Quality of Life Research (2006) 15: 1383–1394
DOI 10.1007/s11136-006-0018-8

Single-factor scoring validation for the Health Assessment Questionnaire-Disability Index (HAQ-DI) in patients with systemic sclerosis and comparison with early rheumatoid arthritis patients

Jason C. Cole[1,2], Dinesh Khanna[3,4,5], Philip J. Clements[6], James R. Seibold[7], Donald P. Tashkin[8], Harold E. Paulus[6], Michael R. Irwin[9], Sarosh J. Motivala[9], Daniel E. Furst[6] & and on behalf of the Scleroderma Lung Study (SLS) and Relaxin Study

[1]*Consulting Measurement Group, Huntington Beach, CA, USA (E-mail: jcole@webcmg.com);* [2]*Quality-Metric, Lincoln, RI, USA;* [3]*Division of Immunology, Department of MedicineUniversity of Cincinnati, Cincinnati, OH, USA;* [4]*Institute for the Study of HealthUniversity of Cincinnati, Cincinnati, OH, USA;* [5]*Veterans Affairs Medical Center, Cincinnati, OH, USA;* [6]*Division of Rheumatology, Department of MedicineUniversity of California, Los Angeles – David Geffen School of Medicine at UCLA, Los Angeles, CA, USA;* [7]*Division of RheumatologyUniversity of Michigan Scleroderma Program, Ann Arbor, MI, USA;* [8]*Division of Pulmonary and Critical Care Medicine, Department of MedicineUniversity of California, Los Angeles – David Geffen School of Medicine at UCLA, Los Angeles, CA, USA;* [9]*University of California, Los Angeles – Cousins Center for Psychoneuroimmunology, Los Angeles, CA, USA*

Accepted in revised form 4 May 2006

Abstract

Objective Structural validity for the Health Assessment Questionnaire-Disability Index (HAQ-DI) has recently been provided for patients with rheumatoid arthritis (RA). The goal of the current study was to examine the structural validity of the HAQ-DI in patients with systemic sclerosis (SSc, scleroderma) and to compare its performance with that in patients with RA. *Methods* The HAQ-DI structural validity was first assessed in a sample of 100 scleroderma patients using confirmatory factor analysis. Second, the similarity of factor structures between SSc patients (n = 291) and RA patients (n = 278) was tested using a multigroup structural validity model to assure that comparison of scores between these two diagnostic groups is appropriate. *Results* Results yielded a single-factor HAQ-DI score which favored the current scoring system of the HAQ-DI (model fit was CFI = 0.99 and RMSEA = 0.04). Moreover, even the most stringent model of multigroup structural validity affirmed the similarity between SSc and RA patients on the HAQ-DI (model fit was CFI = 0.99 and RMSEA = 0.04) nor was it different from a model without any demands on group similarity: CFI difference = 0.007; χ^2 = 4.29, df = 26, $p = 0.99$. *Conclusion* The current results indicate that a single-factor HAQ-DI is appropriate for future clinical trials in scleroderma and, in addition, HAQ-DI scores among patients with SSc and early RA can be compared legitimately with one another.

Key words: Confirmatory factor analysis, HAQ-DI, Latent analysis, Rheumatoid arthritis, Systemic sclerosis

Systemic sclerosis (SSc, scleroderma) is a connective tissue disease of unknown etiology characterized by microvascular injury, variable fibrosis of the skin, and distinctive visceral involvement including the heart, lungs, kidneys, and gastrointestinal tract [1]. SSc has little effective treatment and no cure, and patients must cope with pain, disfigurement, disability, and feelings of helplessness.

Given the impact of SSc on activities of daily living, measures of this impact have been developed. The Health Assessment Questionnaire-Disability Index (HAQ-DI; published in 1980 by [2]) is the most utilized of musculoskeletal-targeted measures. It has been used extensively in rheumatoid arthritis (RA; [3]) and in SSc [4, 5]. Scores from the HAQ-DI have been shown to be reliable and responsive to change in a SSc clinical trial [5], and to predict morbidity and mortality in patients with diffuse SSc [6].

Previous researchers have shown the HAQ-DI to be reliable and convergently valid, yet empirical confirmation for the factor structure and scoring system of the HAQ-DI is limited. According to Messick in his seminal paper on validity [7], the manner in which a test is scored must match the underlying latent structure of the test in order to have structural fidelity (i.e., a valid fit between the latent structure and scoring system). In 2005, Cole et al. [3] provided the first evidence for structural fidelity of the HAQ-DI, examining its performance as a single total score in RA patients (i.e., a single-factor model was confirmed).

However, findings in RA patients regarding the scoring validity of the HAQ-DI do not necessarily generalize to other diagnostic groups. Indeed, Haynes et al. [8] noted that the validity of a test is necessarily specific to each subgroup of users: inferences about the structural validity of the HAQ-DI in diagnostic groups other than RA are not appropriate without empirical evidence. Moreover, to understand the similarity of structural validity of the HAQ-DI for more than one diagnostic group, the similarity of the latent structures must be examined between the two diagnostic groups [9]. When the performance of the HAQ-DI is found to be structurally invariant between two diagnostic groups, comparisons on the same scoring system are psychometrically justified [10].

Given the increased utilization of the HAQ-DI in SSc clinical trials, the goal of the current study was to examine the structural validity of the HAQ-DI in patients with SSc. Confirmatory factor analysis (CFA) was used to examine the fit of the current scoring system of the HAQ-DI, already validated for patients with RA patients, with the responses from SSc patients [4, 11]. The similarity of the structural validities between SSc and RA was then tested to determine structural invariance between the two groups. Results of this analysis should be useful in guiding and clarifying interpretation procedures for HAQ-DI domains from SSc patients. Results should also be informative as to the appropriateness of comparing HAQ-DI scores between RA patients and SSc patients.

Methods

Participants

Patients were those enrolled in one of three studies, as described next. The organization of the participants within each of the studies (including subgroups of diffuse and limited SSc) is shown in Figure 1.

Relaxin study

Participants from the relaxin study were in a double blind, placebo-controlled, multicenter, randomized clinical trial evaluating the safety and efficacy of continuous subcutaneously infused recombinant human relaxin in diffuse SSc over a period of 24 weeks. The details of the study have been published elsewhere [11]. Briefly, participants were randomized to either relaxin 25 μ/kg/day, relaxin 10 μ/kg/day, or placebo in a 2:1:2 ratio. All participants had SSc as defined by the American College of Rheumatology criteria [12] with diffuse disease defined as the presence of thickening proximal as well as distal to the elbows and knees inclusive of the trunk and face [13]. All patients had a disease duration of no more than 5 years, with an average duration of 2.20 years. Two hundred and thirty-nine (239) patients with diffuse SSc were enrolled, with 136 patients receiving relaxin and 103 patients receiving the placebo. Baseline HAQ-DI domains were available for 231 patients of these patients, all of whom were used for the current analyses.

Scleroderma lung study

Participants were from the Scleroderma Lung Study, a double blind, placebo-controlled, multicenter, randomized clinical trial evaluating the safety and efficacy of 1 year of treatment with oral

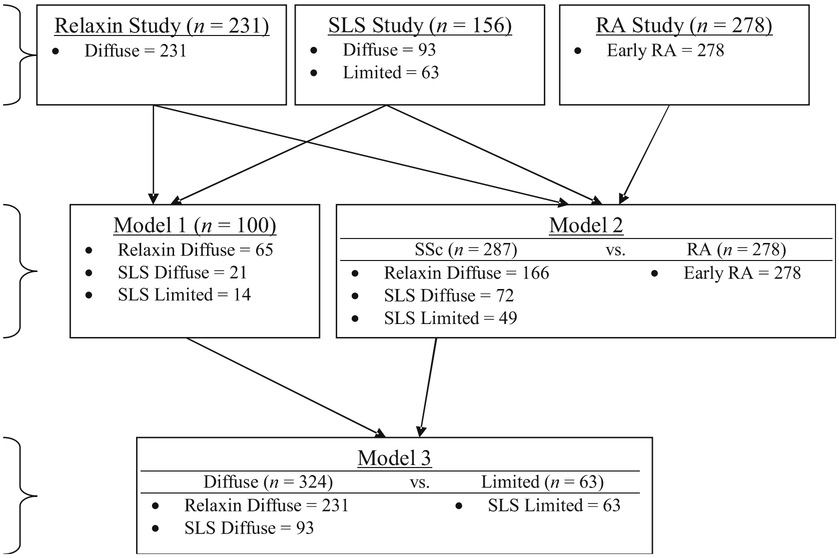

Figure 1. Sample flow throughout analyses.

cyclophosphamide vs. placebo for rapidly progressive active pulmonary alveolitis associated with SSc; it had an additional year of follow up without any treatment [14]. Patients with SSc as defined by the American College of Rheumatology classification criteria [12] with ≤ 7 years duration (onset defined as the date of the first typical non-Raynaud's manifestation) were included in the trial if they also had evidence of SSc-related interstitial lung disease and active alveolitis defined by thoracic high-resolution computed tomography (any ground glass opacification) and bronchoalveolar lavage (≥2% neutrophils and/or 3% eosinophils); average disease duration was 3.10 years. SSc patients were further divided into those with limited and diffuse cutaneous SSc based on the distribution of the skin thickness. Limited disease was characterized by the skin thickness distal to the elbows and knees and proximal to the clavicles (with or without facial involvement). Diffuse disease was characterized by the same criteria as the relaxin study. One-hundred and fifty-eight patients were part of the clinical study. Baseline HAQ-DI domains were available for all but two of these patients (both of whom had limited SSc). Thus, a total of 156 SSc patients were used from the SLS study database, 40.4% of whom had limited SSc. Figure 1 provides clarification for the division of diffuse and limited participants in the current study.

Rheumatoid arthritis sample

RA patients included in the current study are a subset of a group of early RA patients participating in a long-term observational study by the Western Consortium of Practicing Rheumatologists, which is a regional consortium of rheumatology practices in the western United States and Mexico. The consortium has been described in detail in previous publications [15, 16]. Briefly, patients in this subset had a diagnosis of early RA (median duration was 5.2 months since symptom onset, disease duration was less than 15 months), had no previous disease modifying antirheumatic drug treatment, were rheumatoid factor seropositive (RF median 214 IU/ml), and had ≥6 swollen joints and ≥9 tender joints. Baseline HAQ-DI domains were available for 278 patients, all of whom were used for the current analyses. These data were used for a previously published analysis by Cole et al. [3].

Measures

Health Assessment Questionnaire-Disability Index (HAQ-DI)

The HAQ-DI is a musculoskeletal-targeted measure of functional status with demonstrated utility for patients with SSc [17, 18]. The original HAQ-DI was designed as a 20-item self-administered questionnaire that examines difficulties with the performance of activities of daily living on a 0–3 scale (*no disability* to *severe disability*) in eight domains (dressing and grooming, arising, eating, walking, hygiene, reach, grip, and other activities). In the original HAQ-DI, an additional grade of difficulty was added in patients using assistive/adaptive devices (such as canes, walker) as in more recent studies [6, 17], though, we did not modify patients' responses for use of assistive/adaptive devices.

The HAQ-DI is calculated by summing the highest score in each of the eight domains and dividing the sum by 8, giving a score between 0 and 3 on a continuous scale. Both observational studies [19–21] and clinical trials [22–24] have used the HAQ-DI and found its scores to be a reliable and valid predictor of work disability [25], morbidity [25, 26], and mortality [27].

Data analysis

Data were entered and cross-checked by research assistants with ample data entry experience. HAQ-DI scores were obtained per the instructions in Bruce and Fries [28]. No missing data were present for the HAQ-DI domain scores. HAQ-DI domain descriptive statistics are presented in Table 1.

Confirmatory factor analysis

Prior to examination of the structural validity with CFA,[1] the total database was randomly split into two sections. One split contained 100 participants and was reserved for use with the initial CFA (hereinafter, CFA sample), whereas the other split contained the remainder of the participants (invariance sample; n = 291). For the CFA sample, we determined that 100 participants should be sufficient given (a) the stability of the previous HAQ-DI CFA [3, 29] and (b) the size of the path coefficients (i.e., the standardized regression coefficients between the latent factor and the HAQ domains) found in the HAQ-DI for RA patients (0.62–0.81), both of which have a marked impact on stability for CFA [30].[2] The CFA model examined (referred to as Model 1) was based on the current HAQ-DI scoring system wherein a unidimensional latent total score impacts the way each patient scores on the HAQ-DI domains, as shown in Figure 2. Within this model, the rectangular blocks represent HAQ-DI domains with ovals to their left which represent each domain's residual (i.e., any variance of each HAQ-DI domain not measured by its relationship to the latent variable). The circle to right of the domains represents the overall HAQ-DI latent variable of disease impact.

CFA was conducted with AMOS structural equation modeling software [31]. Maximum likelihood (ML) was used to estimate the CFA model. The purpose of the CFA was to measure the extent to which the current scoring system explains the way in which patients respond to the HAQ-DI questions, thus providing evidence for the structural fidelity of the scoring system with the HAQ-DI's latent construct (i.e., does the scoring system of the HAQ-DI fit with the latent constructs underlying the HAQ-DI?). Because the

Table 1. HAQ domain pearson correlations and descriptive statistics for all SSc participants

	Dressing and grooming	Arising	Eating	Walking	Hygiene	Reaching	Grip	Activity
Dressing and grooming	–	0.58	0.63	0.40	0.69	0.66	0.53	0.57
Arising		–	0.51	0.50	0.61	0.59	0.51	0.58
Eating			–	0.42	0.68	0.65	0.62	0.56
Walking				–	0.44	0.49	0.46	0.54
Hygiene					–	0.73	0.54	0.64
Reaching						–	0.57	0.70
Grip							–	0.54
Activity								–
Mean	1.08	0.83	1.12	0.76	1.24	1.30	0.94	1.30
SD	0.91	0.83	0.96	0.84	1.13	1.02	0.82	0.95
Range	0–3	0–3	0–3	0–3	0–3	0–3	0–3	0–3

Note: Correlations are provided for descriptive purposes and were, therefore, not analyzed for significance. N = 391, HAQ-DI mean = 1.07 and *SD* = 0.74.

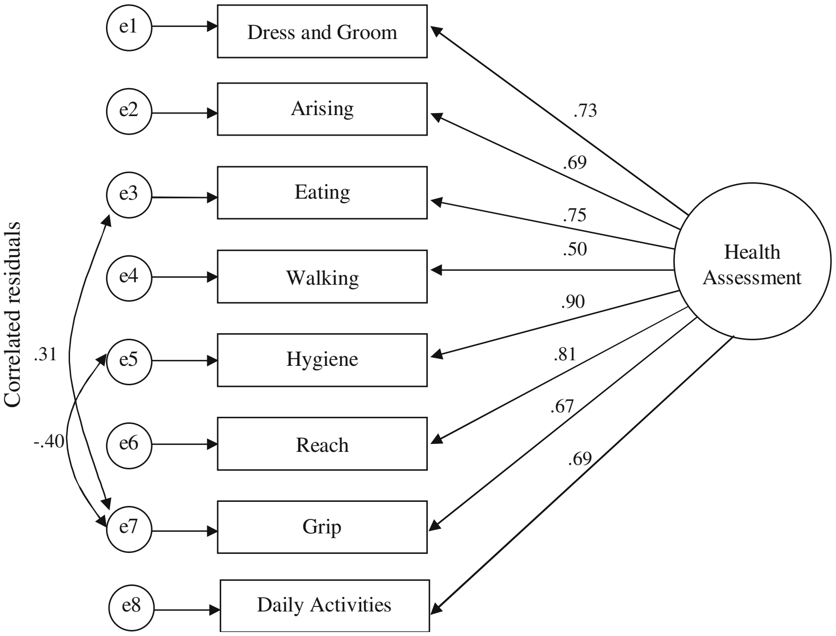

Figure 2. HAQ single-factor model from the confirmatory factor analysis (Model 1) with standardized path coefficients.

relationships among health-outcomes variables from many fields are typically distributed nonnormally [32, 33], and because latent analyses using ML assume normal distributions [34], adjustments need to be made to the model calculations to control for nonnormality. Two thousand Bollen–Stine [35] bootstraps were used during model estimation to control for multivariate nonnormality, per the recommendations of Nevitt and Hancock [34]. The process of bootstrapping takes multiple random subsamples from the current sample in order to smooth over any inaccuracies in the model's standard errors due to nonnormality. Although nonparametric extraction is also appropriate for the HAQ scales, these techniques require far larger samples [36]. Moreover, the Bollen–Stine bootstrap with ML extraction has been shown to perform very well, controlling estimation problems [37] with at least four-ordered categories [36].

Model fit statistics in CFA (and generally for structural equation modeling – SEM) provide measures of the strength of relationship between the theoretical model and the data. Schumacker and Lomax [38] suggested that it is best to review multiple model fit statistics in order to examine the model from various perspectives [38]. Therefore, in the current study, four fit indexes were used: goodness of fit (GFI; ranges from 0 to 1 with larger values indicating better fit), comparative fit index (CFI; ranges from 0 to 1 with larger values indicating better fit), nonnormed fit index (NNFI; ranges from 0 to 1 with larger values indicating better fit, although values can at times exceed 1.0 and then suggest overfitting), and root mean squared error of approximation (RMSEA; ranges from 0 to 1 with smaller values indicating better fit). GFI measures the amount of variance and covariance in the data that is reproduced by the tested model. CFI, NNFI, and RMSEA provide estimates of Type I (CFI and NNFI) and Type II error (RMSEA). CFI is a measure of Type I error in that it specifies the amount of difference between the examined model and the independence model (i.e., a standard comparison model that asserts none of the components in the model are related), with higher scores indicating larger differences. NNFI conducts the same task as CFI, but takes into consideration the number of parameters in a model, an aspect that can inflate CFI [39]. RMSEA is complementary to CFI and NNFI as it is a measure of Type II error, determining how well the examined model reproduces the saturated model (i.e., another standard model that asserts that all of the variance and covariance of the dataset is explained), with lower scores indicating greater similarity. If each of these four fit indices meets or surpasses these thresholds, then the model can be considered satisfactory. GFI was evaluated with a minimum criterion of 0.90 [40], and CFI and NNFI were to be no less than 0.95 [41]. RMSEA yields both a score and a 90% confidence interval; good fit was indicated when the score was 0.06 or lower [41].

Structural invariance testing

Upon the successful completion of the CFA and acceptance of a well-fit model, one has provided evidence of structural validity (the validity that the theoretical latent structure underlying a test is supported by the data). Next, comparison of the HAQ-DI structural validities was conducted for SSc patients and RA patients. The group of SSc patients was comprised of the 291 participants from the random split of the SSc dataset, and the group of RA patients was comprised of the 278 patients previously analyzed on the HAQ-DI by Cole et al. [3]. Therefore, multigroup SEM was employed to examine the structural invariance of the HAQ-DI for the SSc patients and RA patients (referred to as Mode 2). Multigroup SEM is tested through a series of models, to determine where difference may be present if the fully equated model is not invariant (i.e., if the models are not fully equal on all SEM parameters between the two groups). We examined the fully equated model, and planned to examine lesser-constrained models only if the fully equated model did not meet goodness of fit criteria.[3] As before, the multigroup SEM model was evaluated using the goodness of fit criteria CFI, NNFI, and RMSEA (using the aforementioned criterion levels). As it is common to assess multigroup SEM models with the inclusion of latent means [9], GFI was not examined because SEM models with mean structure in AMOS are not calculated because of the theoretical problems associated with integrated implied means and intercepts into the formula (personal communication, James Arbuckle – AMOS author, April 13th, 2006).

Comparison of the fully constrained invariance model to the multigroup model without constraints that force the groups to be equal was conducted in order to determine if the full invariance model placed unrealistic restrictions of the model. This comparison was completed through the use of two statistics: the change in CFI (ΔCFI) and a χ^2 difference test. Cheung and Rensvold [42] recommended a criterion of no more than 0.02 CFI difference for ΔCFI. Moreover, the χ^2 difference test should be nonsignificant between the models.

Model refinement

Often a model's fit indices may come close to reaching these thresholds, but not close enough to be considered satisfactory. In such a case, minor adjustments to the relationships in the model can be made and the model can then be retested. The determination of which adjustments to make can be guided by using modification indices, which provide an estimate of the improvement in model fit that will occur by adding a given relationship (e.g., a correlation between the residuals of domains of eating and grip), including direct paths and correlations [38]. A standard approach of using a modification index of approximately 10 was used (relating to a reduction in chi-square by 10, which indicates better model fit); paths with a modification index much lower than 10 may to be too weak to provide substantive benefit. Modification of the model after an initial analysis would only be conducted if the modification met statistical criteria *and* fits with the theoretical understanding of the HAQ-DI [38]. When modifications are added to a model, the model will be rerun and interpreted with the new fit indices [43]. For more applied information on this process, see Cole et al. [3].

Results

Confirmatory factor analysis of the HAQ-DI in SSc patients

A total of 387 SSc patients had domain-level information on the HAQ-DI, for which no missing data were present. The final sample of SSc patients had a mean age was 47.75 years (*SD* = 11.28 years) and a mean HAQ-DI score of 1.07 (*SD* = 0.74). The descriptive statistics for the HAQ-DI domain scores in the combined SSc samples are provided in Table 1, including correlations among the HAQ-DI domains (in the upper matrix of Table 1) as well as the mean, *SD*, and range of scores for each domain. Correlations among all of the domains were mostly large (according to criteria from [44]), ranging from the mid 0.40s to the low 0.70s. Each of the eight HAQ-DI domain scores range from 0 to 3, with the means and *SD*s near 1.0 for most scales. Additionally, Figure 1 shows the random assignment of participants to the analyses, including diffuse and limited SSc patients for each analysis. The Model 1 and Model 2 sample splits resulted in groups with no statistical differences on age, gender, HAQ total score, or race.

A CFA was run on the single-factor model using CFA sample as had been previously found for the HAQ [3]. In this analysis, the single-factor model was close but did not meet adequate fit criteria (GFI = 0.92, CFI = 0.95, NNFI = 0.93, and RMSEA = 0.10). Whereas GFI and CFI were acceptable, NNFI and RMSEA were not. These results indicate that the model was missing some significant relationships and that minor adjustments in the model were needed. Thus, to find unmodeled relationships that have both statistical and theoretical importance to the HAQ-DI model [38], modification indices were inspected. In most models, variable relationship can be added as unidirectional (i.e., regression paths) or bidirectional (i.e., correlational). Because the current model contained only a single factor, additional paths could only be added as correlations, specifically as correlated residuals [45].

Residuals refer to the variance that is not accounted for by the relationship of a particular domain to its latent variable. For example, when examining Grip and Health Assessment, the residual of Grip is all of the variance not otherwise accounted for by the path coefficients (i.e., the current modeled relationship) from Health Assessment to Grip, or 1 – the square of the standardized coefficient (i.e., $0.67^2 = 0.449$) for standardized values (Figure 2). This residual value is influenced by many sources of variance, such as method variance, shared content beyond the primary factor, and measurement error [45]. Therefore, a correlation between two residuals occurs when aspects of these residual terms are strongly related (although correlations between residuals are not generally assumed to arise from correlated measurement error as this should be random [46]). The examination of fit indices revealed relatively high scores between the residuals for Grip and Eating (modification index = 8.91) as well as between Grip and Hygiene (modification index = 7.67), resulting in correlations of $r = 0.31$ and −0.40, respectively. Both of these correlated residuals appear to have a content relationship in that each focuses on hand-based motor skills, an important limitation for SSc patients. For example, if one has fine-motor skills problems then these may exacerbate scores on grip and activities dependent on grip, such as eating and hygiene, more than the scores on other domains. This differential between scores on related domains compared to all of the HAQ domains results in a secondary relationship which is reflected in the correlated residuals. Although these are occasionally spurious [47], substantiation of correlated residuals across multiple samples can also enhance greater understanding of the construct and inspire new lines of research [3].

It should be noted that we used modification indices with smaller values than normally used. As the modification index is a measure of how much reduction in chi-square can be achieved with the addition of the path (and thus, improvement in data-model fit), the aforementioned modification indices were estimated to reduce chi-square by 22.4 and 19.3%, respectively. It is not altogether uncommon to adjust this criterion when chi-square is low, and thus small modification indices still account for a large percent of chi-square change [48, 49].

The modified CFA model generated satisfactory fit statistics for all model fit criteria with GFI, CFI, and NNFI each greater than 0.95 and RMSEA = 0.04 (see Table 2 results for Model 1). Figure 2 shows the final factor structure of the HAQ-DI, including the standardized path coefficients for the HAQ-DI latent variable on each of the HAQ-DI domains, as well as the level of correlation between the two domain residuals. If one squares the path coefficients, an estimate of the variance explained by the latent construct for each HAQ domain is provided. For example, squaring the coefficient between HAQ-DI and Hygiene takes the standardized coefficient of 0.90 to an explained variance of 81%.

Table 2. Fit statistics for all structural models

Model	χ^2	df	GFI	CFI	NNFI	RMSEA	RMSEA 90% CI
Model 1. CFA (modified)	20.94*	18	0.95	0.99	0.99	0.04	0.00–0.10
Model 2. Structural invariance SEM – scleroderma vs. RA (fully constrained)	112.27	60	–	0.98	0.99	0.04	0.03–0.05
Model 3. *Post hoc* structural invariance SEM – diffuse vs. limited (fully constrained)	85.84	36	0.93	0.95	0.95	0.06	0.05–0.07

*$p > 0.05$.

Note: GFI – goodness of fit; CFI – comparative fit index; NNFI – nonnormed fit index; RMSEA – root mean square error of approximation; CI – confidence interval.

Therefore, the latent construct explains 81% of the variance for Hygiene. Conversely, the lowest coefficient in the model, between HAQ-DI and Walking, explains 25% of the variance. Although this value is lower than all others, it is considered to be between "fair" and "good" according to criteria from Comrey and Lee [50] and it is not surprising to have a wide range of strong loadings given the diversity of activities covered by the HAQ.

Overall, model fit provided substantial evidence for the use of a single total score on the HAQ-DI. Additionally, bootstrapped confidence intervals showed a 1–2% change from the unadjusted estimates, resulting in moderately better (and sufficient) multivariate normality.

Structural invariance

The fully constrained structural invariance model (i.e., the model forcing all relationships to be equal between SSc and RA patients) fit the data well, based on all of the model-fit criteria (Table 2 displays the fit statistics for Model 2). Thus, when forcing all parameters of the model to be equal between SSc patients and RA patients, the model still showed strong relationship to the data. Bootstrapped confidence intervals again showed a 1–2% change from the unadjusted estimates, resulting in moderately better (and sufficient) multivariate normality.

Additionally, comparison between fully constrained invariance model and the nested unconstrained model showed no substantial differences between the models. CFI for the unconstrained model was 0.977, and thus ΔCFI = 0.007. Moreover, the χ^2 difference test was also nonsignificant: χ^2 difference = 4.29 (df = 26, p = 0.99).

Discussion

The current results provide an expansion to the understanding of the HAQ-DI's structural validity and fidelity with its scoring system using rigorous psychometric methods. Along with prior results examining the structural validity for RA patients, single-factor scoring of the HAQ-DI has now been demonstrated to be appropriate for SSc patients. Moreover, comparison of HAQ-DI scores between SSc patients and RA patients is psychometrically justified. Finally, by examining the structural validity of the HAQ's latent structure in both random splits of the SSc patient database (i.e., in the CFA and invariance samples), this study also presented latent analysis in a two-step cross validation. As factor analysis is a sample-dependent technique, the validity of a factor structure must be tested on an independent sample in order for one to have confidence in the results. Therefore, the cross-validation built into the current study should enhance the generalizability of results for SSc patients.

An interesting finding of our study is that the additional impact on motor skills can be interpreted by examining pairs of scores on Grip and Eating as well as Grip and Hygiene. However, it should be noted that the correlation between residuals for Grip and Hygiene was negative (−0.40). Whereas the other correlated residual pair has a more logical interpretation, interpreting negatively correlated residuals between Grip and Hygiene is more difficult and should be examined further with other measures of manual dexterity and hygiene. This same negative residual correlation is consistent with the correlation found by Cole et al. [3] for RA patients on the HAQ-DI.

Fit indices for the current study were based on rigorous criteria from strongly validated studies

[41, 51]. The use of such rigorous criteria is likely to have been appropriate given the research by Marsh et al. [52] in their investigation of the appropriateness for various cutoff rules with model fit in SEM. Although more complex models often warrant various relaxations to the cutoff rules, the current model is simple enough to warrant stricter use for assuring the psychometric efficacy of this tool.

A possible limitation to the current study is that the items for each HAQ-DI domain differ from person to person. This is a necessary and expected aspect of the HAQ-DI and all related psychometric evaluations of the HAQ-DI, as HAQ-DI scoring criteria require one to use the score of the highest item to create the score for the HAQ-DI domains. The influence of this aspect of the HAQ-DI should also be tested. Unfortunately, such a model would be so complex that the sample size demands would make it particularly difficult for SSc studies. Additionally, it was possible that patients with diffuse and limited SSc differed in their response patterns, ultimately requiring individual scoring models. Because the current pooled database had only 65 patients with limited SSc and 326 with diffuse SSc, such differences would be a limitation for analysis of the current database. Both populations feature finger and hand skin thickening and clinical features of peripheral vascular injury including cold sensitivity and ischemic ulcers. Although we did not consider these differences to provide a marked difference between the response patterns between the two SSc subgroups, this was a subjective assumption. Therefore, a *post hoc* and exploratory structural invariance analysis was undertaken to confirm that interpretation of the HAQ-DI did not differ by this particular scleroderma subtype.[4] Results indicated appropriate fit for the fully constrained structural invariance model (see Table 2 results for Model 3). Because these results use the combined CFA and invariance samples, and because the sample size for the limited SSc group is small, these results should be interpreted with appropriate caution.

Finally, one should be cautious regarding the interpretation of the structural invariance between SSc patients and RA patients as implying that the mean HAQ-DI score for these two groups is, and shall remain, identical. Instead, the positive finding for structural invariance provides evidence that the way patients respond to HAQ-DI domains items is the same between these two groups, allowing researchers to compare the means between SSc patients and RA patients.

Two key areas can be addressed in future research: exploration of the validity of the HAQ-DI in other diagnostic groups and exploration of the negatively correlated residual between Grip and Hygiene domains. The current study expanded the structural validity and scoring fidelity for the HAQ-DI to now include SSc patients and RA patients. However, the HAQ-DI also is used to determine the disease-specific functional abilities and QoL in other diagnostic groups, such as osteoarthritis, systemic lupus erythematosus, and other musculoskeletal conditions. Although the current results provide a rationale for using the original RA structural validation model with other disease populations, there is no empirical evidence to suggest that current results are necessarily generalizable to other disease conditions, including other rheumatologic conditions (see 8 for a discussion of sample-specific validity). Byrne [53] has recommended that one should first determine the latent structure of a test on a single and appropriate sample prior to testing the similarity for CFA results across various subgroups. HAQ-DI structural validity is now available on two samples: SSc and RA. Hereafter, it would be beneficial for other research to affirm the latent structure of the HAQ for other disease populations and measure the consistency between those groups with RA or SSc groups [53].

Notes

1. Confirmatory factor analysis is a type of factor analysis that is used to examine the fit between a theoretical model and the data. Typically, CFA is used to confirm that a single (or multiple set of) latent construct(s) (sometimes called factors) are responsible for the way patients respond to items on a test. Because latent constructs cannot be measured directly, CFA uses a set of statistics to measure the correlations between items to determine the fit between the theoretical latent model and the data (for more on this process, see [49]). For the current analyses, we hypothesized that a single latent construct of "Burden on Activities of Daily Living" is present. Although we cannot directly measure this burden, we can infer its presence from the HAQ-DI domains scores.
2. In order to obtain a good likelihood of model convergence and stable results, Bentler and Chou [54] recommended that at least five participants per free parameter be used. The weakest

model in the current study, the revised CFA, had 18 free parameters, and therefore we had 5.56 participants per variable. Although successful convergence and stable results are not a measure of statistical power [55], previous results with RA patients lead us to believe that the very large coefficients and fit would be sufficient if the model results were stable.

3. It should be noted that the mean of the latent construct on the HAQ-DI was allowed to vary between the groups, as mean differences are expected between the scleroderma and RA patients (and this is a useful difference between the models). Moreover, please note that the mean of the latent construct is different from the mean of the total score, although this discussion is beyond the scope of the current manuscript. For more information, please see Byrne [53].
4. This model was run without the latent mean structure incorporated into the model in order to limit the degrees of freedom for the smaller limited SSc group.

Appendix A

On behalf of the Scleroderma Lung Study (SLS) research group, the following investigators also participated:
University of California Los Angeles, Los Angeles, California: Philip J. Clements, MD, MPH; Donald P. Tashkin, MD; Robert Elashoff, PhD; Jonathan Goldin, MD, PhD; Michael Roth, MD; Daniel Furst, MD; Ken Bulpitt, MD; Dinesh Khanna, MD; Wen-Ling Joanie Chung, MPH; Sherrie Viasco, RN; Mildred Sterz, RN, MPH; Lovlette Woolcock; Xiaohong Yan, MS; Judy Ho, Sarinnapha Vasunilashorn; Irene da Costa.
University of Medicine & Dentistry of New Jersey, New Brunswick, New Jersey: James R. Seibold, MD*; David J. Riley, MD; Judith K. Amorosa, MD; Vivien M. Hsu, MD; Deborah A. McCloskey, BSN; Julianne E. Wilson, RN. * Current address: University of Michigan Scleroderma Program, Ann Arbor, Michigan.
University of Illinois Chicago, Chicago, Illinois: John Varga, MD; Dean Schraugnagel, MD; Andrew Wilbur, MD; David Lapota, MD; Shiva Arami, MD; Patricia Cole-Saffold, MS.
Boston University, Boston, Massachussettes: Robert Simms, MD; ArthurTheodore, MD; Peter Clarke, MD; Joseph Korn, MD; Kimberley Tobin, Melynn Nuite, BSN.
Medical University of South Carolina, Charleston, South Carolina: Richard Silver, MD; Marcie Bolster, MD; Charlie Strange, MD; Steve Schabel, MD; Edwin Smith, MD; June Arnold; Katie Caldwell; Michael Bonner.
Johns Hopkins School of Medicine, Baltimore, Maryland: Robert Wise, MD; Fred Wigley, MD; Barbara White, MD; Laura Hummers, MD; Mark Bohlman, MD; Albert Polito, MD; Gwen Leatherman, MSN; Edrick Forbes, RN; Marie Daniel.
Georgetown University, Washington, DC: zVirginia Steen, MD; Charles Read, MD; Cirrelda Cooper, MD; Sean Wheaton, MD; Anise Carey; Adriana Ortiz.
University of Texas Houston, Houston, Texas: Maureen Mayes, MD, MPH; Ed Parsley, DO; Sandra Oldham, MD; Tan Filemon, MD; Samantha Jordan, RN; Marilyn Perry.
University of California San Francisco, San Francisco, California: Kari Connolly, MD; Jeffrey Golden, MD; Paul Wolters, MD; Richard Webb, MD; John Davis, MD; Christine Antolos; Carla Maynetto.
University of Alabama, Birmingham, Alabama: Barri Fessler, MD; Mitchell,Olman, MD; Colleen Sanders, MD; Louis Heck, MD; Tina Parkhill.
University of Connecticut Health Center, Farmington, Connecticut: Naomi Rothfield, MD; Mark Metersky, MD; Richard Cobb, MD; Macha Aberles, MD; Fran Ingenito, RN; Elena Breen;
Wayne State University, Detroit, Michigan: Maureen Mayes, MD; Kamal Mubarak, MD; Jose L Granda, MD; Joseph Silva, MD; Zora Injic, RN, MS; Ronika Alexander, RN.
Virginia Mason Research Center, Seattle, Washington: Daniel Furst, MD; Steven Springmeyer, MD; Steven Kirkland, MD; Jerry Molitor, MD; Richard Hinke, MD; Amanda Mondt, RN.
University of Alabama, Birmingham: Mitchell Olman, MD; Barri Fessler, MD; Colleen Sanders, MD; Louis Heck, MD; Tina Parkhill.
On behalf of the Relaxin study, the following investigators also participated: J Korn, MD, R Simms, MD, P Merkel, MD, Boston University, Boston, MA; NF Rothfield, University of Connecticut Health Center, Farmington, CT; F Wigley, MD, Johns Hopkins University, Baltimore, MD; M Ellman, MD University of Chicago, Chicago, IL; Y Kim, MD, Stanford University, Palo Alto, CA; L Moreland, University of Alabama at Birmingham, Birmingham, AL; RW Silver, University of South Carolina, Charleston, SC; VD Steen, Division of Rheumatology, Georgetown University Medical Center, Washington, DC; M Weisman, MD, Cedar Sinai Medical Center, Los Angeles, CA; GS Firestein, MD, AF Kavanaugh, MD University of California, San Diego, CA; ME Csuka, MD Medical College of Wisconsin, Madison, WI; MD Mayes, University of Texas Health Science Center, Houston, TX; D Collier, University of Colorado, Denver, CO; TA Medsger, Jr., University of Pittsburgh, Pittsburgh, PA; and Vivian M Hsu, UMDNJ-Scleroderma Program.

Acknowledgements

Dr. Khanna was supported in part by the Arthritis and Scleroderma Foundations (Physician Scientist Development Award), the Scleroderma Foundation (New Investigator Grant), a National Institutes of Health K12 BIRWCH Scholar Award, the Scleroderma Clinical Trials Consortium, and the Scleroderma Lung Study by Grant No. UO1 HL60587.

References

1. Medsger TA Jr. System sclerosis: Clinical aspects. In: Koopman W (ed.), Arthritis and Allied Conditions. Baltimore:Williams and Wilkins, 19971433–1464.

2. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980; 23: 137–145.
3. Cole JC, Motivala SJ, Khanna D, Lee JY, Paulus HE, Irwin MR. Validation of a single-factor structure and the scoring protocol for the Health Assessment Questionnaire-Disability Index (HAQ-DI). Arthritis Care Res 2005; 53: 536–542.
4. Khanna D, Clements PJ, Furst DE, Chon Y, Elashoff R, Roth MD. Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with system sclerosis and active alveolitis: Results from the Scleroderma Lung Study. Arthritis Rheum 2005; 52: 592–600.
5. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, et al. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. J Rheumatol 2005; 32: 832–840.
6. Clements PJ, Wong WK, Hurwitz EL, Furst DE, Mayes MD, White B, et al. The disability index of the Health Assessment Questionnaire is a predictor and correlate of outcome in the high-dose versus low-dose penicillamine in system sclerosis trial. Arthritis Rheum 2001; 44: 653–661.
7. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol 1995; 50: 741–749.
8. Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. Psychol Assess 1995; 7: 238–247.
9. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organ Res Method 2000; 3: 4–70.
10. Meade AW, Lautenschlager GJ. A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. Organ Res Method 2004; 7: 361–388.
11. Seibold JR, Clements PJ, Korn JH, Ellman M, Rothfield N, Wigley FM, et al. US phase III trial of relaxin in diffuse scleroderma [Abstract]. J Rheumatol 2001; 63: T-64.
12. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. J Rheumatol 1980; 23: 581–590.
13. Clements PJ, Hurwitz EL, Wong WK, Seibold JR, Mayes MD, White B, et al. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: High-dose versus low-dose penicillamine trial. Arthritis Rheum 2000; 43: 2445–2454.
14. Tashkin DP, Elashoff D, Clements PJ, Golding JM, Roth MD et al. Cyclophosphamide versus placebo in scleroderma lung disease. New Engl J Med (in press).
15. Khanna D, Ranganath VK, Fitzgerald J, Park GS, Altman RD, Elashoff D, et al. Increased radiographic damage scores at the onset of seropositive rheumatoid arthritis in older patients are associated with osteoarthritis of the hands but not with more rapid progression of damage. Arthritis Rheum 2005; 52: 2284–2292.
16. Wu H, Khanna D, Park GS, Gersuk V, Nepom GT, Wong WK, et al. Interaction between RANKL and HLA-DRB1 genotypes may contribute to younger age at onset of seropositive rheumatoid arthritis in an inception cohort. Arthritis Rheum 2004; 50: 3093–3103.
17. Clements PJ, Wong WK, Hurwitz EL, Furst DE, Mayes MD, White B, et al. Correlates of the disability index of the Health Assessment Questionnaire: A measure of functional impairment in systemic sclerosis. Arthritis Rheum 1999; 42: 2372–2380.
18. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Seibold JR. Responsiveness of the health related quality of life instruments (sf-36 and HAQ-DI) in a systemic sclerosis clinical trial [Abstract]. Arthritis Rheum 2003; 48: S398.
19. Kumar A, Malaviya AN, Pandhi A, Singh R. Validation of an Indian of the Health Assessment Questionnaire in patients with rheumatoid arthritis. Rheumatology (Oxford) 2002; 41: 1457–1459.
20. El Meidany YM, El Gaafary MM, Ahmed I. Cross-cultural adaptation and validation of an Arabic Health Assessment Questionnaire for use in rheumatoid patients. Joint Bone Spine 2003; 70: 195–202.
21. El Meidany YM, Youssef S, El Gaafary MM, Ahmed I. Evaluating changes in health status: sensitivity to change of the modified Arabic Health Assessment Questionnaire in patients with rheumatoid arthritis. Joint Bone Spine 2003; 70: 509–514.
22. Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, Keystone EC, et al. A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. New Engl J Med 2000; 343: 1586–1593.
23. Lipsky PE, Heijde DMvan der, St. Clair EW, Smolen JS, Furst JS, Kalden JR, et al. 102-week clinical and radiologic results from the ATTRACT trial: A 2-year, randomized, controlled, phase 3 trial of infliximab in patients with active RA despite MTX. Arthritis Rheum 2000; 43: S269.
24. Weinblatt ME, Keystone EC, Furst DE, Moreland LW, Weisman MH, Birbara CA, et al. Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patient taking concomitant methotrexate: The ARMADA trial. Arthritis Rheum 2003; 48: 35–45.
25. Wolfe F, Hawley DJ. The long term outcomes of rheumatoid arthritis: Work disability: A prospective 18 year study of 823 patients. J Rheumatol 1998; 25: 2108–2117.
26. Wolfe F. The determination and measurement of functional disability in rheumatoid arthritis. Arthritis Res 2002; 4: S11–S15.
27. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patient with rheumatoid arthritis. Arthritis Rheum 2003; 48: 1530–1542.
28. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: Dimensions and practical applications. Health Qual Life Outcomes 2003; 1: 1–6.
29. Westhovens R, Cole JC, Li T, Martin M, MacLean R, Lin P et al. Improved health-related quality of life for

rheumatoid arthritis patients treated with abatacept who have inadequate response to anti-TNF therapy in a double-blind, placebo-controlled, multicenter randomized clinical trial. Rheumatology (in press).

30. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance modeling. Psychol Method 1996; 1: 130–149.
31. Arbuckle JL. Amos. In. 6.0 ed. Chicago: Small Waters, 2005.
32. Cole JC, Motivala SJ, Dang J, Lucko A, Lang N, Levin MJ, et al. Structural validation of the Hamilton Depression Rating Scale. J Psychopathol Behav Assess 2004; 26: 241–254.
33. Cole JC, Rabin AS, Smith TL, Kaufman AS. Development and validation of a Rasch-derived CES-D short form. Psychol Assess 2004; 16: 360–372.
34. Nevitt J, Hancock GR. Improving the root mean square error of approximation for nonnormal conditions in structural equation modeling. J Exp Educ 2000; 68: 251–268.
35. Bollen K, Stine RA. Bootstrapping goodness-of-fit measures in structural equation models. Sociol Method Res 1992; 21: 205–229.
36. Finney SJ, DiStefano C. Nonnormal and categorical data in structural equation modeling. In: Hancock GR, Mueller RO (eds.), Structural Equation Modeling: A Second Course. Greenwich, CT:IAP, 2006269–314.
37. Nevitt J, Hancock GR. Evaluating small sample approaches for model test statistics in structural equation modeling. Multivar Behav Res 2004; 39: 439–478.
38. Schumacker RE, Lomax RG. A Beginner's Guide to Structural Equation Modeling. Mahwah, NJ: Lawrence Erlbaum, 1996.
39. Marsh HW, Hau K-T, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Struct Equation Model 2004; 11: 320–341.
40. Bentler PM, Bonett DG. Significance tests and goodness-of-fit in the analysis of covariance structures. Psychol Bull 1980; 88: 588–606.
41. Hu L-t, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equation Model 1999; 6: 1–55.
42. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equation Model 2002; 9: 223–255.
43. Arbuckle JL, Wothke W. Amos 4.0 User's Guide, 4.01 ed. Chicago: Small Waters, 1999.
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences., 2nd ed., Hillsdale: Lawrence Erlbaum, 1988.
45. Palmer RF, Graham JW, Taylor B, Tatterson J. Construct validity in health behavior research: Interpreting latent variable models involving self-report and objective measures. J Behav Med 2002; 25: 525–550.
46. Anastasi A, Urbina S. Psychological Testing., 7th ed., Upper Saddle River, NJ: Prentice Hall, 1998.
47. Hershberger SL. The problem of equivalent structural models. In: Hancock GR, Mueller RO (eds.), Structural Equation Modeling: A Second Course. Greenwichm, CT:IAP, 200613–42.
48. Muthén BO, Muthén LK. Mplus 3.0 User's Guide. Los Angeles: Muthén and Muthén, 2004.
49. Schumacker RE, Lomax RG. A Beginner's Guide to Structural Equation Modeling., 2nd ed., Mahwah, NJ: Lawrence Erlbaum, 2004.
50. Comrey AL, Lee HB. A First Course in Factor Analysis., 2nd ed., Hillsdale, NJ: Lawrence Erlbaum, 1992.
51. Hu L-t, Bentler PM. Evaluating model fit. In: Hoyle RH (ed.), Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA:Sage, 1995.
52. Marsh HW, Hau K-T, Grayson D. Goodness of fit in structural equation models. In: Maydeu-Olivares A, McArdle JJ (eds.), Contemporary Psychometrics. Mahwah, NJ:Lawrence Erlbaum, 2005275–340.
53. Byrne BM. Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming. Mahwah, NJ: Lawrence Erlbaum, 2001.
54. Bentler PM, Chou C. Practical issues in structural modeling. Sociol Method Res 1987; 16: 78–117.
55. Hancock GR. Power analysis in covariance structure modeling. In: Hancock GR, Mueller RO (eds.), Structural Equation Modeling: A Second Course. Greenwich, CT:IAP, 200669–118.

Address for correspondence: Consulting Measurement Group 7071 Warner Ave., #F-400 Huntington Beach CA 91403 USA
Phone: + 866-782-8799
E-mail: jcole@webcmg.com)