisfaction with appearance (SWAP subscale score)	0.44	0.47	0.027

Note. ps<0.001 for all rs.

statistical problems when models are judged by stringent criteria, but that the implications for the overall validity of DAS-24 scores using the *standard method* may be of less consequence.

The DAS-24's reasonably strong functional performance, as assessed via convergent validity analyses suggests that there is likely excessive redundancy within the measure. When items rated as NA were modelled using the *standard method*, factor loadings were substantially impaired, and, for many items, clearly unacceptable. Nonetheless, convergent validity was only minimally reduced, and internal consistency reliability was very high (SSc α=0.92; visible difference α=0.96). It is well-established that very high Cronbach's alphas (eg, α >0.90) tend to reflect overly redundant items.[39–42] Thus, it seems that there is a core group of DAS-24 items that function well, and the inclusion of unnecessary items and even problematically scored items does not have a substantive negative impact on the practical performance of the measure or clinical inferences that can be drawn from it.

Nonetheless, the problems of unnecessary item redundancy and problematically scored items are concerns in research and clinical contexts because they reduce measurement precision and increase survey length and time burden on in populations with physical appearance concerns, like SSc. These concerns are particularly salient in collaborative research contexts where medical and patient-reported outcome data are collected from large numbers of patients across span countries, languages and clinical settings, as is the case with the SPIN cohort. Research on the DAS-24 should focus on reducing the number of items and addressing existing scoring issues with the goal of maximising item parsimony and retaining the full construct domain of distress and dysfunction related to concerns with physical appearance as previously defined.[12 13] Recently, optimal test assembly methods have been used to shorten measures based on objective, prespecified criteria,[43] and these methods could be used to identify a shorter version of the DAS-24 that performs similarly and eliminates its problematic scoring.

This study contributes to the growing literature on the DAS-24 using two large samples composed of respondents with diverse physical appearance concerns that are generalisable to other visible difference populations. There are also limitations that should be considered in the context of the current results. First, the SSc data did not contain the convergent validity measures used in the original DAS-24 validation study, and thus measures of closely related constructs were used as proxies. Second, convergent validity analyses could not be conducted in the visible difference sample as these measures were not available.

In sum, these findings suggest that scoring the DAS-24 according to the *standard method* leads to a suboptimal measurement model. The *missing method* yields a superior one-factor structure but does not lead to substantively different relationships with theoretically related constructs. Thus, while it is preferable to treat NA responses as missing, particularly in research settings, it is unlikely that using the *standard method* greatly impacts the practical interpretation of DAS-24 scores. The development of a short form that removes item redundancy while retaining adequate psychometric properties is needed and would particularly benefit research in populations with visible differences in appearance such as patients with SSc.

Author affiliations
[1]Department of Psychology, California State University, Dominguez Hills, Carson, California, USA
[2]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada
[3]Department of Psychiatry, McGill University, Montreal, Quebec, Canada
[4]Behavioral Science Institute, Clinical Psychology, Radboud University, Nijmegen, The Netherlands
[5]San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, California, USA
[6]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA
[7]Department of Educational and Counseling Psychology, McGill University, Montreal, Quebec, Canada
[8]Centre for Appearance Research, Faculty of Health and Life Science, University of the West of England, Bristol, UK
[9]Department of Internal Medicine, Division of Rheumatology, University of Texas McGovern Medical School, Houston, Texas, USA
[10]Division of Rheumatology, Geffen School of Medicine at the University of California, Los Angeles, California, USA
[11]Scleroderma Foundation, Los Angeles, California, USA
[12]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada
[13]Department of Medicine, McGill University, Montreal, Quebec, Canada
[14]Department of Psychology, San Diego State University, San Diego, California, USA

Collaborators SPIN Investigators: Murray Baron, McGill University, Montreal, Quebec, Canada; Susan J Bartlett, McGill University, Montreal, Quebec, Canada; Frank van den Hoogen, Radboud University Medical Center and Sint Maartenskliniek, Nijmegen, The Netherlands; Luc Mouthon, Université Paris Descartes, Paris, France; Warren R Nielson, St. Joseph's Health Care, London, Ontario, Canada; Robert Riggs, Scleroderma Foundation, Danvers, Massachusetts, USA; Maureen Sauve, Scleroderma Society of Ontario, Hamilton, Ontario; Fredrick Wigley, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; Isabelle Boutron, Université Paris Descartes, and Assistance Publique Hôpitaux de Paris, Paris, France; Angela Costa Maia, University of Minho, Braga, Portugal; Ghassan El-Baalbaki, Université du Québec à Montréal, Montreal, Quebec, Canada; Carolyn Ells, McGill University, Montreal, Quebec, Canada; Cornelia van den Ende, Sint Maartenskliniek, Nijmegen, The Netherlands; Kim Fligelstone, Scleroderma Society, London, UK; Catherine Fortune, Scleroderma Society of Ontario, Hamilton, Ontario, Canada; Tracy Frech, University of Utah, Salt Lake City, Utah, USA; Dominique Godard, Association des Sclérodermiques de France, Sorel-Moussel, France; Daphna Harel, New York University, New York, New York, USA; Marie Hudson, McGill University, Montreal, Quebec, Canada; Ann Impens, Midwestern University, Downers Grove, Illinois, USA; Yeona Jang, McGill University, Montreal, Quebec, Canada; Sindhu R Johnson, Toronto Scleroderma Program, Mount Sinai Hospital, Toronto Western Hospital, and University of Toronto, Toronto, Ontario, Canada; Ann Tyrell Kennedy, Federation of European Scleroderma Associations, Dublin, Ireland; Annett Körner, McGill University, Montreal, Quebec, Canada; Maggie Larche, McMaster University, Hamilton, Ontario, Canada; Catarina Leite, University of Minho, Braga, Portugal; Carlo Marra, Memorial University, St. John's, Newfoundland, Canada; Karen Nielsen, Scleroderma Society of Ontario, Hamilton, Ontario, Canada; Janet Pope, University of Western Ontario, London, Ontario, Canada; Alexandra Portales, Asociación Española de Esclerodermia, Madrid, Spain; Tatiana Sofia Rodriguez Reyna, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico; Anne A Schouffoer, Leiden University Medical Center, Leiden, The Netherlands; Russell J Steele, Jewish General Hospital and McGill University, Montreal, Quebec, Canada; Maria E Suarez-Almazor, University of Texas MD Anderson Cancer Center, Houston, Texas, USA; Joep Welling, NVLE Dutch patient organisation for systemic autoimmune diseases, Utrecht, The Netherlands; Durhane Wong-Rieger, Canadian Organisation for Rare Disorders, Toronto, Ontario, Canada; Christian Agard, Centre Hospitalier Universitaire-Hôtel-Dieu de Nantes, Nantes, France; Alexandra Albert, Université Laval, Quebec, Quebec, Canada; Marc André,Centre Hospitalier Universitaire Gabriel-Montpied, Clermont-Ferrand, France; Guylaine Arsenault, Université de Sherbrooke , Sherbrooke, Quebec, Canada; Nouria Benmostefa, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Ilham Benzida, Assistance Publique Hôpitaux de Paris-Hôpital St-Louis, Paris, France; Sabine Berthier, Centre Hospitalier Universitaire Dijon Bourgogne, Dijon, France; Lyne Bissonnette, Université de Sherbrooke , Sherbrooke, Quebec, Canada; Gilles Boire, Université de Sherbrooke, Sherbrooke, Quebec, Canada; Alessandra Bruns, Université de Sherbrooke, Sherbrooke, Quebec, Canada; Patricia Carreira, Servicio de Reumatologia del Hospital 12 de Octubre, Madrid, Spain; Marion Casadevall, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Benjamin Chaigne, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Lorinda Chung, Stanford University, Stanford, California, USA; Pascal Cohen, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Pierre Dagenais, Université de Sherbrooke, Sherbrooke, Quebec, Canada; Christopher Denton, Royal Free London Hospital, London, UK; Robyn Domsic, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; James V Dunne, St. Paul's Hospital and University of British Columbia, Vancouver, British Columbia, Canada; Dominique Farge-Bancel, Assistance Publique Hôpitaux de Paris-Hôpital St-Louis, Paris, France; Paul Fortin, Université Laval, Quebec, Quebec, Canada; Anna Gill, Royal Free London Hospital, London, UK; Jessica Gordon, Hospital for Special Surgery, New York City, New York, USA; Brigitte Granel-Rey, Aix Marseille Université, and Assistance Publique Hôpitaux de Marseille-Hôpital Nord, Marseille, France; Claire Grange, Centre Hospitalier Lyon Sud, Lyon, France; Genevieve Gyger, Jewish General Hospital and McGill University, Montreal, Quebec, Canada; Eric Hachulla, Centre Hospitalier Régional Universitaire de Lille -Hôpital Claude Huriez, Lille, France; Pierre-Yves Hatron, Centre Hospitalier Régional Universitaire de Lille-Hôpital Claude Huriez, Lille, France; Ariane L Herrick, University of Manchester, Salford Royal NHS Foundation Trust, Manchester, UK; Adrian Hij, Assistance Publique Hôpitaux de Paris-HôpitalSt-Louis, Paris, France; Monique Hinchcliff, Northwestern University, Chicago, Illinois, USA; Alena Ikic, Université Laval, Quebec, Quebec, Canada; Niall Jones, University of Alberta, Edmonton, Alberta, Canada; Artur Jose de B Fernandes, Université de Sherbrooke, Sherbrooke, Quebec, Canada; Suzanne Kafaja, University of California, Los Angeles, California, USA; Nader Khalidi, McMaster University, Hamilton, Ontario, Canada; Benjamin Korman, Northwestern University, Chicago, Illinois, Marc Lambert, Centre Hospitalier Régional Universitaire de Lille-Hôpital Claude Huriez, Lille, France; David Launay, Centre Hospitalier Régional Universitaire de Lille-Hôpital Claude Huriez, Lille, France; USA; Patrick Liang, Université de Sherbrooke, Sherbrooke, Quebec, Canada; Jonathan London, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Joanne Manning, Salford Royal NHS Foundation Trust, Salford, UK; Thierry Martin, Les Hôpitaux Universitaires de Strasbourg-Nouvel Hôpital Civil, Strasbourg, France; Ariel Masetto, Université de Sherbrooke, Sherbrooke, Quebec, Canada; Arsene Mekinian, Assistance Publique Hôpitaux de Paris-Hôpital St-Antoine, Paris, France; Romain Paule, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Alexis Regent, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Sébastien Rivière, Assistance Publique Hôpitaux de Paris-Hôpital St-Antoine, Paris, France; David Robinson, University of Manitoba, Winnipeg, Manitoba, Canada; Sophie Roux, Université de Sherbrooke, Sherbrooke, Quebec, Canada; Perrine Smets, Centre Hospitalier Universitaire Gabriel-Montpied, Clermont-Ferrand, France; Doug Smith, University of Ottawa, Ottawa, Ontario, Canada; Vincent Sobanski, Centre Hospitalier Régional Universitaire de Lille-Hôpital Claude Huriez, Lille, France; Robert Spiera, Hospital for Special Surgery, New York, New York, USA; Virginia Steen, Georgetown University, Washington, DC, USA; Evelyn Sutton, Dalhousie University, Halifax, Nova Scotia, Canada; Benjamin Terrier, Assistance Publique Hôpitaux de Paris-Hôpital Cochin, Paris, France; Carter Thorne, Southlake Regional Health Centre, Newmarket, Ontario, Canada; John Varga, Northwestern University, Chicago, Illinois, USA; Pearce Wilcox, St. Paul's Hospital and University of British Columbia, Vancouver, British Columbia, Canada; Vanessa C Delisle, Jewish General Hospital and McGill University, Montreal, Quebec, Canada; Brooke Levis, Jewish General Hospital and McGill University, Montreal, Quebec, Canada; Mia R Pepin, Jewish General Hospital, Montreal, Quebec, Canada.

Contributors ELM, LK, BDT and VLM were responsible for the study concept and design, data analysis and interpretation and drafting the manuscript. LK, MEC, SG, SDM, RSF, LRJ, HW, DH, SA, DEF, KG, MDM, TPM, BDT, VLM and SPIN were responsible for data collection. All authors were responsible for the critical revision and final approval of the manuscript. ELM is the guarantor.

Funding The Scleroderma Patient-centered Intervention Network (SPIN) is funded by the Canadian Institutes of Health Research (CIHR; PI=Thombs, TR3-119192; PI=Thombs, PJT- 148504; PIs=Thombs, Mouthon, Poiraudeau, PJT-149073) and the Arthritis Society (SOG-16- 380, PI=Thombs). In addition, SPIN has received institutional contributions from the Lady Davis Institute for Medical Research of the Jewish General Hospital, Montreal, Quebec, Canada and from McGill University, Montreal, Canada. SPIN has also received support from the Scleroderma Society of Ontario, Scleroderma Canada and Sclérodermie Quebec. LK was supported by a CIHR Banting Postdoctoral Fellowship. RSF was supported by a National Cancer Institute Training Grant (5T32CA193193). LRJ was supported by a CIHR Doctoral Research Award. BDT was supported by an Investigator Salary Award from the Arthritis Society.

Competing interests The authors have read and understood the BMJ policy on declaration of interests and declare the following interests: the DAS-24 is commercially available from TPM.

Patient consent Detail has been removed from this case description/these case descriptions to ensure anonymity. The editors and reviewers have seen the detailed information available and are satisfied that the information backs up the case the authors are making.

Ethics approval Research Ethics Committee of the Jewish General Hospital, Montreal, Canada and by the Institutional Reviews Boards of each participating center.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Anonymised patient level DAS-24 data from the SPIN sample are evaluated by the SPIN steering committee and are available on reasonable request. Anonymised patient level DAS-24 data from the Visible Difference sample (15) are available from the Peer journal website at https://peerj.com/articles/1070/#supplemental-information.

REFERENCES

1. Medsger TA. Natural history of systemic sclerosis and the assessment of disease activity, severity, functional status, and psychologic well-being. *Rheum Dis Clin North Am* 2003;29:255–73.
2. Gholizadeh S, Fox RS, Mills SD. *et al*Coping with the disfigurement of scleroderma: facial, skin, and hand Changes. In: Varga J, Denton CP, Wigley FM, . eds. *Scleroderma*. New York: Springer, 2017:713–21.
3. Rumsey N, Harcourt D. Body image and disfigurement: issues and interventions. *Body Image* 2004;1:83–97.
4. Jewett LR, Kwakkenbos L, Carrier ME, *et al*. Examination of the association of sex and race/ethnicity with appearance concerns: a Scleroderma Patient-centered Intervention Network (SPIN) Cohort study. *Clin Exp Rheumatol* 2016;34(Suppl 100):92–9.
5. Benrud-Larson LM, Heinberg LJ, Boling C, *et al*. Body image dissatisfaction among women with scleroderma: extent and relationship to psychosocial function. *Health Psychol* 2003;22:130–9.
6. Heinberg LJ, Kudel I, White B, *et al*. Assessing body image in patients with systemic sclerosis (scleroderma): validation of the adapted satisfaction with appearance scale. *Body Image* 2007;4:79–86.
7. Mills SD, Fox RS, Merz EL, *et al*. Evaluation of the satisfaction with appearance scale and its short form in systemic sclerosis: analysis from the UCLA scleroderma quality of life study. *J Rheumatol* 2015;42:1624–30.
8. Malcarne VL, Hansdottir I, Greenbergs HL, *et al*. Appearance self-esteem in systemic sclerosis. *Cognit Ther Res* 1999;23:197–208.
9. Jewett LR, Hudson M, Haythornthwaite JA, *et al*. Development and validation of the brief-satisfaction with appearance scale for systemic sclerosis. *Arthritis Care Res* 2010;62:1779–86.
10. Malcarne VL, Fox RS, Mills SD, *et al*. Psychosocial aspects of systemic sclerosis. *Curr Opin Rheumatol* 2013;25:707–13.
11. Thombs BD, van Lankveld W, Bassel M, *et al*. Psychological health and well-being in systemic sclerosis: State of the science and consensus research agenda. *Arthritis Care Res* 2010;62:1181–9.
12. Carr T, Moss T, Harris D. The DAS24: a short form of the Derriford Appearance Scale DAS59 to measure individual responses to living with problems of appearance. *Br J Health Psychol* 2005;10:285–98.
13. Moss T, Harris D, Carr T. *Manual for the Derriford Appearance Scale 24 (DAS24)*. Bradford on Avon: Musketeer Press, 2004.
14. Carr T, Harris D, James C. The Derriford Appearance Scale (DAS-59): a new scale to measure individual responses to living with problems of appearance. *Br J Health Psychol* 2000;5:201–15.
15. Moss TP, Lawson V, White P. Identification of the underlying factor structure of the derriford appearance scale 24. *PeerJ* 2015;3:e1070.
16. Holman R, Glas CA, Lindeboom R, *et al*. Practical methods for dealing with 'not applicable' item responses in the AMC Linear Disability Score project. *Health Qual Life Outcomes* 2004;2:29.
17. 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. *Ann Rheum Dis* 2013;72:1747–55.
18. Peters L, Sunderland M, Andrews G, *et al*. Development of a short form Social Interaction Anxiety (SIAS) and Social Phobia Scale (SPS) using nonparametric item response theory: the SIAS-6 and the SPS-6. *Psychol Assess* 2012;24:66–76.
19. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther* 1998;36:455–70.
20. Kroenke K, Strine TW, Spitzer RL, *et al*. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord* 2009;114:163–73.
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
22. Milette K, Hudson M, Baron M, *et al*. Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: internal consistency reliability, convergent validity and clinical correlates. *Rheumatology* 2010;49:789–96.
23. Carleton RN, McCreary DR, Norton PJ, *et al*. Brief fear of negative evaluation scale-revised. *Depress Anxiety* 2006;23:297–303.
24. Leary MR. A brief version of the fear of negative evaluation scale. *Pers Soc Psychol Bull* 1983;9:371–5.
25. Lawrence JW, Heinberg LJ, Roca R, *et al*. Development and validation of the satisfaction with appearance scale: assessing body image among burn-injured patients. *Psychol Assess* 1998;10:64–70.
26. Tashkin DP, Elashoff R, Clements PJ, *et al*. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006;354:2655–66.
27. Kahaleh MB, Sultany GL, Smith EA, *et al*. A modified scleroderma skin scoring method. *Clin Exp Rheumatol* 1986;4:367–9.
28. Hachulla E, Launay D. Diagnosis and classification of systemic sclerosis. *Clin Rev Allergy Immunol* 2011;40:78–83.
29. Muthén LK, Muthén BO. *Mplus user's guide*. 7th edn. Los Angeles, 2015.
30. Hoyle RH. Confirmatory factor analysis. In: Tinsley HEA, Brown SD, eds. *Handbook of applied multivariate statistics and mathematical modeling*. San Diego: Academic Press, 2000:465–97.
31. Bentler PM. On tests and indices for evaluating structural models. *Pers Individ Dif* 2007;42:825–9.
32. Steiger JH. Structural model evaluation and modification: an interval estimation approach. *Multivariate Behav Res* 1990;25:173–80.
33. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;107:238–46.
34. Hu Li-tze, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 1999;6:1–55.
35. IBM. *IBM SPSS statistics*. Armonk, NY: IBM Corporation, 2016.
36. Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull* 1980;87:245–51.
37. Lee IA, Preacher KJ. Calculation for the test of the difference between two dependent correlations with one variable in common [computer software]. 2013. http://quantpsy.org
38. Hopwood CJ, Donnellan MB. How should the internal structure of personality inventories be evaluated? *Pers Soc Psychol Rev* 2010;14:332–46.
39. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011;2:53–5.
40. Boyle GJ. Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Pers Individ Dif* 1991;12:291–4.
41. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
42. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess* 2003;80:99–103.
43. Levis AW, Harel D, Kwakkenbos L, *et al*. Using optimal test assembly methods for shortening patient-reported outcome measures: development and validation of the cochin hand function scale-6: a scleroderma patient-centered intervention network cohort study. *Arthritis Care Res* 2016;68:1704–13.