

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	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
AUTHORS	Merz, Erin; Kwakkenbos, Linda; Carrier, Marie-Eve; Gholizadeh, Shadi; Mills, Sarah; Fox, Rina; Jewett, Lisa; Williamson, Heidi; Harcourt, Diana; Assassi, Shervin; Furst, Daniel; Gottesman, Karen; Mayes, Maureen; Moss, Tim; Thombs, Brett; Malcarne, Vanessa

VERSION 1 – REVIEW

REVIEWER	Jan Passchier VU University Amsterdam, The Netherlands
REVIEW RETURNED	07-Sep-2017

GENERAL COMMENTS	<p>The manuscript regards the differences between two methods of scoring of the Derriford Appearance Scale-24: one treating the ‘not applicable’ (NA) responses as standard and one treating it as missing data. This study has been carried out with two large samples of individuals with physical disfigurement (one clinical and one from the community).</p> <p>While the question if ‘NA’ should be treated as missing seems to be an important point , I doubt if this specific population and this specific scale are suitable to arrive at conclusions which can be generalized to the general population and other scales as well. If not, I wonder if this investigation is interesting enough for a broader public than those who use this scale and population. It suggests that the study is more data-driven that designed for the research question concerned. Anyway, the authors do not discuss this point.</p> <p>The introduction part is focused on the DAS-24, its scoring and its factor structure. A more broad elaboration of the theoretical notions of ‘NA’ versus ‘missing’ is lacking. F.i: if a subject is stating that a problem area is not applicable: does not that imply that for him/her this problem does not exist; in other words that the stress/dysfunction concerned is zero (like the standard way of scoring implies too)? Treating it as ‘missing’, ignores this outcome but indicates that there seems to be a problem which the subject forgot/does not want to reveal, etc..</p>
-------------------------	--

	<p>Previous discussions in the literature on the NA phenomenon (f.i. Holman et al, Health Qual Life Outcomes, 2004;2;29) are missing too.</p> <p>In the Methods parts, the 'missing method' way of scoring is not clear. Which type of missing value imputation is used: complete case, LOCF, regression analysis? One might come to different conclusion, depending on the way the missings are handled. Hypotheses about associations of the DAS-24 are stated without explicit reasons. The sample size appears to be sufficient for the statistical analysis.</p> <p>The Results section again shows how specific this (mainly female and advanced age) sample is. The many dermatological sub categories are probably mainly interesting for experts in that field. It is concluded that the factor loadings under the Missing Method scoring are larger than under the NA scoring. However, this is understandable because in the authors only use the 1st factor for determining the loadings, while in the NA scoring, two factors are found. So, it is an 'unfair' comparison. I expect that in the NA scoring the variance explained by the items for the two factors together will be larger than for one factor only; and perhaps larger than the one factor loading in the Missing Method scoring.</p> <p>In the Discussion, the authors appear to consider the one factor solution of the DAS-24 better ('logically more consistent') than the two factor solution because of the original conceptualization of the DAS-24 as representing a single latent construct. One may alternatively consider the two factor outcome from the NA scoring as a falsification of this conceptualization.</p> <p>A final word about the 5 pages with authors and affiliations. I was a bit surprised to see that so many authors (approaching 100) were involved in this article of limited scope and cannot imagine that each of them delivered important input for the article, apart from supplying a number of patients. I wonder if their contribution is sufficient for an authorship, or will be more qualified for a mentioning in the acknowledgment sections. But it is up to the editors' policy to decide upon this point.</p>
--	---

REVIEWER	Prof. Dr. Noha A. Azab Rheumatology and Rehabilitation Department, Faculty of Medicine, Cairo University, Egypt
REVIEW RETURNED	24-Nov-2017

GENERAL COMMENTS	<p>This study was carried out to investigate the possible influence of "not applicable" responses in The Derriford Appearance Scale-24 (DAS-24) in patients with SSc and volunteers with general visible differences using two methods; the standard method and the missing method. The study is very impressive and very well written , however there are few comments</p> <ul style="list-style-type: none"> • Page 9 line 4: the part of "STRENGTHS AND LIMITATIONS" is better moved to the last section of the discussion • Page 13 line 10: Why was the submission online and repeated every 3 months, this needs to be more clarified, wouldn't it be better if the form was submitted in the center with the investigator? <p>Finally ,I would like to thank the authors for this great work .</p>
-------------------------	---

REVIEWER	Anna Hoffmann-Vold Oslo University Hospital, Norway
REVIEW RETURNED	25-Nov-2017

GENERAL COMMENTS	<p>The authors address the important fact of changes in visual appearance in SSc patients and point out the problems associated with the existing assessment tools and scoring systems.</p> <p>My major concern is that the manuscript is difficult to read in its present form and should be shortened and more structured. The aim of the study should be pointed out more clearly and the influence of changes in visual appearance in patients with SSc more clearly discussed.</p> <ol style="list-style-type: none"> 1. Introduction: Should be majorly shortened and subheadings should be removed. I am missing the link between the assessment tools and SSc as the chosen disease. 2. M&M should also be shortened. Include all SSc definitions under the SSc subheading. Describe clearly the aim of the study. Are you aiming to explore only the DAS 24 factor structure using two scoring methods or do you aim to assess the magnitude of impaired visual appearance in SSc patients? Or both? 3. Results: should be structured in the same way as M&M. Where are results from the other than DAS 24 investigations shown (SIAS-6, PHQ-8, BFNE-II, SWAP)? If you also aim to address the magnitude of impaired visual appearance in SSc, please include a subheading with the results for both analyzing methods. 4. Discussion: the link between the differ scoring methods and their effect on SSc is missing. What consequences do the findings have for SSc patients? What would be suggested to do in clinics? What is the impact of changes in visual appearance in SSc?
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1 COMMENTS

COMMENT 1: While the question if 'NA' should be treated as missing seems to be an important point, I doubt if this specific population and this specific scale are suitable to arrive at conclusions which can be generalized to the general population and other scales as well. If not, I wonder if this investigation is interesting enough for a broader public than those who use this scale and population. It suggests that the study is more data-driven than designed for the research question concerned. Anyway, the authors do not discuss this point.

RESPONSE: We appreciate the reviewer's concerns regarding the generalizability of the current findings. The current study utilizes two large samples: the first sample is comprised of patients with systemic sclerosis, a disease characterized by significant physical disfigurement; the second sample is comprised of a combined community/clinical sample of patients who self-identified as having visible differences. That is, both samples are characterized by physical appearance concerns and thus are typical populations with whom the DAS-24 is used. Given the diversity of participants, and the similar results derived from each separate sample, we believe that the current results are likely to be generalizable to other populations with physical appearance concerns. We have added language to our sample descriptions within the results section (pp. 16-17) to underscore the physical appearance concerns in both samples. We have also added language to the discussion section supporting the generalizability of the current findings to other visible difference samples (p. 20).

Unfortunately there was not an opportunity to conduct convergent validity analyses with the more representative visible differences sample, which we have emphasized further in the discussion (p. 21).

The study was most definitely not “data-driven.” Rather, based on review of published evidence on the DAS-24, we formulated a hypothesis and used a confirmatory approach to test the hypothesis. The idea of the study was generated because the DAS-24 is used as a key measure of the SPIN cohort, which collects data from over a thousand SSc patients with physical appearance concerns. Concerns were raised by team members when it was discovered that there were high rates of NA responses with possible implications on scoring. Thus we sought to test this scoring feature and determine what would happen if we accounted for it. It may be that the reviewer is suggesting that the study was “data driven” because funding was not separately obtained and data not separately collected to answer this question in a specifically designed sample. We emphasize that testing our hypothesis separately in two large existing cohorts was an efficient approach to answering an important research question.

COMMENT 2: The introduction part is focused on the DAS-24, its scoring and its factor structure. A more broad elaboration of the theoretical notions of ‘NA’ versus ‘missing’ is lacking. F.i: if a subject is stating that a problem area is not applicable: does not that imply that for him/her this problem does not exist; in other words that the stress/dysfunction concerned is zero (like the standard way of scoring implies too)? Treating it as ‘missing’, ignores this outcome but indicates that there seems to be a problem which the subject forgot/does not want to reveal, etc. Previous discussions in the literature on the NA phenomenon (f.i. Holman et al, Health Qual Life Outcomes, 2004;2;29) are missing too.

RESPONSE: We thank the reviewer for this suggestion. We have added relevant language to the introduction (p. 10) and discussion sections (p. 19) with reference to previous literature on “not applicable” responses. In keeping with the request from Reviewer 3 to shorten the manuscript, and specifically the introduction (Comments 1 and 2) we have kept the discussion brief.

COMMENT 3: In the Methods parts, the ‘missing method’ way of scoring is not clear. Which type of missing value imputation is used: complete case, LOCF, regression analysis? One might come to different conclusion, depending on the way the missings are handled. Hypotheses about associations of the DAS-24 are stated without explicit reasons. The sample size appears to be sufficient for the statistical analysis.

RESPONSE: We have added clarifying language within the Scoring section (p. 15) pointing the reader to the methods for handling missingness within the Analytic Plan section (p. 15). We note that the WLSMV procedure for ordinal data in Mplus was used and, because there were not covariates, this is analogous to pairwise present (Muthén & Muthén, 2015). This approach has been used to handle missing data when WLSMV estimation is used in numerous other recent articles (e.g., Mejia, Filus, Calam, Morawska, & Sanders, 2016; Wang, Chassin, Eisenberg, & Spinrad, 2015).

COMMENT 4: The Results section again shows how specific this (mainly female and advanced age) sample is. The many dermatological sub categories are probably mainly interesting for experts in that field.

RESPONSE: We have removed language in the sample description paragraph within the results section regarding the disease subtypes (p. 16); we agree that this may not be of significant interest to researchers who are interested in physical appearance more broadly. However, we have left information regarding disease characteristics related to physical appearance in Table 1. This information affirms the significant physical appearance concerns in the SSc sample.

We have also added more information describing the Visible Difference sample to support the generalizability of this sample (p. 12). Language has also been added to the discussion section to support the generalizability of the findings to other populations with visible differences (pp. 20-21)

COMMENT 5: It is concluded that the factor loadings under the Missing Method scoring are larger than under the NA scoring. However, this is understandable because in the authors only use the 1st factor for determining the loadings, while in the NA scoring, two factors are found. So, it is an 'unfair' comparison. I expect that in the NA scoring the variance explained by the items for the two factors together will be larger than for one factor only; and perhaps larger than the one factor loading in the Missing Method scoring.

RESPONSE: The reviewer is correct that in the current study the factor loadings are stronger when the "not applicable" responses are treated as missing, rather than scored as 0 (a very low score). The DAS-24 is hypothesized to have a single underlying construct of distress and dysfunction related to physical appearance concerns, but only one study (Moss et al., 2015) used exploratory (i.e., data driven) methods and found the second factor. Our concern is that this finding was not theoretically-driven, which is underscored by the fact that the second factor was not hypothesized a priori, is not supported by the long-form version of the measure (DAS-59; Carr, Harris, & James, 2000), and has only been found in a single sample. Moreover, studies using the DAS (both long and short forms) report very high Cronbach's alphas, providing further support for unidimensionality.

Our primary concern regarding Moss et al.'s (2015) findings of a second factor was that this factor was comprised of 6 items with a high amount of "not applicable" responses (on average, 25.97% of responses for each item were marked "not applicable"), whereas the first factor was comprised of 10 items that did not have the "not applicable" response option, and of the items that did have this response option, only 13.11% were marked as "not applicable". To corroborate our interpretation, we tested the two-factor CFA model using the SSc data (scoring data according to the Standard Method, with "not applicable" responses scored as 0) to confirm this and found that the two-factor model did not fit well according to RMSEA: $\chi^2(251) = 2067.7$, $p < .001$; CFI = .91, RMSEA = .09 and clearly did not improve upon the one-factor model. Moreover the interfactor correlation was very high (.77, $p < .001$), suggesting that the two factors are so highly associated they are practically indistinguishable. Taken together we do not see utility in pursuing the two-factor model further, given the lack of theoretical support and lack of fit in confirmatory (i.e., theory driven) models. Thus, we have chosen to focus on testing the single factor, theoretically relevant model, in the current study.

COMMENT 6: In the Discussion, the authors appear to consider the one factor solution of the DAS-24 better ('logically more consistent') than the two factor solution because of the original conceptualization of the DAS-24 as representing a single latent construct. One may alternatively consider the two factor outcome from the NA scoring as a falsification of this conceptualization.

RESPONSE: We undertook the current study due to our concerns that the DAS-24 scoring method is logically untenable—specifically, counting items that are not relevant to a respondent as a "0" score (indicating no appearance-related distress and dysfunction) could distort the scoring and interpretation of the measure, when the primary response scale ranges from 1-4. That is, we were concerned that the scoring of the DAS-24 may be influencing scores so substantially that it could drive a second factor. Indeed, in the article by Moss et al. (2015), all 6 items on the second factor had large amounts of data with the "not applicable" option chosen for both in the SPIN (SSc) and Visible Difference samples (19.1-48.6%). Given this concern, we re-analyzed the data from the Moss study (Visible Differences sample) and also a new data set (SPIN cohort of SSc patients with physical appearance concerns).

Because the one-factor structure fit well in two independent samples here, we believe there is not a reason to introduce more complexity particularly given that the theoretical construct underlying the DAS (both the long [59-item; Carr, Harris, & James, 2000] and short [24-item; Carr, Moss, & Harris, 2005] forms) is hypothesized to be a single factor representing appearance-related distress and dysfunction. We do not believe that a model based on a scoring method with significant concerns (i.e., undervalued total score of a person for whom multiple items are not applicable) invalidates the assumption that a single latent factor underlies the DAS items.

COMMENT 7: A final word about the 5 pages with authors and affiliations. I was a bit surprised to see that so many authors (approaching 100) were involved in this article of limited scope and cannot imagine that each of them delivered important input for the article, apart from supplying a number of patients. I wonder if their contribution is sufficient for an authorship, or will be more qualified for a mentioning in the acknowledgment sections. But it is up to the editors' policy to decide upon this point.

RESPONSE: The Scleroderma Patient-centered Intervention Network (SPIN) Cohort is a large, multi-center study that depends on substantive contributions from a large number of contributors. All meet ICMJE criteria for authorship. This is denoted under author contributions (p. 22). See also, Editor Comment 1.

REVIEWER 2 COMMENTS

COMMENT 1: This study was carried out to investigate the possible influence of "not applicable" responses in The Derriford Appearance Scale-24 (DAS-24) in patients with SSc and volunteers with general visible differences using two methods; the standard method and the missing method. The study is very impressive and very well written, however there are few comments. Page 9 line 4: the part of "STRENGTHS AND LIMITATIONS" is better moved to the last section of the discussion

RESPONSE: We thank the referee for the positive comments about our study. We apologize for any confusion regarding the initial "Strengths and Limitations" section following the abstract. We included this section following the guidelines for BMJ OPEN (<http://bmjopen.bmj.com/pages/authors/>) which state "A section, placed after the abstract, consisting of the heading 'Strengths and limitations of this study', and containing up to five short bullet points, no longer than one sentence each, that relate specifically to the methods. They should not include the results of the study." These are summarized in the discussion as well (pp. 20-21).

COMMENT 2: Page 13 line 10: Why was the submission online and repeated every 3 months, this needs to be more clarified, wouldn't it be better if the form was submitted in the center with the investigator?

RESPONSE: SPIN participants complete online surveys every 3 months as part of participation in the overall study. However, only baseline data was used for the current study (p. 12). We agree that in many cases it is ideal to conduct in-person assessments. However, for this large cohort study comprised of respondents with a rare disease, the online platform is preferred in order to collect data from large amounts of patients in a relatively cost-effective and timely manner, not requiring frequent in-person trips to study centers.

REVIEWER 3 COMMENTS

COMMENT 1: My major concern is that the manuscript is difficult to read in its present form and should be shortened and more structured. The aim of the study should be pointed out more clearly and the influence of changes in visual appearance in patients with SSc more clearly discussed.

RESPONSE: We have made considerable edits to the manuscript, removing superfluous words to shorten it, and clarifying our language throughout, reducing the manuscript by 2 pages. We have more clearly pointed out the aim of the study (p. 11). We have included a discussion of physical appearance change concerns to the introduction (p. 9), specifically noting that appearance changes are common and have been linked with numerous quality of life outcomes including body image dissatisfaction, psychosocial functioning, anxiety, and depression.

COMMENT 2: Introduction: Should be majorly shortened and subheadings should be removed. I am missing the link between the assessment tools and SSc as the chosen disease.

RESPONSE: We have shortened and made significant revisions to the introduction (pp. 9-11). This includes removing language, and modifying language to more concisely narrow the existing literature to the point of the current study. As requested, we have removed all subheadings from the introduction. We have shortened language in the first paragraph in particular to more explicate that SSc is a disease with considerable physical appearance concerns. We have also bolstered our language linking the DAS-24 with SSc (p. 9).

COMMENT 3: M&M should also be shortened. Include all SSc definitions under the SSc subheading. Describe clearly the aim of the study. Are you aiming to explore only the DAS 24 factor structure using two scoring methods or do you aim to assess the magnitude of impaired visual appearance in SSc patients? Or both?

RESPONSE: We have shortened the Methods section by removing text and clarifying our wording throughout (pp. 12-16). We considered moving SSc-related measures under the SSc Sample heading as requested but ultimately decided against this as these definitions are relevant to all SSc patients, not just the current SSc sample. However, we have added parentheses next to each set of variables (Main Study Measure, Convergent Validity Measures, SSc Measures) indicating which sample provided which measures in order to help clarify (pp. 12-14). We have clarified that the aim of the study is to determine whether the traditional scoring method, that is, scoring "not applicable" as a 0 affects the factor structure and convergent validity of DAS-24 scores (p. 11).

COMMENT 4: Results: should be structured in the same way as M&M. Where are results from the other than DAS 24 investigations shown (SIAS-6, PHQ-8, BFNE-II, SWAP)? If you also aim to address the magnitude of impaired visual appearance in SSc, please include a subheading with the results for both analyzing methods.

RESPONSE: We have made edits to our results and methods sections so they are uniformly presented according to the "Data Analysis" section header (pp. 12-18). We have retained the sample descriptions and material descriptions earlier in the methods section which is in alignment with the journal's standards and the structure of recently published articles at BMJ Open. We presented the bivariate correlations with the SIAS-6, PHQ-8, BFNE-II, and SWAP subscales to evaluate the convergent validity of the DAS-24 scores across the two scoring methods to determine whether scoring method influenced the coefficients, and, whether there are clinical implications for the traditional method of DAS-24 scoring. This is reflected in the subheading "Convergent Validity: SSc Sample" (p. 18). However, one limitation, which we note in our manuscript (p. 21), is that the convergent validity measures were only available in the SSc sample, and thus could not be tested in the visible difference sample. The purpose of the study is not to address physical appearance concerns in SSc, but to evaluate the DAS-24 in this sample and another sample with physical appearance concerns, given that the DAS-24 is a measure that is of interest to populations with physical appearance concerns (like SSc).

COMMENT 5: Discussion: the link between the differ scoring methods and their effect on SSc is missing. What consequences do the findings have for SSc patients? What would be suggested to do in clinics? What is the impact of changes in visual appearance in SSc?

RESPONSE: As described above we have made significant changes throughout the manuscript to support the argument that the current findings are relevant to SSc, and also to any population with visible differences (See Reviewer 1, Comments 1 and 4). The use of the SSc sample (in addition to the Visible Difference sample) was to include participants who identified as having physical appearance concerns, not to identify specific consequences for SSc patients. We have added language to our discussion section noting that the primary differences between scoring methods are with statistical precision, but that clinical inferences are likely to be so similar across methods that there is no reason to think that the DAS-24 cannot be used in clinical or research contexts using the current scoring protocol (p. 20). We have also added language regarding measurement precision and participant burden noting that future research should work towards shortening the measure to reduce patient burden (p. 20).

Again, we thank the reviewers and the editor for their time and thoughtful comments. Please let us know if there are additional comments that need to be addressed. We look forward to receiving your feedback on our revised manuscript.

VERSION 2 – REVIEW

REVIEWER	Anna Hoffmann-Vold Oslo University Hospital, Norway
REVIEW RETURNED	04-Jan-2018
GENERAL COMMENTS	The authors have adequately revised the manuscript that is significantly enhanced. I have no further comments.

Correction: 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

Merz EL, Kwakkenbos L, Carrier M on behalf of the SPIN Investigators, *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

In the Collaborator list, the name 'Ilham Benzida' should have been spelt 'Ilham Benzidia'.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

BMJ Open 2018;8:e018641corr1. doi:10.1136/bmjopen-2017-018641corr1

