Defining Hospital Anxiety and Depression Scale (HADS) structure by confirmatory factor analysis: a contribution to validation for oncological settings

M. A. Annunziata1*, B. Muzzatti1 & G. Altoè2

1Unit of Oncological Psychology, Centro di Riferimento Oncologico, IRCCS Istituto Nazionale Tumori, Aviano; 2Department of Developmental and Social Psychology, University of Padova, Padova, Italy

Received 16 March 2010; revised 23 June 2010 & revised 23 August 2010; accepted 29 November 2010

Background: Despite its popularity, not a definitive word has yet been said about the latent structure of the Hospital Anxiety and Depression Scale (HADS). The present work is a contribution to this issue: statistically identifying the best tool structure helps in understanding which constructs the tool actually detects.

Participants and methods: Five hundred and twelve Italian consecutive hospitalized cancer patients completed the HADS together with a form for the collection of personal and clinical data. Two confirmatory factor analyses (CFAs) were carried out to test the validity of both two- and one-factor models, whereas qualitative and quantitative (i.e. Akaike information criterion) indices were used to assess which model among them would fit better with the observed data. Finally, two multigroup CFAs were carried out to test the factorial invariance across gender and disease phase (diagnostic, therapeutic) of the best-fitting model.

Results: Although both considered models provide a good fit to the observed data, the two-factor model is more adequate; it is invariant across gender and disease phase.

Conclusions: The present study gives evidence for using HADS to detect anxious and depressive states separately as originally suggested by its authors. Given that this work involved only Italian cancer inpatients, replications in different cultural/national contexts are recommended.

Key words: confirmatory factor analysis, Hospital Anxiety and Depression Scale, multigroup confirmatory factor analysis, neoplasm, psychometrics

Introduction

The Hospital Anxiety and Depression Scale (HADS) [1–3] is a well-known emotional distress self-report measure and it is one of the most frequently used in oncological settings [4–6] as well as in other somatic diseases (e.g. cardiology, brain injury, general medicine). Originally, it was designed to screen emotional disorders of patients in nonpsychiatric settings by detecting the two most frequent distress expressions: anxious and depressive states. Since it is specific to patients with organic disease, HADS excludes somatic symptoms of emotional distress (e.g. headache, weight loss, insomnia) that could be caused by the illness itself rather than being emotional distress expressions [1, 2, 4, 5]. Furthermore, severely psychopathological symptoms are not covered to improve HADS’s sensitivity to medical conditions [1, 2, 4]. Thus, HADS is a prolonged state rather than a trait measure [4] and it is not recommendable in detecting psychopathological disorders.

It is a Likert scale composed of 14 items to which patients respond through a 4-point scale (from 0 to 3) referring to overt symptoms within the last week.

HADS seems to be the scale of choice for clinical practice in oncology due to the following characteristics: (i) its mentioned specificity to patients with organic disease [1, 2, 4, 5], (ii) it is short and rapid in administration (taking 2–5 min) [2, 4], (iii) its psychometric properties are described as good by most pertinent literature (e.g. [2–6]), and (iv) it is generally well accepted by users [4–6].

In spite of this, there are still two open questions concerning its use in daily practice: first, not a definitive word has yet been said about the latent structure of HADS; thus, it is not clear if it is better used as a one-dimensional measure (of major depression and of both adjustment and depressive disorders) or as separate anxiety and depression indexes. Secondly, the literature provides a large range of cut-off rankings used by different authors for distinguishing cases from noncases (see [6] for an updated review).

The present work is a contribution to the definition of the HADS latent structure: statistically identifying the best (i.e. better fit with data) tool structure helps in understanding
which constructs the tool actually detects. Furthermore, it represents the necessary precondition to tackle the question about cut-off, i.e. to indicate above which HADS score is the users’ distress clinically relevant and, consequently, which patients would benefit from psychological supportive care.

As mentioned, Zigmond and Snith [1] designed the instrument to screen two different emotional conditions: anxious and depressive states. Several authors (e.g. [7–9]) used HADS in this way, whereas others (e.g. [10–13]) used the scale as one dimensional: identifying major depression and adjustment plus depressive disorders. In addition, other authors [14] identified a trifactor structure (corroborated by research in nononcological clinical samples, e.g. [15, 16]). Finally, several authors [17] recommended using HADS as a one-dimensional measure despite providing statistical evidence for a three-factor structure, and others provided both global and anxiety/depression scores (e.g. [18–20]).

Although many papers (e.g. [5, 6, 14]) have repeatedly highlighted a HADS structure inconsistency throughout literature and, consequently, in the measured psychological constructs, only few research papers address the question in a methodologically appropriate way. One of these studies was conducted by Smith et al. [21] who demonstrated in a very large sample of cancer patients (N = 1474) the adequacy of the two-factor structure (anxiety and depression subscales) by performing a principal component analysis with Oblimin axes rotation; in replicating the same analysis across subgroups, they concluded that this structure was stable despite gender, age, and disease stage (see also [22]). In a subsequent study, Rodgers et al. [14] tested seven different factorial models (i.e. one, bi, and three) for the scale by confirmatory factor analyses (CFAs) and concluded that the best fit with the observed data was provided by a three-factor structure. The relatively small sample size (i.e. 110 subjects) together with its great homogeneity (all participants were undergoing adjuvant treatment of breast cancer) should induce caution in generalization. These two limitations were overcome by Smith et al. [23] who analyzed HADS data from 1855 patients affected by a large range of cancers. They carried out testing on Rasch models and concluded that the HADS formed a one-dimensional construct of emotional distress and that the two subscales were one-dimensional scales or factors embedded within this structure. In spite of its very large sample and the complex analyses carried out, this study cannot be considered conclusive about the HADS dimensional structure because it involves a technique (Rasch model), which is more adequate in confirming one dimensionality than in comparing different dimensional structures.

The discussed HADS characteristics and properties together with the mentioned limitations in previous studies prompt us to further deepen knowledge on HADS. The present work is first aimed to verify the HADS factorial structure by applying CFA to a large sample of data, collected in an oncological setting, and comparing the resulting structure with a one-dimensional model. Secondly, the structure invariance of the best obtained factorial model is tested by multigroup analyses according to the most relevant distress moderators (i.e. gender and disease phase). Finally, because differences in latent structure could suffer in part from cross-cultural and/or cross-national factors, a supplementary interest of the present study concerns the international debate on cross-national and cross-cultural validity of psychological measures.

method

participants

Five hundred and forty-four consecutive oncolgical patients participated in the study. The eligibility criteria of the incidental cases were the following: (i) being hospitalized in a same oncological hospital in northeastern Italy, (ii) age between 18 and 75 years, (iii) being in a diagnostic phase (within a month of diagnosis) or in a therapeutic phase (3–6 months from diagnosis), (iv) the absence of handicaps, psychiatric syndromes, or temporary inability to compile forms, and (v) having a good understanding of the Italian language.

Potential participants were selected by consulting clinical files, whereas actual involvement in the research was subordinated to the signing of a written permission. The study obtained permission through the Institute’s ethical review board.

Data from 52 participants were omitted from the final dataset since they did not complete the HADS entirely; consequently, the complete data are composed of 512 statistical units. The final sample included 299 males (58.4%) and 213 females (41.6%) whose median age was 54 years (range 18–74). Concerning the disease data, 316 patients (61.7%) were in a diagnostic phase and 196 (38.6%) were in a therapeutic phase. The most representative types of cancer were gastrointestinal tumors (30%), respiratory tumors (24.7%), and hematological tumors (13.7%).

materials and procedure

The HADS (together with the personal identification and clinical data collection form and the informed consent form for trial participation) was administered by a psychologist the day after the patient’s hospitalization and they were collected the following day at which time the researcher was available to hear doubts or problems related by the participants.

statistical analyses

First, two CFAs [24] were carried out on the whole sample to test the validity of the two-factor model presented in Figure 1 and of a one-factor model that simply loaded all 14 items on to a single latent construct. Since the observed variables included in the two models were ordinal, the weighted least squares method in the LISREL program (LISREL 8.71, 2004) was used; starting from the polychoric correlation matrix and the estimated asymptotic covariance matrix as suggested by Jöreskog and Sörbom [24]. Assessment of fit was based on several indices. Since the χ² statistic is dependent on sample size, two relative fit indices have been considered: the nonnormed fit index (NNFI) and the comparative fit index (CFI), as they both perform well with small and large samples. Values above 0.95 are usually considered satisfactory [25]. The root mean square error of approximation (RMSEA) was also used. This is an absolute fit index assessing approximation of parameter estimates to true parameters in the population. Values below 0.08 reflect acceptable fit [25].

Secondly, the fit of the two-factor model was compared with that of the one-factor model to select the best-fitting model. The models were compared using both a qualitative evaluation of the fit indices and the Akaike information criterion (AIC) [26, 27]. As suggested by Schermelle et al. [25], AIC can be used to compare the competing models. In particular, given a set of models for the same data, the model with the minimum AIC value is regarded as the best-fitting model.

Finally, two multigroup confirmatory factor analyses (MGCFAs) were carried out to test the factorial invariance across gender and disease phase for the best-fitting model [24]. Two aspects of factorial invariance were assessed: configural invariance and metric invariance.
As configural invariance requires that both genders display the same number of factors as well as identical corresponding items, a baseline model implying an equal number of factors across gender was carried out. In this case, a good-fitting model suggests configural invariance. In addition to the requirements of configural invariance, metric invariance has the requisite of equivalent factor loadings across groups. The establishment of metric invariance ensures that items have the same meaning for groups being compared. In order to assess metric invariance, a nested model test, in which every single factor loading was constrained to be equal across gender, was carried out. The ΔCFI index (i.e. the difference in the CFIs between configural invariance model and metric invariance model) was used to examine metric invariance. An absolute value of ΔCFI ≤ 0.01 supports the hypothesis of metric invariance [28]. The same procedure was used to test factorial invariance for the disease phase.

results

The two-factor model provided a good fit to the observed data. Although the Satorra–Bentler chi-square was significant ($\chi^2_{SB} (76, N = 512) = 194.14, P < 0.01$), the other indices (NNFI = 0.98; CFI = 0.99; RMSEA = 0.06) pointed to a satisfactory fit, having NNFI and CFI above 0.97 and RMSEA below 0.08 [25]. Furthermore, as it is shown in Table 1, all estimated factor loadings were significant and generally high, confirming good levels of internal factor structure validity. On the other hand, the one-factor model provided less adequate fit indices ($\chi^2_{SB} (77, N = 512) = 370.19, P < 0.01$; NNFI = 0.96; CFI = 0.97; RMSEA = 0.09). In particular, the AIC model comparison (AIC$_{two-factor} = 252.14 < AIC$_{one-factor} = 426.19) showed that the two-factor model could be considered the best-fitting model and, in the present case, it is more likely to reflect the underlying factor structure that generated the observed data than the one-factor model.

Both the baseline models tested to verify the configural invariance of the two-factor model across gender ($\chi^2_{SB} (152, N_{males} = 299, N_{females} = 213) = 256.84, P < 0.01$; NNFI = 0.99; CFI = 0.99; RMSEA = 0.06) and disease phase ($\chi^2_{SB} (152, N_{diagnostic} = 316, N_{therapeutic} = 196) = 253.06, P < 0.01$; NNFI = 0.99; CFI = 0.99; RMSEA = 0.05) showed a good fit to the observed data. Thus, it could be stated that the same number of factors reflected by the same set of indicators were present across gender and disease phase. Finally, for metric invariance exams, CFIs of the baseline models were compared with the corresponding CFIs of the metric invariance models. Both for gender and disease phase, the absolute value of ΔCFI was <0.01, supporting the hypothesis of metric invariance.

In summary, the two-factor model (i) provides a good fit to the observed data, (ii) is more adequate than the one-factor model, and (iii) is invariant across gender and disease phase.

discussion

HADS is one of the most frequently used emotional distress measures in oncological settings [4–6]. Despite it, several questions are still open about its validity; in particular, literature is not consistent about the HADS latent structure [1, 5–14, 17–23]. Given that identification of the best tool structure helps in understanding which constructs the tool actually detects and that understanding the actual detected construct is the necessary prerequisite to define the clinically relevant cut-off (i.e. to indicate above which HADS score users’ distress is clinically relevant and, consequently, which patients would benefit from psychological supportive care), identifying HADS latent structure is both theoretically and practically relevant. The present study is a contribution to this important question.

CFAs carried out on a large sample data showed that both the one-factor and the two-factor models are fitting well with the observed data. However, the considered fit indices were better in the two-factor than in the one-factor model; this datum was also confirmed by the AIC. Thus, according to this data, using HADS as a two-factor scale is more appropriate on Italian oncological patients.

Table 1. Standardized factor loadings and error variances of the two-factor model

<table>
<thead>
<tr>
<th>Item</th>
<th>Anxiety</th>
<th>Depression</th>
<th>Error variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>0.80</td>
<td>–</td>
<td>0.36</td>
</tr>
<tr>
<td>Item 2</td>
<td>0.73</td>
<td>–</td>
<td>0.47</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.74</td>
<td>–</td>
<td>0.46</td>
</tr>
<tr>
<td>Item 4</td>
<td>0.64</td>
<td>–</td>
<td>0.59</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.72</td>
<td>–</td>
<td>0.48</td>
</tr>
<tr>
<td>Item 6</td>
<td>0.53</td>
<td>–</td>
<td>0.72</td>
</tr>
<tr>
<td>Item 7</td>
<td>0.75</td>
<td>–</td>
<td>0.44</td>
</tr>
<tr>
<td>Item 8</td>
<td>–</td>
<td>0.68</td>
<td>0.53</td>
</tr>
<tr>
<td>Item 9</td>
<td>–</td>
<td>0.73</td>
<td>0.47</td>
</tr>
<tr>
<td>Item 10</td>
<td>–</td>
<td>0.80</td>
<td>0.36</td>
</tr>
<tr>
<td>Item 11</td>
<td>–</td>
<td>0.55</td>
<td>0.70</td>
</tr>
<tr>
<td>Item 12</td>
<td>–</td>
<td>0.55</td>
<td>0.69</td>
</tr>
<tr>
<td>Item 13</td>
<td>–</td>
<td>0.65</td>
<td>0.58</td>
</tr>
<tr>
<td>Item 14</td>
<td>–</td>
<td>0.64</td>
<td>0.60</td>
</tr>
</tbody>
</table>

All standardized factor loadings, error variances, and the correlation between anxiety and depression (0.81) estimated are significantly above the $P <0.01$ level ($N = 512$).

Figure 1. The theoretical two-factor model ($N = 512$).
In addition, two distinct MGCFAs were run describing the two-factor model as invariant across gender and disease phase (diagnostic, i.e. within a month of diagnosis or therapeutic, i.e. 3–6 months from diagnosis). In other words, gender neither determined differences in number of factors nor in corresponding items (configural invariance) and items had the same meaning for males and females (metric invariance). Analogous patterns were found in the disease phase (diagnostic or therapeutic).

In summary, a rigorous methodological approach gives evidence for using HADS to detect anxiety and depression separately as originally suggested by its authors [1]. Furthermore, because the present data concern Italy (i.e. a national context different from the context originally studied), they gave some evidence also on the cross-national and cross-cultural HADS validity. Since the disposition for expressing psychological features (e.g. quality of life, emotions, mood states) can suffer from nurture as well as cultural habits, checking cross-cultural and cross-national validity of a tool is recommended in psychology [29].

The main strength of the present work is the nature of the analyses and the large sample data that have made them possible. Conversely, the main limitation depends on the medical characteristics of the involved patients: we tested the HADS structural invariance through two disease phases (i.e. diagnostic and therapeutic), further research should take into account subsequent phases such as advanced cancer or relapse as well as the subgroups could be based upon the extent of the disease matters (e.g. local disease, locoregional disease, metastatic disease) and/or the different cancer types. Analogously, replications for other somatic disease groups would surely be useful. Finally, the Italian background of all participants suggests corroboration of present findings from other national and cultural contexts.

In conclusion, the present work supports using HADS as separate indexes of anxious and depressive states. Furthermore, its findings can represent the first step in tackling the other mentioned open question about HADS, i.e. the determination of cut-off capable of distinguishing clinically relevant cases from noncases.

disclosure

The authors declare no conflict of interest.

references