Royal College of Surgeons in Ireland

repository@rcsi.com

The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis

AUTHOR(S)

Sam Norton, Theodore Cosco, Frank Doyle, John Done, Amanda Sacker

CITATION

Norton, Sam; Cosco, Theodore; Doyle, Frank; Done, John; Sacker, Amanda (2012): The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis. Royal College of Surgeons in Ireland. Journal contribution. https://hdl.handle.net/10779/rcsi.10772612.v2

HANDLE

10779/rcsi.10772612.v2

LICENCE

CC BY-NC-ND 4.0

This work is made available under the above open licence by RCSI and has been printed from https://repository.rcsi.com. For more information please contact repository@rcsi.com

URL

https://repository.rcsi.com/articles/journal_contribution/The_Hospital_Anxiety_and_Depression_Scale_A_meta_confirmatory_factor_analysis/10772612/2

Please cite this article as: Norton S, Cosco T, Doyle F, Done J & Sacker A. The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis. *Journal of Psychosomatic Research*, 2012, http://dx.doi.org/10.1016/j.jpsychores.2012.10.010

The Hospital Anxiety and Depression Scale: a meta confirmatory factor analysis

Sam Norton, PhD

Psychology Department, Institute of Psychiatry, King's College London, UK

Theodore Cosco, MSc

Department of Public Health & Primary Care, Institute of Public Health, University of Cambridge, UK

Frank Doyle, PhD

Division of Population Health Sciences, Royal College of Surgeons in Ireland, Ireland

John Done, PhD

School of Psychology, University of Hertfordshire, UK

Amanda Sacker, PhD

Institute for Social & Economic Research, University of Essex, UK

Correspondence should be addressed to Sam Norton, Psychology Department, Institute of Psychiatry, King's College London, 5th Floor Bermondsey Wing, Guy's Hospital Campus, London Bridge, London, SE1 9RT. Tel: +44 (0) 20 7188 0198. E-mail: sam.norton@kcl.ac.uk

Abstract

Objective: To systematically evaluate the latent structure of the Hospital Anxiety and Depression Scale (HADS) through reanalysis of previous studies and meta confirmatory factor analysis (CFA).

Method: Data from 28 samples were obtained from published studies concerning the latent structure of the HADS. Ten models were considered, including eight previously identified models and two bifactor models. The fit of each model was assessed separately in each sample and by meta CFA. Meta CFA was conducted using all samples and using subgroups consisting of community samples, cardiovascular disease samples and samples from studies administering the English language version of the HADS.

Results: A bifactor model including all items loading onto a general distress factor and two orthogonal anxiety and depression group factors provided the best fit for the majority of samples. Meta CFA provided further support for the bifactor model with two group factors. This was the case using all samples, as well as all subgroup analyses. The general distress factor explained 73% of the covariance between items, with the (autonomic) anxiety and (anhedonic) depression factors explaining 11% and 16%, respectively.

Conclusion: A bifactor structure provides the most acceptable empirical explanation for the HADS correlation structure. Due to the presence of a strong general factor, the HADS does not provide good separation between symptoms of anxiety and depression. We recommend it is best used as a measure of general distress.

Keywords: Hospital Anxiety and Depression Scale; factor analysis; meta-analysis; anxiety; depression

Introduction

The Hospital Anxiety and Depression Scale (HADS) is a widely used measure of psychological distress designed for use in non-psychiatric patient populations [1]. Numerous studies examining its construct validity using exploratory (EFA) and confirmatory (CFA) factor analysis and item response theory (IRT) methods in clinical and non-clinical populations have been published. The findings of these studies have been summarised in several reviews [2–5]. However, disagreement about the underlying dimensionality of the HADS remains and concerns regarding the apparent lack of consistency between studies have led to calls for the abandonment of the HADS [6]. Inconsistency is likely to be partially due to the application of different methodologies between studies.

As we have previously suggested, the reason for the apparent inconsistency between methods may also be due to the presence of a general distress factor [7]. Two alternative hierarchical models – higher-order and bifactor – have been proposed to represent the structure of scales with a general factor. The most widely supported three factor structure in the Cosco et al. [5] review of the HADS is a higher-order model, proposed by Dunbar and colleagues [8], based on the tripartite theory of anxiety and depression [9]. The tripartite theory posits that a higher-order general somatopsychic distress trait, negative affectivity (NA), accounts for the observed association between anxiety and depression. The Tripartite theory stipulates that the specific component of anxiety is autonomic arousal marked by somatic symptoms, and that the main component of depression is anhedonia characterized by low positive affect—a loss of pleasure and interest in life and inability to feel please even when engaging in pleasurable activities [10]. As was noted by Dunbar and colleagues [8], several of the HADS items fit these constructs well.

The crucial difference between higher-order and bifactor models lies in the ability to separate the variance accounted for by the general factor from the other factors. Higher-order models are composed of first-order factors (e.g. autonomic arousal, anhedonia) onto which the observed items load and higher-order factors (e.g. distress/NA) onto which the first-order factors load. This superordinate higher-order factor is assumed to cause the correlation between the observed items loading on different scales. Where the higher-order factor explains a large proportion of the variance in the lower-order factors (i.e. the first-order factors are highly correlated) the use of the total score, calculated by summing across the subscales, is a valid measure of the higher-order factor. In this situation it is difficult to disentangle the separate effects of the first-order factors from any higher-order factor. For example, it is unclear whether the association between depression, anxiety and anger with cardiovascular disease is due to specific components of these overlapping affective constructs or by some general negative affective disposition [11].

Bifactor models consist of a broad general factor, such as distress or depression, onto which all observed items load and conceptually narrower group factors onto which observed items with related content load [12]. The essential difference is that the general factor is at the same level conceptually as the group factors, thus allowing for the parsing of the variance explained by the general and group factors. This enables the evaluation of the specific components independent contribution to prediction, of say incident cardiovascular disease, controlling for the general factor. As with the higher-order model, where there is a strong general factor a combined overall score is a valid measure of the general factor. A bifactor version of the tripartite model would consist of a general factor and group factors relating to autonomic arousal and anhedonia, and would allow for the separation of item

variance into the variance explained by the general factor and each of the group factors. Also known as *group-factor* models, bifactor models were initially developed in intelligence research [13] but have since been applied to the study of psychological distress [14–18]. A bifactor structure has not yet been considered in relation to the HADS.

The present study involves a re-analysis and meta-analysis of previously published studies considering the latent structure of the HADS. The goal is to examine whether uncertainty regarding the latent structure of the HADS is due to differences between the methods and samples of earlier studies. Some of the ambiguity in previous studies may have arisen from the methods used, both in terms of the overarching method (EFA, CFA and IRT) and also differences in the structures compared within CFA studies. A re-analysis will enable the comparison of all previously identified 'best fitting' latent structures using CFA, thus minimizing any ambiguity that might have arisen from the application of different methods. As well as the 'best fitting' structures, the bifactor structure will also be considered since this has shown optimal fit for similar instruments. The meta-analysis involves pooling inter-item correlation matrices from samples used in previous studies, which are then subjected to CFA. Subgroup analysis considering samples drawn from specific populations (community samples, cardiovascular disease samples) and also only studies using English translations of the HADS will further allow for the consideration of whether ambiguity may be due to different latent structures across populations.

Methods

Sample

Data were drawn from studies selected for inclusion in a recent systematic review of the HADS [5], which included studies published between 2002 and 2010. Where the inter-item

correlation matrices were not included in the original publication, the corresponding author was contacted by email. Where no response was received from the corresponding author within 4 weeks, a follow-up email was sent. If the corresponding author's email address was no longer active, attempts were made to identify a current email address via their affiliated institution or by contacting other authors.

In addition, seven studies meeting the recruitment criteria for the Cosco et al. review but published since the literature search was conducted were also included [19–26]. Again, the corresponding author was contacted if the inter-item correlation matrices were not included in the publication.

In total, we attempted to obtain summary data relating to 54 published studies concerning the latent structure of the HADS (Figure 1). We failed to obtain any response in relation to 26 studies and the corresponding author was not able to provide the required information for a further 8 studies. Data concerning 21,820 individuals across 21 studies with 28 unique samples were obtained. Information regarding the 21 studies that provided data is given in Table 1.

Hospital Anxiety and Depression Scale

The HADS is a self-administered scale consisting of 14 items split across anxiety and depression subscales, each with a four-point ordinal response format. To reduce the risk of a false positive bias, the scale does not assess symptoms of anxiety and depression related to physical disorder, such as fatigue and insomnia. The HADS has been shown to have adequate diagnostic accuracy. A recent meta-analysis of diagnostic test accuracy studies reported that, using a score of 8 or more as the cut-off, the HADS depression scale gave 82%

sensitivity and 74% specificity for detecting major depressive disorder; and the anxiety scale gave 78% sensitivity and 74% specificity for detecting generalised anxiety disorder [27].

Statistical analysis

A total of eight different factor structures have been suggested in the literature as providing the *'best fit'* to the HADS item structure (1 unidimensional distress, 2 bidimensional consisting anxiety and depression, and 5 tridimensional consisting of anxiety depression and restlessness/agitation/negative-affectivity) [1,8,28–32]. Furthermore, two bifactor structures were considered since this model has been shown to provide the best fit in other studies jointly assessing symptoms of anxiety and depression [17,18]. The first bifactor model included two group-factors consisting of the anxiety and depression items, respectively. The second bifactor model consisted of three group-factors – depression, anxiety and restlessness. The restlessness factor involved the items common to the restlessness/agitation/negative-affectivity factors of the previously identified three factor models with related meaning – A7 "*I can sit at ease and feel relaxed*", A11 "*I feel restless as if I have to be on the move*", and D14 "*I can enjoy a good book or radio or TV programme"*. In total, ten different structures were considered. The pattern of item loadings on each factor is provided in Figure 2.

Higher-order models with two or three lower-order factors may only be estimated using equality constraints on the loadings of the first-order factors onto the second-order factor [33,34]. The resulting model is mathematically equivalent to a model without the higher-order and allowing the lower-order factors to correlate [34]. Therefore, we can consider the comparison of the models nine and ten, the two bifactor models, to models two through seven as a direct comparison of a bifactor versus a higher-order structure. It is also

important to note that the higher-order model can be considered nested within the bifactor model [34], with the direct effects of the general factor on the observed items constrained to zero. Standard likelihood ratio tests can be applied to assess whether this restriction is appropriate only when four or more lower-order factors are present, hence its omission from the current analysis.

Recently, it has been suggested that the inclusion of an item wording method factor may further increase model fit in CFA of the HADS [25,26]. Therefore, each model was also estimated with the inclusion of a method factor to account for the positive wording of items A7, D2, D4, D6, D12, D14.

Initially each of the structures was considered in relation to the item correlation matrices for each of the 28 separate samples. The fit statistics for each model in each of the samples, along with the sample correlation matrices, is provided in Appendix A.1. Goodness-of-fit was assessed using the χ^2 test of exact fit, the root mean squared error of approximation (RMSEA), standardised root mean squared residual (SRMR), Comparative Fit Index (CFI), and the Tucker–Lewis Index (TLI). Values of RMSEA and SRMR closer to 0 indicate better fit, values less than .08 considered *acceptable* fit [35]. For CFI and TLI values closer to 1 values indicate better fit, values greater than 0.95 indicating *good* fit [35]. To aid model comparison the Akaike (AIC) and Bayesian Information Criterion (BIC) are also presented, smaller values indicate better fit.

Inter-item correlation matrices were pooled using the method described by Cheung and Chan [36], implemented using the R package metaSEM [37]. Pooled matrices were generated for all studies, and also for subgroups consisting of English language studies only, studies consisting of community samples and cardiovascular disease samples. The fit of each of the ten structures described above was then tested in each pooled correlation matrix (a meta

CFA), again with and without a wording method factor. Appendix A.2 provides information concerning model fit for the pooled analysis.

Results

Within sample analysis

Confirmatory factor analysis was applied to the inter-item correlations matrices for each of the 28 samples. The eight models derived from the literature and the two bifactor models were considered, all with and without item wording methods factors. The frequency with which each of the models was selected as superior by various goodness-of-fit criteria are presented in Table 2, for models without the inclusion of an item wording method factor. Appendix A.1 provides model fit statistics for each of the models in each of the samples.

The bifactor model with all items loading onto a general distress factor and two orthogonal anxiety and depression group-factors (model 9) was indicated as providing the best fit to the majority of studies – according to RMSEA (85.7%), SRMR (96.4%), CFI (89.3%), TLI (78.6%), and AIC (82.1%). Considering BIC, although the bifactor model with two group-factors was still most frequently identified as superior (39.3%), the higher-order Dunbar model (model 8) showed a superior fit in six samples (21.3%). BIC penalizes for model complexity and it would seem that, by this measure, the increased fit observed for the bifactor model was not always balanced by the number of additional parameters estimated.

Considering only the eight factor structures considered previously in the literature (models 1 to 8), the Dunbar model and its higher-order variant were observed to provide superior fit in around half of the samples. Notably however differences in fit were often marginal for several of the proposed three factor structures. For the Caci, Friedman, Dunbar

and Dunbar higher-order models fit statistics were typically equal to two decimal places. This goes some way to explaining the apparent lack of agreement in previous studies.

Item wording method factor

Including an item wording method factor in general improved model fit, compared to the same model without a method factor. Again, there was strong support for the bifactor model with two group-factors (model 9). According to RMSEA, SRMR, CFI, TLI, and AIC the bifactor model with two group-factors provided superior fit in over 80% of the samples. BIC however favoured the bifactor model in only eight samples (28.6%), with the higher-order variant of the Dunbar model (model 8) favoured in nine samples (32.1%).

Meta CFA

Using the method described by Cheung and Chan [36] the correlation matrices from each of the 28 samples were pooled. The pooled correlation matrix provided an approximate fit to the original matrices with statistics within acceptable limits (RMSEA = .058; SRMR = .111; CFI = .925; TLI = .928). This indicates that the matrices were homogenous enough for meta-analysis to be appropriate.

Confirmatory factor analysis of the pooled correlation matrix across all samples indicated the bifactor model with two group-factors (model 9) as that providing superior fit, compared to the eight previously suggested structures and the bifactor model with three group-factors (Table 3). All fit statistics were within acceptable limits. Including a method factor to account for item wording effects resulted in improvement in model fit for all models. The fit of the bifactor model with two group-factors remained superior according to all criteria.

A further reason to support the bifactor model was the high inter-factor correlations observed for those models involving only lower-order factors (models 2 to 7). For the two factor models (models 2 & 3) the correlation between the anxiety and depression factors were >.73 (>.80 controlling for wording effect). For the three factor models (models 4, 5, 6 & 7) correlations between anxiety and depression were >.65 (>.73 controlling for wording effect). Correlations between the anxiety and restlessness/agitation/negative-affect factors were >.80 (>.85 controlling for wording effect). This indicates poor separation between the factors, particularly when considering the anxiety and restlessness/agitation/negative-affect.

For the bifactor model with two group-factors, with and without a wording method factor, the proportion of the common and total variance explained by the factors, the factor loadings and the communality estimates for the items (R^2) are presented in Table 4. The loadings of each item on the general factor were all moderate to high (range .37 to .69) but tended to be low to moderate for the group-factors. Only items A3, A9 and A13 loaded greater than .3 on the anxiety factor, and only items D2, D4 and D12 loaded greater than .3 on the depression factor. This suggests that the other items tap into the general distress component underlying anxiety and depression rather than the specific components of autonomic arousal and anhedonia the HADS was designed to measure.

The general factor accounted for 70% and 73% of the common variance with and without an item wording method factor, respectively, indicating that the saturation of the HADS by the general factor is relatively high. The anxiety and depression group-factors were associated with relatively modest contributions to the common variance accounting for 11% and 16%, respectively, dropping to 6% and 13% with the inclusion of an item wording method factor.

Sub-group analyses

Analyses were run on the pooled correlation matrices for several more homogenous groups, consisting of all English language samples, community samples and cardiovascular samples. The pooled correlation matrices for each subgroup provided an approximate fit to the original matrices (English language: RMSEA = .047; SRMR = .095; CFI = .954; TLI = .958; Community: RMSEA = .046; SRMR = .086; CFI = .940; TLI = .947; Cardiovascular: RMSEA = .048; SRMR = .057; CFI = .955; TLI = .966). As would be expected, the fit statistics for the pooled versus original matrices for each subgroup were improved, indicating less heterogeneity, compared to the previous analysis of all samples. Nevertheless, for each sub group meta CFA the bifactor model with anxiety and depression group factors was observed to provide the best fit, by all goodness of fit indices. Details of model fit for each sub group meta CFA are provided in Appendix A.2.

Discussion

Meta CFA of the latent structure of the HADS provided evidence for a bifactor structure consisting of a general distress factor and anxiety and depression group factors. This structure was also supported by reanalysis of the majority of previous studies considering the latent structure of the HADS. This finding indicates that, although tapping into autonomic arousal and anhedonia as originally intended, the HADS is saturated by the presence of a strong general factor, accounting for over 70% of the common variance. In addition, the inclusion of an item wording method factor improved model fit.

The findings of this study are broadly in line with the conclusions of previous reviews of CFA studies of the HADS, where a three factor structure splitting the anxiety subscale into two factors was generally preferred [2,4,5]. In both the reanalysis and meta-analysis reported

here, such a three factor structure was generally observed to provide the best fit to the data when only the eight latent structures considered previously in the literature were considered, However, a bifactor structure not previously considered provides a more optimal solution than any previously considered structure. Furthermore, a bifactor structure goes someway to explaining the conflicting findings of previous studies using different analytic methods. The first unrotated factor in EFA studies indicates a general factor but the aim of finding a simple structure typically led to rotation and the anxiety-depression factorization [3]. CFA studies have generally supported a three factor structure [5], often involving a higher-order negative affectivity factor or otherwise two extremely highly correlated anxiety factors. Both solutions are understandable in the presence of a strong general factor. The interpretation of the general factor in the bifactor model is equivalent to the interpretation of the general factor in the Dunbar higher-order model [34], and the extraction of a third factor related to restlessness or agitation is likely to represent over extraction due to the failure to account for the general factor. Furthermore, IRT studies find that a unidimensional solution is appropriate [5], which is reasonable given the high saturation of the HADS by the general factor.

Dunbar and colleagues suggested that as well as measuring autonomic anxiety and anhedonic depression, the HADS taps into negative affectivity [8]. Although several items that are closely related to the negative affectivity construct had high loadings on the general factor and low loadings on the anxiety factor (tension: A1; restlessness: A7 and A11; worry: A5), all HADS items had high loadings on the general factor, including those specific to anxiety and depression. This would seem to suggest that although the general factor does capture elements of negative affectivity it is better conceptualized as a measure of general psychological distress – the shared variance between symptoms of anxiety and depression.

This interpretation of the general factor and support for a bifactor structure is provided by recent research extending the Tripartite theory. Simms and colleagues used IRT based full-information factor analysis on the Inventory of Depression and Anxiety Symptoms [17]. The authors found that a bifactor solution fitted the data better than a unidimensional Rasch model, identifying 13 specific factors relating to symptom groups such as suicidality, dysphoria, panic and generalized anxiety. Comparing the general factor to other instruments they concluded that the general factor measured general distress, although it was strongly associated with anxiety, depression and negative affectivity. It would seem likely that this is also true of the HADS. This corroborates previous studies examining the hierarchical structure of mood and anxiety disorders [38–40]. Specifically, the present study confirms that much of the variance in psychological distress is shared across a range of symptoms, though groups of symptoms do have a unique component that should not be ignored. Furthermore, due to the equivalence of models two through seven to their higher-order variants [34], we can conclude that the bifactor model provides a better hierarchical explanation of the structure of the HADS than a standard higher-order model.

The HADS has been in use for over 30 years and there are growing calls for it to be abandoned in favour of more contemporary tools [6]. It would appear that the HADS suffers from being saturated by a general distress factor, due in part to its narrow focus on anhedonia and autonomic arousal, resulting in problems with distinguishing between anxiety and depression. This explains the finding that the HADS anxiety subscale may act equally as well as the depression subscale as a screening tool for depressive disorder [41]. It is not clear to what extent general factor saturation is an issue with contemporary measures of anxiety and depression, such as the Patient Health Questionnaire (PHQ-9; [42]) and the Generalised Anxiety Disorder scale (GAD-7; [43]). However, it appears to be the case for

other widely used instruments. For example, a combined EFA of the Beck Depression Inventory II and Beck Anxiety Inventory also suggested the a general factor explained more than 50% of the common variance between items of both scales [18]. The presence of a strong general factor underlying the HADS is not surprising. Anxiety and depression are commonly comorbid [39,44,45] and, from a dimensional perspective, there is considerable research concerning the overlap of symptoms of anxiety and depression [39,46].

For research purposes, the use of the bifactor model in order to parse the general and specific factors is recommended. However, this is complicated and the use of a summed HADS total score appears to provide an adequate estimate of general psychological distress. In clinical practice, where it is important to distinguish between symptoms of anxiety and depression, the use of the HADS is not recommended. Due to its narrow focus the HADS has poor trait coverage and therefore brief tools with a broad coverage of symptoms of anxiety and depression, such as the PHQ-9 [42] and GAD-7 [43], may be more appropriate. Where the focus is detailed examination of the specific components of anxiety and depression, where high discriminant validity is essential, the use of longer tools that include multiple items for each symptom, such as the IDAS [47], are advisable. The use of such tools, would help to further develop our understanding of the relationship between specific components of affective disorder and outcomes, such as cardiovascular mortality [48].

This study has some notable limitations. Firstly, the response rate was 39%, which may have introduced some bias into the results – although there is no reason to suspect this to be the case. Despite the low response rate, it was possible to conduct a meta CFA on a large number of separate samples with a total size of over 20,000. To assist future meta CFAs, and meta-analyses and reanalyses in general, we implore the researchers to provide access to their raw data at the time of publication, or at the very least appropriate summary data such

as means, standard deviations and correlations. To this end, we have provided the correlation matrices from each of the studies and the pooled correlation matrices in the online supplemental material (Appendices A.1 & A.2).

A further limitation of the study is that many of the primary studies were based on the analysis of Pearson correlation matrices. Analyses based on such matrices assume that the underlying data are continuous and follow a multivariate normal distribution. However, the HADS has a four point ordinal response format and the response distributions are typically skewed. In this situation factor loadings and total variance explained will be underestimated and model fit estimates will be reduced [49]. Therefore, the factor loadings and communalities presented in this study must be considered conservative.

In conclusion, a bifactor structure provides an acceptable empirical explanation for the HADS item correlation structure and explains apparent inconsistencies in previous studies. Future research concerning the latent structure of the HADS should consider this structure, as well as accounting for item wording effects. The HADS does not provide good separation between anxiety and depression, but it is likely that this is a problem for most instruments where symptoms of anxiety and depression are jointly measured. Mounting concern over the conceptualisation and construction of the HADS may mean clinicians and researchers wanting to separately assess symptoms of anxiety and depression increasingly use other tools. However, the findings of studies using the HADS are not invalidated but must be considered in terms of general distress rather than specifically anxiety or depression.

Acknowledgments

We thank all the authors of the original papers that were able to provide information that enabled this analysis.

S. Norton was supported by an Economic and Social Research Council and Medical Research council interdisciplinary studentship and, in part, by the NIHR CLAHRC for Cambridgeshire and Peterborough. A. Sacker was supported by Grant RES-596-28-0001 from the Economic and Social Research Council.

References

1 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* 1983;67:361–70.

2 Bjelland I, Dahl AA, Haug TT, *et al.* The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *Journal of Psychosomatic Research* 2002;52:69–77.

3 Herrmann C. International experiences with the Hospital Anxiety and Depression Scale - A review of validation data and clinical results. *Journal of Psychosomatic Research* 1997;42:17–41.

4 Martin CR. What does the Hospital Anxiety and Depression Scale (HADS) Really Measure in Liaison Psychiatry Settings? *Current Psychiatry Reviews* 2005;1:69–73.

5 Cosco TD, Doyle F, Ward M, *et al.* Latent structure of the Hospital Anxiety And Depression Scale: A 10-year systematic review. *Journal of Psychosomatic Research* 2012;72:180–4.

6 Coyne JC, van Sonderen E. No further research needed: Abandoning the Hospital and Anxiety Depression Scale (HADS). *Journal of Psychosomatic Research* 2012;72:173–4.

7 Norton S, Sacker A, Done DJ. Further research needed: A comment on Coyne and van Sonderen's call to abandon the Hospital Anxiety and Depression Scale. *Journal of Psychosomatic Research* 2012;73:75–6.

8 Dunbar M, Ford G, Hunt K, *et al.* A confirmatory factor analysis of the Hospital Anxiety and Depression scale: Comparing empirically and theoretically derived structures. *British Journal of Clinical Psychology* 2000;39:79–94.

9 Watson D, Pennebaker JW. Health complaints, stress, and distress: exploring the central role of negative affectivity. *Psychological review* 1989;96:234–54.

10 Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology* 1991;100:316–36.

11 Suls J, Bunde J. Anger, anxiety, and depression as risk factors for cardiovascular disease: the problems and implications of overlapping affective dispositions. *Psychological bulletin* 2005;131:260–300.

12 Reise S, Morizot J, Hays R. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research* 2007;16:19–31.

13 Schmid J, Leiman JM. The development of hierarchical factor solutions. *Psychometrika* 1957;22:53–61.

14 Ward LC. Comparison of factor structure models for the Beck Depression Inventory--II. *Psychological assessment* 2006;18:81–8.

15 Thombs BD, Ziegelstein RC, Beck CA, *et al.* A general factor model for the Beck Depression Inventory-II: validation in a sample of patients hospitalized with acute myocardial infarction. *Journal of psychosomatic research* 2008;65:115–21.

16 Chilcot J, Norton S, Wellsted DM, *et al.* A confirmatory factor analysis of the beck depression inventory-II in end-stage renal disease patients. *Journal of Psychosomatic Research* Published Online First: March 2011. doi:10.1016/j.jpsychores.2011.02.006

17 Simms LJ, Gras DF, Watson D, *et al.* Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depression and Anxiety* 2008;25:E34–E46.

18 Steer RA, Clark LA, Beck AT, *et al.* Common and specific dimensions of self-reported anxiety and depression: the BDI-II versus the BDI-IA. *Behaviour research and therapy* 1999;37:183–90.

19 Annunziata MA, Muzzatti B, Altoè G. Defining hospital anxiety and depression scale (HADS) structure by confirmatory factor analysis: a contribution to validation for oncological settings. *Annals of oncology* 2011;22:2330–3.

20 Cosco TD, Doyle F, Watson R, *et al.* Mokken scaling analysis of the Hospital Anxiety and Depression Scale in individuals with cardiovascular disease. *General Hospital Psychiatry* 2012;34:167–72.

21 Emons WHM, Sijtsma K, Pedersen SS. Dimensionality of the Hospital Anxiety and Depression Scale (HADS) in Cardiac Patients: Comparison of Mokken Scale Analysis and Factor Analysis. *Assessment* Published Online First: 14 October 2010. doi:10.1177/1073191110384951

22 Gibbons CJ, Mills RJ, Thornton EW, *et al.* Rasch analysis of the hospital anxiety and depression scale (HADS) for use in motor neurone disease. *Health and quality of life outcomes* 2011;9:82.

23 Helvik A-S, Engedal K, Skancke RH, *et al.* A psychometric evaluation of the Hospital Anxiety and Depression Scale for the medically hospitalized elderly. *Nordic journal of psychiatry* 2011;65:338–44.

24 Lambert SÃ, Pallant JF, Girgis A. Rasch analysis of the Hospital Anxiety and Depression Scale among caregivers of cancer survivors : implications for its use in psycho-oncology. *Psycho-oncology* 2011;20:919–25.

25 Schönberger M, Ponsford J. The factor structure of the Hospital Anxiety and Depression Scale in individuals with traumatic brain injury. *Psychiatry research* 2010;179:342–9.

26 Wouters E, Booysen FLR, Ponnet K, *et al.* Wording Effects and the Factor Structure of the Hospital Anxiety & Depression Scale in HIV/AIDS Patients on Antiretroviral Treatment in South Africa. *PloS one* 2012;7:e34881.

27 Brennan C, Worrall-Davies A, McMillan D, *et al.* The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *Journal of psychosomatic research* 2010;69:371–8.

28 Moorey S, Greer S, Watson M, *et al.* The factor structure and factor stability of the hospital anxiety and depression scale in patients with cancer. *The British Journal of Psychiatry* 1991;158:255–9.

29 Friedman S, Samuelian J-C, Lancrenon S, *et al.* Three-dimensional structure of the Hospital Anxiety and Depression Scale in a large French primary care population suffering from major depression. *Psychiatry Research* 2001;104:247–57.

30 Caci H, Baylé FJ, Mattei V, *et al.* How does the Hospital and Anxiety and Depression Scale measure anxiety and depression in healthy subjects? *Psychiatry research* 2003;118:89–99.

31 Brandberg Y, Bolund C, Sigurdardottir V, *et al.* Anxiety and depressive symptoms at different stages of malignant melanoma. *Psycho-Oncology* 1992;1:71–8.

32 Razavi D, Delvaux N, Farvacques C, *et al.* Screening for adjustment disorders and major depressive disorders in cancer in-patients. *The British Journal of Psychiatry* 1990;156:79–83.

33 Chen FF, Sousa K, West SG. Testing measurement invariance of second-◎order factor models. *Structural Equation Modeling* 2009;12:471–92.

34 Chen FF, West SGS, Sousa KH. A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research* 2006;41:189–225.

35 Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6:1–55.

36 Cheung MWL, Chan W. Meta-analytic structural equation modeling: a two-stage approach. *Psychological methods* 2005;10:40–64.

37 Cheung MWL. metaSEM: Meta-analysis: A structural equation modeling approach. 2011.

38 Simms LJ, Prisciandaro JJ, Krueger RF, *et al.* The structure of depression, anxiety and somatic symptoms in primary care. *Psychological medicine* 2012;42:15–28.

39 Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. *Annual review of psychology* 1998;49:377–412.

40 Brown TA, Chorpita BF, Barlow DH. Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of abnormal psychology* 1998;107:179–92.

41 Mitchell AJ, Meader N, Symonds P. Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. *Journal of affective disorders* 2010;126:335–48.

42 Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 2001;16:606–13.

43 Spitzer RL, Kroenke K, Williams JBW, *et al.* A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 2006;166:1092–7.

44 Schoevers RA, Beekman ATF, Deeg DJH, *et al.* Comorbidity and risk-patterns of depression, generalised anxiety disorder and mixed anxiety-depression in later life: results from the AMSTEL study. *International Journal of Geriatric Psychiatry* 2003;18:994–1001.

45 Byers AL, Yaffe K, Covinsky KE, *et al.* High occurrence of mood and anxiety disorders among older adults: The National Comorbidity Survey Replication. *Archives of general psychiatry* 2010;67:489–96.

46 Watson D. Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. *Journal of abnormal psychology* 2005;114:522–36.

47 Watson D, O'Hara MW, Simms LJ, *et al.* Development and validation of the Inventory of Depression and Anxiety Symptoms (IDAS). *Psychological assessment* 2007;19:253–68.

48 Doyle F, Conroy R, McGee H. Differential predictive value of depressive versus anxiety symptoms in the prediction of 8-year mortality after acute coronary syndrome. *Psychosomatic medicine* 2012;74:711–6.

49 Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, *et al.* Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity* 2008;44:153–66.

50 Chan Y-F, Leung DYP, Fong DYT, *et al.* Psychometric evaluation of the Hospital Anxiety and Depression Scale in a large community sample of adolescents in Hong Kong. *Quality of Life Research* 2010;19:865–73.

51 Dawkins N, Cloherty ME, Gracey F, *et al.* The factor structure of the Hospital Anxiety and Depression Scale in acquired brain injury. *Brain injury* 2006;20:1235–9.

52 Forjaz MJ, Rodriguez-Blázquez C, Martinez-Martin P, *et al.* Rasch analysis of the hospital anxiety and depression scale in Parkinson's disease. *Movement Disorders* 2008;24:525–31.

53 Gale CR, Allerhand M, Sayer AA, *et al.* The structure of the Hospital Anxiety and Depression Scale in four cohorts of community-based, healthy older people: the HALCyon program. *International psychogeriatrics / IPA* 2010;22:559–71.

54 Gough K, Hudson P. Psychometric properties of the Hospital Anxiety and Depression Scale in family caregivers of palliative care patients. *Journal of pain and symptom management* 2009;37:797–806.

55 Hunt-Shanks T, Blanchard C, Reid R, *et al.* A psychometric evaluation of the Hospital Anxiety and Depression Scale in cardiac patients: addressing factor structure and gender invariance. *British journal of health psychology* 2010;15:97–114.

56 Marinus J, Leentjens AFG, Visser M, *et al.* Evaluation of the hospital anxiety and depression scale in patients with Parkinson's disease. *Clinical neuropharmacology* 2002;25:318–24.

57 Michopoulos I, Douzenis A, Kalkavoura C, *et al.* Hospital Anxiety and Depression Scale (HADS): validation in a Greek general hospital sample. *Annals of general psychiatry* 2008;7:4.

58 Olssøn I, Mykletun A, Dahl A a. The Hospital Anxiety and Depression Rating Scale: a cross-sectional study of psychometrics and case finding abilities in general practice. *BMC psychiatry* 2005;5:46.

59 Osborne RH, Elsworth GR, Sprangers MAG, *et al.* The value of the Hospital Anxiety and Depression Scale {(HADS)} for comparing women with early onset breast cancer with population-based reference women. *Quality of Life Research* 2004;13:191–206.

60 Quintana JM, Padierna a, Esteban C, *et al.* Evaluation of the psychometric characteristics of the Spanish version of the Hospital Anxiety and Depression Scale. *Acta psychiatrica Scandinavica* 2003;107:216–21.

61 Roberts SB, Bonnici DM, Mackinnon AJ, *et al.* Psychometric evaluation of the Hospital Anxiety and Depression Scale (HADS) among female cardiac patients. *British journal of health psychology* 2001;6:373–83.

62 Smith AB, Selby PJ, Velikova G, *et al.* Factor analysis of the Hospital Anxiety and Depression Scale from a large cancer population. *Psychology and psychotherapy* 2002;75:165–76.

Table 1. Studies included in the meta confirmatory factor analysis

Study	Year	Language	Population	N	Method	Structure
Annunziata et al. [19]	2011	Italian	Cancer	512	CFA	2
Caci et al. [30]	2003	French	Community	195	CFA	3
Chan et al. [50]	2010	Chinese	Community	5857	CFA	2
Cosco et al. [20]	2012	English	Cardiovascular disease	893	IRT	1
Dawkins et al. [51]	2006	English	Brain Injury	140	EFA	3
Emons et al. [21]	2010	Dutch	Cardiovascular disease	534	IRT+EFA+CFA	3
Forjaz et al . [52]	2009	Spanish	Parkinsons Disease	387	IRT	1
Gale et al. [53]	2010	English	Community	1028	EFA+CFA	2
			Community	357		
			Community	3221		
			Community	547		
Gibbons et al. [22]	2011	English	Motor Neurone Disease	298	IRT	1
Gough & Hudson [54]	2009	English	Caregivers	106	EFA	2
Helvik et al. (22)	2011	Norwegian	Elderly inpatient	484	EFA	2
Hunt-Shanks et al. [55]	2010	English	Cardiovascular	801	CFA	3
Marinus et al. [56]	2002	Dutch	Parkinsons Disease	177	EFA	2
Michopoulos et al. [57]	2008	Greek	Elderly outpatient	521	EFA	2
Olsson et al. [58]	2005	Norwegian	Inpatient	1781	EFA	2
Osborne et al. [59]	2004	English	Breast cancer	763	CFA	1
			Community	153		
Quintana et al. [60]	2003	Spanish	Community	256	EFA	2
			Eating disorder	131		
			Inflammatory Bowel Disease	154		
			Respiratory Disease	144		
Roberts et al. [61]	2001	English	Cardiovascular disease	167	EFA+CFA	2
Schönberger et al. [25]	2010	English	Brain injury	294	CFA	3
Smith et al. [62]	2002	English	Cancer	1474	EFA	2
Wouters et al. [26]	2012	Sesotho	HIV/AIDS	716	CFA	1

Note. EFA = exploratory factor analysis; CFA = confirmatory factor analysis; IRT = item response theory

Table 2. Best fitting models by goodness-of-fit criteria for the 28 samples, no wording effect

	RMSEA		SRMR		CFI		TLI		AIC		BIC	
	N	%	N	%	N	%	N	%	N	%	N	%
1. Razavi	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
2. Zigmond & Snaith	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
3. Moorey	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	2	7.1
4. Friedman	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	3.6
5. Caci	1	3.6	0	0.0	0	0.0	1	3.6	2	7.1	3	10.7
6. Brandberg	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
7. Dunbar	1	3.6	0	0.0	0	0.0	1	3.6	1	3.6	2	7.1
8. Dunbar, higher-order	1	3.6	0	0.0	0	0.0	0	0.0	0	0.0	6	21.4
9. Bifactor, 2 group-factors	24	85.7	27	96.4	25	89.3	22	78.6	23	82.1	11	39.3
10. Bifactor, 3 group factors	5	17.9	2	7.1	4	14.3	5	17.9	2	7.1	3	10.7

Note. Frequencies for RMSEA, SRMR, CFI and TLI sum to >28 due to equal fit criteria in some samples. N = number of samples, out of 28; RMSEA = root mean squared error of approximation; SRMR = root mean squared residual; CFI = comparative fit index; TLI = Tucker-Lewis index; AIC = Akaike information criterion; BIC = Bayesian information criterion

Table 3. Meta confirmatory factor analysis, model fit statistics (28 samples, N=21,820)

	χ^2	DF	RMSEA	SRMR	CFI	TLI	AIC	BIC
No item wording method factor								
1. Razavi	13506.6	77	0.089	0.059	0.863	0.838	782547	782771
2. Zigmond & Snaith	5670.9	76	0.058	0.041	0.943	0.932	774714	774945
3. Moorey	5196.3	76	0.056	0.038	0.948	0.937	774239	774471
4. Friedman	5071.6	74	0.056	0.038	0.949	0.937	774118	774366
5. Caci	4878.4	74	0.055	0.038	0.951	0.940	773925	774173
6. Brandberg	5894.2	74	0.060	0.041	0.941	0.927	774941	775189
7. Dunbar	4599.9	74	0.053	0.037	0.954	0.943	773647	773894
8. Dunbar, higher-order	4678.3	75	0.053	0.037	0.953	0.943	773723	773963
9. Bifactor, 2 group-factors	**2561.2**	**63**	**0.043**	**0.022**	**0.974**	**0.963**	**771630**	**771965**
10. Bifactor, 3 group factors	3905.3	64	0.052	0.033	0.961	0.944	772972	773300
Item wording method factor								
1. Razavi	5170.8	71	0.057	0.041	0.948	0.933	774223	774495
2. Zigmond & Snaith	3438.3	70	0.047	0.027	0.966	0.955	772493	772773
3. Moorey	3719.8	70	0.049	0.029	0.963	0.951	772774	773054
4. Friedman	3170.9	68	0.046	0.026	0.968	0.958	772230	772525
5. Caci	2914.3	68	0.044	0.025	0.971	0.961	771973	772269
6. Brandberg	3048.4	68	0.045	0.026	0.970	0.959	772107	772403
7. Dunbar	2607.1	68	0.041	0.024	0.974	0.965	771666	771961
8. Dunbar, higher-order	2609.9	69	0.041	0.024	0.974	0.966	771667	771954
9. Bifactor, 2 group-factors	**1612.7**	**57**	**0.035**	**0.017**	**0.984**	**0.975**	**770693**	**771077**
10. Bifactor, 3 group factors	2309.5	58	0.042	0.027	0.977	0.964	771388	771764

Note. DF = degrees of freedom; RMSEA = root mean squared error of approximation; SRMR = root mean squared residual; CFI = comparative fit index; TLI = Tucker-Lewis index; AIC = Akaike information criterion; BIC = Bayesian information criterion

Table 4. Factor loadings, Bifactor model with two group-factors with and without wording method factor

	No wording effect				Wording effect				
	General factor	Anxiety factor	Depression factor	R^2	General factor	Anxiety factor	Depression factor	Wording factor	R^2
A1	0.67	0.23		0.50	0.73	0.02			0.51
A3	0.60	0.43		0.43	0.67	0.26			0.40
A5	0.69	0.26		0.54	0.75	0.03			0.49
A7	0.69	-0.12		0.44	0.59	-0.05		0.34	0.41
A9	0.55	0.34		0.54	0.59	0.28			0.54
A11	0.46	0.13		0.46	0.47	0.10			0.42
A13	0.62	0.46		0.50	0.68	0.44			0.43
D2	0.47		0.46	0.27	0.41		0.44	0.27	0.30
D4	0.50		0.44	0.42	0.44		0.33	0.41	0.40
D6	0.62		0.29	0.20	0.57		0.25	0.27	0.21
D8	0.49		0.18	0.23	0.48		0.30		0.22
D10	0.37		0.26	0.54	0.36		0.34		0.49
D12	0.50		0.55	0.59	0.44		0.48	0.32	0.61
D14	0.43		0.23	0.24	0.37		0.14	0.35	0.24
Common variance explained	73.0%	11.1%	15.9%		70.2%	5.8%	13.4%	10.6%	
Total variance explained	30.1%	4.7%	6.7%		30.1%	2.5%	5.8%	4.6%	

Figure 1. Flow chart showing origin of included studies

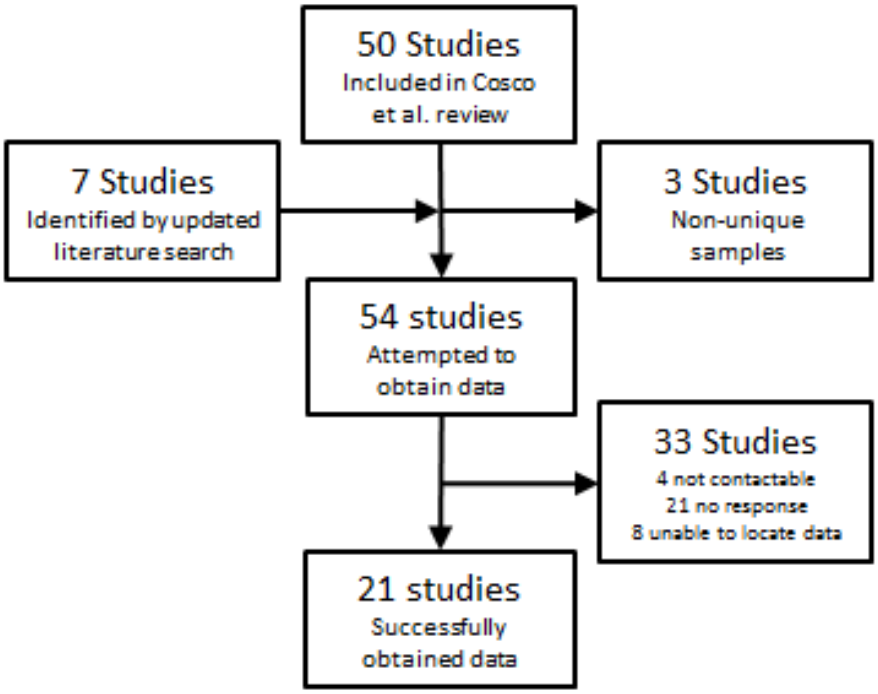


Figure 2. Factor loading patterns for each model

