

Title: Limited validity of the Hospital Anxiety and Depression Scale (HADS) in dementia: Evidence from a confirmatory factor analysis

Stott, J., Spector, A., Orrell, M., Scior, K., Sweeney, J. & Charlesworth, G. (2016). Limited validity of the Hospital Anxiety and Depression Scale (HADS) in dementia: Evidence from a confirmatory factor analysis. *International Journal of Geriatric Psychiatry*, 32 (7), pp. 805-813.

Link to repository:

<https://repository.nottinghamshirehealthcare.nhs.uk/handle/123456789/1024>

Additional information:

This is the peer reviewed version of the following article: Stott, J., Spector, A., Orrell, M., Scior, K., Sweeney, J. & Charlesworth, G. (2016). Limited validity of the Hospital Anxiety and Depression Scale (HADS) in dementia: Evidence from a confirmatory factor analysis. *International Journal of Geriatric Psychiatry*, 32 (7), pp. 805-813, which has been published in final form at <http://dx.doi.org/10.1002/gps.4530>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Publisher: Wiley

Version note:

The version presented here may differ from the published version or from the version of record. If you wish to cite the following item, it is advised to consult the publisher version. Access to the publisher version can be found via the repository URL listed above.

For more information about this article, or the research repository, please contact repository@nottshc.nhs.uk

Please cite the published version

Nottinghamshire Healthcare NHS Foundation Trust
Institutional Repository
repository.nottinghamshirehealthcare.nhs.uk

Limited validity of the Hospital Anxiety and Depression Scale (HADS) in dementia: evidence from a confirmatory factor analysis

Running Head –Limited validity of HADS in Dementia

Authors: Joshua Stott^{1*} Aimee Spector¹, Martin Orrell², Katrina Scior¹, Joanne Sweeney³ Georgina Charlesworth¹

¹ Research Department of Clinical, Educational and Health Psychology

University College London WC1E 7HB

² Institute of Mental Health University of Nottingham

³ Division of Psychology and Language Science, University College London

*Corresponding author – address contact details as above, telephone 02076795950,
email: j.stott@ucl.ac.uk

Key Words: HADS, Hospital Anxiety and Depression Scale, validity, dementia, factor analysis

Key Points

- The HADS is acceptable and feasible for use with mild to moderate dementia
- Structural validity of HADS in dementia is unclear, making interpretation difficult.
- It is, however, preferable to use the HADS to measure two factors of anxiety and depression rather than one single distress factor.

Funding:

This analysis was completed as part of a fellowship awarded to Joshua Stott by the Alzheimer's society. Grant number 236 (AS-CTF-14-005).

This was a secondary data analysis of data obtained from the Can't Forget to Worry trial and the Support at Home – Interventions to Enhance Life in Dementia (SHIELD) trial.

Funding for the Can't forget to worry trial is detailed in the following statement: This article presents independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (grant reference number PB-PG-0609-18230). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Funding for the Shield trial is detailed in the following statement: This article presents independent research funded by the National Institute for Health Research [Programme Grant Reference RP-PG-0606-1083. Support at Home – Interventions to Enhance Life in Dementia (SHIELD)]. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Word count: 3495

Abstract

Objectives

The Hospital Anxiety and Depression Scale (HADS) is a well-validated, self-report measure of both anxiety and depression. It is frequently used with people with dementia. However, its structural validity has never been examined in this population. The current study used confirmatory factor analysis (CFA) to assess this.

Methods

Baseline data from two intervention studies for people with mild to moderate dementia were combined (N = 268). CFA was used to test whether a one, two or three factor structure best fit the data. Indices of model misspecification were examined to test for poor quality items, and models re-specified accordingly. Finally, measurement invariance across gender and different levels of cognitive impairment was assessed.

Results

A one-factor structure did not fit the data. Two and three factor structures fitted the data equally well. Model fit was improved by removal of two items. Measurement invariance was adequate across gender, but poor across groups with differing levels of cognitive impairment.

Conclusion

The HADS is acceptable and feasible but difficult to interpret in a dementia population. We suggest that it should be interpreted as measuring two separate factors of anxiety and depression and not one 'distress' factor. However, two items may need to be removed, affecting cut-off scores. Poor measurement invariance means the HADS may not be a good tool for measuring differences in anxiety and depression between those with mild and those with moderate cognitive impairment.

Introduction

Depression and anxiety are common yet under-diagnosed comorbidities in dementia (Enache et al., 2011, Wolitzky-Taylor et al., 2010) and are associated with negative outcomes (Gibbons et al., 2002). Self-rating of mood in those with cognitive impairment is complex (Feher et al., 1992). However, dementia diagnosis is now typically made at an earlier stage in disease progression with associated increased self-awareness (Grimmer et al., 2015). Self-report measures are therefore increasingly relevant for people with dementia and are particularly useful for measuring mood in people with dementia who have no available informant (Alzheimer's Association, 2012).

The Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983) is a 14-item self-report measure of anxiety and depression. It is appealing for use in dementia as it is relatively brief, measures both anxiety and depression and can be used in those with comorbid physical health problems. Evaluation of measurement properties of the HADS in dementia is important, given its use in clinical practice and dementia research (e.g. Clare et al. (2012))

The utility of any measure stands or falls on its reliability and validity, both of which are multi-faceted constructs. The HADS performs well on some aspects of validity and reliability across different populations, for example, those with physical health problems or psychiatric inpatients (Bjelland et al., 2002).

Structural validity, the degree to which item scores are an adequate reflection of dimensional structure is an important aspect of validity without which measurements cannot be adequately interpreted (Mokkink et al., 2010). Evidence for structural validity of the HADS in populations without dementia is mixed, with studies

suggesting that the HADS measures one single ‘distress’ factor (Razavi et al., 1990), separate ‘anxiety’ and ‘depression’ factors (Zigmond and Snaith, 1983) or even three factors, following Clark and Watson (1991)’s tripartite model of anxiety, depression and negative affectivity (Cosco et al., 2012).

While the structural validity of the HADS has not been specifically examined in a dementia population, work in a medically-hospitalised older sample (Helvik et al., 2011) and a cognitively-intact nursing home sample (Haugan and Drageset, 2014) favours a two-factor structure. Such findings are not generalizable to a dementia population (Cosco et al., 2012), as anxiety and depression present differently (Banerjee et al., 2011) and some items (e.g. the fourth item on the depression subscale (I feel slowed down)) may be confounded by cognitive functioning (Haugan and Drageset, 2014). The current study seeks to inform the use of and interpretability of the HADS in dementia through assessment of its structural validity in a dementia sample.

Method

Participants

Data analysed in this study were the combined baseline data for participants with mild to moderate dementia (diagnosed according to DSM-IV) taken from two clinical trials, one examining home based support for people with dementia (Orrell et al., In press), and the other CBT for anxiety in dementia (Spector et al., 2015). The recruitment procedures and samples for these trials have been described in detail elsewhere (Orrell et al., In press, Spector et al., 2015). All participants in both trials gave written informed consent. Ethical approval was obtained from ‘East London 3 Research Ethics Committee’ (reference number 10/H0701/124) for use of the Spector

et al. (2015) data and from the Outer North East London Research Ethics Committee (reference number: 09/H0701/54) for use of the Orrell et al. (In press) data.

Age, gender, Mini Mental State (MMSE)(Folstein et al., 1975) scores and dementia diagnosis (only recorded in the Orrell et al. (In press) study) of the combined sample are presented in Table 1.

The HADS

The HADS comprises 14 items each rated from 0-3, with higher scores indicating greater anxiety/depression. The anxiety and depression subscales each have seven items and a maximum score of 21. (Zigmond and Snaith, 1983)

Statistical Analyses

Confirmatory factor analysis (CFA) was used to test the fit of the HADS data from the combined dataset with the three most commonly proposed factor structures: the two-factor model of Zigmond and Snaith (1983), the one-factor model of Razavi et al. (1990), and the three-factor non-hierarchical model of Dunbar et al. (2000). Diagrams illustrating these models are shown in Figure 1.

- Insert Figure 1 about here-

CFA was performed in R(version 3.2.2) (R Core Team, 2013) statistical software using Lavaan (Rosseel, 2012) semplots and semtools packages (SemTools Contributors, 2015).

Indices of model fit

In line with the literature (Hu et al., 1992, Dunbar et al., 2000), model fit was assessed by several indices with cut-off scores used to determine good, adequate or poor fit.

Two of the indices used here - the Standardized Root Mean Square Residual (SRMR) and the Root Mean Square Error of Approximation (RMSEA) - have cut-off scores of good fit, <0.05; adequate fit, <0.08; and poor fit > 0.08. Two of the others - the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) - have cut-off scores of good fit, >0.95; adequate fit, >0.9 and poor fit, <0.9. The final index, the Bayesian Information Criterion (BIC), is a comparative fit index, with no cut-off.

Specification searching

Fit indices should be supplemented with information about how well individual items fit within a CFA model (Byrne, 2013). Here, items with standardised residuals with values in excess of 2.58 and high modification indices were classified as misspecified (Byrne, 2013). Both standardised residuals and modification indices were used to adapt and improve, or 're-specify', models through specification searching (Byrne, 2013). To avoid undue influence being given to the idiosyncracies of a particular data set, *a priori* concerns were used to drive specification searching (Byrne, 2013).

Measurement invariance

Measurement invariance of the HADS is necessary if it is to be used to test for differences in anxiety and depression across particular subgroups of people with dementia. Measurement invariance is assumed if individuals in different groups with the same levels of the latent construct have the same expected raw-score on the measure (Drasgow and Kanfer, 1985). To test for measurement invariance of the HADS, the data were split into subgroups according to gender and cognitive impairment. In line with evidence relating MMSE to stage of dementia (Pernecky et al., 2006), mild impairment was defined as $MMSE \geq 21$ and compared to a moderate impairment subgroup ($MMSE \leq 20$). Following this the models were examined for the different types of measurement invariance (configural, metric, strong and strict)

through comparison of progressively more constrained models, with a change in CFI greater than 0.01 taken to indicate change in model fit across constraints and therefore lack of invariance between groups (Chen, 2007).

Results

Data characteristics and initial analyses

Of the combined dataset (N = 339), 65 participants did not attempt the HADS. Of those who attempted the HADS, six were ‘non completers’ (missing data for one or more items).

The data were examined for differences in gender, age and MMSE scores between those who completed, attempted and did not attempt the HADS. Chi square was used to test for differences in gender and ANOVA for differences in MMSE and age.

Groups did not differ for gender or age but did differ in MMSE, ($F_{2,304} = 25.97$ $p < 0.001$) with planned comparisons revealing that non attempters had lower MMSE scores than completers (Games Howell MD = -7.25, $p < 0.001$).

For those who attempted the HADS (N= 274), Little’s MCAR test revealed that data were missing completely at random ($\chi^2 = 50.48$ (36), $p = 0.06$). Under these circumstances, listwise deletion of cases with missing data is acceptable (Graham, 2009). Consequently only those with full data available (N=268) were subject to CFA.

The final CFA sample consisted of 125 males (47%), 142 (52.9%) females and one unstated, mean age was 69 years (Standard Deviation of 12.3) and mean MMSE score was 19.8 (standard deviation, 5.4). Descriptive data for the HADS items are shown in Table 1. Graphical inspection and significant Shapiro Wilk tests for all HADS items

indicated significant univariate non-normality, with the sample generally reporting low levels of depression and anxiety and consequent positive skew. Mardia's test indicated significant multivariate non normality (χ^2 skew = 1784.7, $p < 0.001$, Z Kurtosis = 26.2, $p < 0.001$). Given this non-normal ordinal data, the CFA approach of Maximum Likelihood (ML) methods with Satorra Bentler corrected (robust) chi square was used to examine fit of all models (Finney and DiStefano, 2006).

- Insert Table 1 about here -

Confirmatory Factor Analysis results

Parameter estimates

Parameter estimates were deemed adequate using Byrne (2013)'s three criteria: consistency with underlying theory, values falling inside admissible ranges, and parameters being statistically significant.

Initial model fit

Table 2 shows the fit indices for all three models. None of the models show a good fit with the data, with the one factor model performing particularly poorly across all indices. The two and three factor models both performed similarly and adequately on three indices (SRMR, CFI, RMSEA) but were poor on one (TLI).

Specification searching

Given the mixed evidence as to the adequacy of fit of the two and three factor models but lack of difference in fit between them, specification searching (examination of modification indices and standardized residuals) of the two and three factor models was conducted to understand sources of model misspecification. This was first done in relation to item four on the depression subscale 'I feel slowed down..', which had been specified *a priori*, as potentially problematic. The highest modification indices

in the two factor (31.7) and three factor models (31.9) were associated with cross-loadings of this item onto latent variables other than depression (anxiety and negative affectivity). Additionally, in both models, this item was associated with the largest standardized residual covariance values (5.16, in the three factor model and 5.01 in the two factor model) and also the highest number of these in excess of 2.58 (five in both the two and three factor models). Given clear evidence that this item was a source of misspecification, it was removed from the analyses, which were then re-run with results and fit indices detailed in Table 2. The fit of both two and three-factor models was improved, such that indices of fit were now ‘adequate’ for some fit indices (TLI, CFI) and ‘good’ for others (RMSEA, SRMR). There was, however, still no discrimination between the models, with both models having almost identical fit indices.

- Insert Table 2 about here -

On inspection, the fourth item on the anxiety subscale (‘I can sit at ease and feel relaxed’) was associated with the next highest modification indices (22.1 in the 3 factor model, 23.1 in the 2 factor model) and next highest number of standardized residuals above 2.58 (three) in both the two and three factor models. While there was not an *a priori* reason for removing this item, given some evidence of misspecification, it was removed (along with the fourth item on the depression subscale) for an exploratory analysis (with results and fit indices detailed in Table 2). Removing item four on the anxiety scale improved model fit in both two and three factor models with all indices now suggestive of ‘good fit’. However, once more there was no difference between models.

Measurement invariance

In order to assess whether the HADS can be validly used to measure differences in depression and anxiety across groups who differ in cognitive functioning or gender, we assessed measurement invariance of the HADS across these groups. Measurement invariance assessment was conducted on both the two and three factor models with the fourth item of the depression subscale removed (the models with the fourth item on the anxiety subscale also removed were not subjected to this analysis as the removal of this item was exploratory). The data were first divided into subgroups according to gender (male N = 142) and separately by MMSE score. For five participants, MMSEs were missing so the sample size for this analysis was 263, with N = 142 falling into the low MMSE group and N=121 into the high MMSE group.

The results of the analysis of the different types of invariance (configural, metric, strong and strict) are shown in Table 3. Measurement invariance of the HADS was adequate across groups who differ in gender but inadequate across MMSE categories. Specifically, the data indicate that for groups differing according to gender, configural invariance criteria were adequate for CFI (three-factor 0.91, two-factor 0.91) and RMSEA (three-factor 0.071, two-factor 0.069) and criteria for all other invariance types were met with CFI change always more than 0.01. For groups differing in cognition, the high cognition group for the three-factor model had a non-positive definite covariance matrix, meaning that it was difficult to interpret invariance for this model, and measurement invariance could not be assumed. For the two factor model, the configural invariance assumption was not met (CFI of 0.88 and a RMSEA of 0.081) although measurement invariance was demonstrated across all other levels ($CFI\Delta < 0.01$) aside from strict invariance where $CFI\Delta = 0.013$.

- Insert Table 3 about here -

Discussion

This study suggests that people with mild to moderate dementia can complete the HADS, but raises concerns about the structural validity and consequent interpretation of HADS scores in dementia.

A single distress factor?

In line with previous CFA studies in other populations (Haugan and Drageset, 2014, Cosco et al., 2012) we found no evidence that the HADS measures a single distress factor. We suggest that it should not be interpreted in this way in people with dementia in clinical or research contexts.

Structural ambiguity and pragmatic use of a two factor model

We could not distinguish between two interpretations of the HADS; that it measures two factors of ‘anxiety’ and ‘depression’ or that it measures three factors of anxiety, depression and negative affectivity. The inability to distinguish between different interpretations has been termed ‘structural ambiguity’ (Wang et al., 2006) and makes understanding HADS scores in dementia difficult.

Structural ambiguity has been found with the HADS in other populations (Wang et al., 2006) and is in line with the general lack of clarity over HADS structure (Cosco et al., 2012). This has led some authors to advocate abandoning it altogether (Coyne and van Sonderen, 2012). One strategy for deciding between structural models is to favour the most parsimonious structure. However, fit indices used in the current study (e.g. BIC) take model parsimony into account (Neath and Cavanaugh, 2012) and did not indicate that the two factor structure should be preferred. A strategy to disambiguate in future research would be to test the indices derived from the two and three factor structures for other forms of validity (e.g. concurrent or criterion validity) in a

dementia sample and to see which performs best. Until this research has been done, we suggest that, if the HADS is to be used in dementia, a two-factor interpretation might be preferred due to its greater simplicity of scoring. Given this and that fit indices are so similar for two and three factor models, the rest of this discussion will focus on the two-factor interpretation.

Removal of items

Model fit was improved by removal of two items (the fourth items on the anxiety and depression subscales). The prediction that the fourth item on the depression subscale (I feel slowed down...) would be confounded by cognitive impairment was supported by the poor fit of this item coupled with its relatively high mean score, which could reflect individuals endorsing it due to cognitive impairment regardless of depression. It was more surprising that the fourth item on the anxiety scale (I can sit at ease and feel relaxed) did not relate to the underlying construct of anxiety. This may be a data idiosyncrasy but has been found before (Haugan and Drageset, 2014) and warrants further consideration.

Given their poor fit to the data, we suggest that HADS users should definitely remove the fourth item on the depression scale and consider removing the fourth item on the anxiety subscale in scoring the HADS for people with dementia. The removal of one or both of these items will affect the ability to use HADS cut-offs for anxiety and depression caseness, so we suggest that future work with the HADS in dementia could also focus on developing cut-offs for shortened HADS subscales excluding these items.

Measurement invariance

The measurement invariance data suggest that, in a dementia population, differences in mean HADS scores between moderate and milder impairment groups may be uninterpretable. This is because such differences may be due either to between group variation in the relationship of raw HADS scores to the latent constructs of anxiety and depression or to between group differences in anxiety and depression themselves (Xu, 2012). The implication of this is that research using the HADS to examine differences in anxiety and depression between mild and moderate impairment groups will be hard to interpret. Similarly, in clinical work, where normative reference groups differ in the degree of cognitive impairment to a patient with dementia, HADS scores will be difficult to meaningfully understand. As measurement invariance is better across gender, comparisons in HADS scores between men and women with dementia can be performed.

Strengths and limitations

This is the first study to examine the structural validity of the HADS in dementia. A strength is the use of a CFA approach to test hypotheses as to which structure proposed in the literature best fits in a dementia sample. Some limitations require noting. The mean score on HADS items was low. Future work should examine this measure in samples where there is more variance and higher levels of depression and anxiety. Dementia is an umbrella term and factor structure may differ across specific dementia diagnoses, which have different patterns of impairment. For example, those with behavioural variant Frontotemporal Dementia may lack insight (Rosen et al., 2014) and under-report anxiety or depression. We did not have the data to examine this, but future research is recommended. A number of individuals in the Orrell et al. (in press) study did not attempt the HADS, and were excluded from the analysis. Consequently, these results are only representative of those people with dementia who

attempt the HADS not the dementia population as a whole. The higher MMSE scores of attempters compared to non-attempters suggests that this population may have higher cognitive functioning, although there were attempters with very low MMSE scores, indicating low MMSE scores should not be used to rule out use of the HADS.

Our relatively small sample size (for a CFA study) may result in the structural ambiguity found here (Wang et al., 2006). Replication with a larger sample is recommended. Finally, although the most frequently proposed structures in the literature were evaluated, not all potential HADS structures were considered. Future research should examine the bi-factor structure (Norton et al., 2013) and the impact of measurement artefacts (Straat et al., 2013). Item Response Theory studies may be useful to conduct in a dementia population as these provide strong evidence of latent variable structure, with the particular advantage of being generalizable beyond a population (Cosco et al., 2012).

Conclusions

This study suggests that the HADS is feasible for use in dementia, but is difficult to interpret. The HADS should not be used to measure one factor of ‘distress’ in this population. While two and three factor structures are equally supported here, we suggest that the HADS is used to measure two factors of anxiety and depression for simplicity of scoring. Two HADS items may not be useful in a dementia population and further work is needed to develop cut-off scores for a reduced item version. Lack of measurement invariance means that the HADS may not be suitable to measure differences in anxiety and depression where groups differ in level of cognitive impairment.

References

- Alzheimer's Association 2012. 2012 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 8, 131-168.
- Banerjee, S., Hellier, J., Dewey, M., Romeo, R., Ballard, C., Baldwin, R., Bentham, P., Fox, C., Holmes, C. & Katona, C. 2011. Sertraline or mirtazapine for depression in dementia (HTA-SADD): a randomised, multicentre, double-blind, placebo-controlled trial. *The Lancet*, 378, 403-411.
- Bjelland, I., Dahl, A. A., Haug, T. T. & Neckelmann, D. 2002. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *Journal of psychosomatic research*, 52, 69-77.
- Byrne, B. M. 2013. *Structural equation modeling with AMOS: Basic concepts, applications, and programming*, Routledge.
- Chen, F. F. 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, 14, 464-504.
- Clare, L., Nelis, S. M., Martyr, A., Roberts, J., Whitaker, C. J., Markova, I. S., Roth, I., Woods, R. T. & Morris, R. G. 2012. The influence of psychological, social and contextual factors on the expression and measurement of awareness in early-stage dementia: testing a biopsychosocial model. *International journal of geriatric psychiatry*, 27, 167-177.
- Clark, L. A. & Watson, D. 1991. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of abnormal psychology*, 100, 316.
- Cosco, T. D., Doyle, F., Ward, M. & McGee, H. 2012. Latent structure of the Hospital Anxiety And Depression Scale: a 10-year systematic review. *Journal of psychosomatic research*, 72, 180-184.
- Coyne, J. C. & van Sonderen, E. 2012. The Hospital Anxiety and Depression Scale (HADS) is dead, but like Elvis, there will still be citings. *Journal of Psychosomatic Research*, 73, 77-78.
- Drasgow, F. & Kanfer, R. 1985. Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70, 662.
- Dunbar, M., Ford, G., Hunt, K. & Der, G. 2000. A confirmatory factor analysis of the Hospital Anxiety and Depression scale: comparing empirically and theoretically derived structures. *British Journal of Clinical Psychology*, 39, 79-94.
- Enache, D., Winblad, B. & Aarsland, D. 2011. Depression in dementia: epidemiology, mechanisms, and treatment. *Current opinion in psychiatry*, 24, 461-472.
- Feher, E. P., Larrabee, G. J. & Crook, T. H. 1992. Factors attenuating the validity of the Geriatric Depression Scale in a dementia population. *Journal of the American Geriatrics Society*, 40, 906-909.
- Finney, S. J. & DiStefano, C. 2006. Non-normal and categorical data in structural equation modeling. *Structural equation modeling: A second course*, 269-314.
- Folstein, M. F., Folstein, S. E. & McHugh, P. R. 1975. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12, 189-198.
- Gibbons, L. E., Teri, L., Logsdon, R., McCurry, S. M., Kukull, W., Bowen, J., McCormick, W. & Larson, E. 2002. Anxiety symptoms as predictors of nursing home placement in patients with Alzheimer's disease. *Journal of Clinical Geropsychology*, 8, 335-342.
- Graham, J. W. 2009. Missing data analysis: making it work in the real world. *Annu Rev Psychol*, 60, 549-76.

- Grimmer, T., Beringer, S., Kehl, V., Alexopoulos, P., Busche, A., Förstl, H., Goldhardt, O., Natale, B., Ortner, M. & Peters, H. 2015. Trends of patient referral to a memory clinic and towards earlier diagnosis from 1985–2009. *International Psychogeriatrics*, 27, 1939-1944.
- Haugan, G. & Drageset, J. 2014. The Hospital Anxiety and Depression Scale— Dimensionality, reliability and construct validity among cognitively intact nursing home patients. *Journal of affective disorders*, 165, 8-15.
- Helvik, A.-S., Engedal, K., Skancke, R. H. & Selbæk, G. 2011. A psychometric evaluation of the Hospital Anxiety and Depression Scale for the medically hospitalized elderly. *Nordic journal of psychiatry*, 65, 338-344.
- Hu, L.-t., Bentler, P. M. & Kano, Y. 1992. Can test statistics in covariance structure analysis be trusted? *Psychological bulletin*, 112, 351.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M. & de Vet, H. C. 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*, 63, 737-745.
- Neath, A. A. & Cavanaugh, J. E. 2012. The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4, 199-203.
- Norton, S., Cosco, T., Doyle, F., Done, J. & Sacker, A. 2013. The Hospital Anxiety and Depression Scale: a meta confirmatory factor analysis. *J Psychosom Res*, 74, 74-81.
- Orrell, M., Hoe, J., Charlesworth, G., Russell, I., Challis, D., Moniz-Cook, E., Knapp, M., Woods, B., Hoare, Z., Aguirre, E., Toot, S., Streater, A., Crellin, N., Whitaker, C. J. & D'Amico, F. In press. The support at Home - Interventions to Enhance Life in Dementia (SHIELD). Final Grant Report for the NIHR
- Pernecky, R., Wagenpfeil, S., Komossa, K., Grimmer, T., Diehl, J. & Kurz, A. 2006. Mapping scores onto stages: mini-mental state examination and clinical dementia rating. *The American Journal of Geriatric Psychiatry*, 14, 139-144.
- R Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Razavi, D., Delvaux, N., Farvacques, C. & Robaye, E. 1990. Screening for adjustment disorders and major depressive disorders in cancer in-patients. *Br J Psychiatry*, 156, 79-83.
- Rosen, H. J., Alcantar, O., Zakrzewski, J., Shimamura, A. P., Neuhaus, J. & Miller, B. L. 2014. Metacognition in the behavioral variant of frontotemporal dementia and Alzheimer's disease. *Neuropsychology*, 28, 436.
- Rosseel, Y. 2012. Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1-36.
- SemTools Contributors 2015. semTools: Useful tools for structural equation modeling. R package version 0.4-9. . Retrieved from <http://cran.r-project.org/package=semTools>.
- Spector, A., Charlesworth, G., King, M., Lattimer, M., Sadek, S., Marston, L., Rehill, A., Hoe, J., Qazi, A. & Knapp, M. 2015. Cognitive-behavioural therapy for anxiety in dementia: pilot randomised controlled trial. *The British Journal of Psychiatry*, 206, 1-8.
- Straat, J. H., van der Ark, L. A. & Sijtsma, K. 2013. Methodological artifacts in dimensionality assessment of the hospital anxiety and depression scale (HADS). *J Psychosom Res*, 74, 116-21.

- Wang, W., Lopez, V. & Martin, C. R. 2006. Structural ambiguity of the Chinese version of the Hospital Anxiety and Depression Scale in patients with coronary heart disease. *Health Qual Life Outcomes*, 4, 6.
- Wolitzky-Taylor, K. B., Castriotta, N., Lenze, E. J., Stanley, M. A. & Craske, M. G. 2010. Anxiety disorders in older adults: a comprehensive review. *Depression and anxiety*, 27, 190-211.
- Xu, K. 2012. Multiple group measurement: invariance analysis in Lavaan. *Cambridge Mass*, 37.
- Zigmond, A. S. & Snaith, R. P. 1983. The hospital anxiety and depression scale. *Acta Psychiatr Scand*, 67, 361-70.

Conflict of interest

None declared

Table 1 showing sample characteristics and descriptive statistics of HADS items

	N(%)	Mean	SD	Skew	Kurtosis	Shapiro Wilk
SAMPLE CHARACTERISTICS						
Male	125(48)					
Age		69.1	12.3			
MMSE		19.8	5.4			
Dementia Diagnosis						
Not recorded (all participants in the Spector et al. (2015) trial and those with missing data in the Orrell et al. (In press) trial.)	70 (26)					
Alzheimer's disease	109 (40)					
Vascular dementia	27(10)					
Frontotemporal dementia	3(1.1)					
Lewy Body Dementia	3(1.1)					
Any other type of dementia	23(8.6)					
HADS ITEMS						
1. I feel tense or 'wound up'		.74	0.78	1.07	1.029	0.769
2. I still enjoy the things I used to enjoy		.81	0.93	0.95	-.011	0.787
3. I get a sort of frightened feeling		.72	0.92	0.98	-.174	0.754
4. I can laugh and see the funny side of things		.51	.781	1.49	1.520	0.675
5. Worrying thoughts go through my mind		.75	.893	1.06	.308	0.769
6. I feel cheerful		.56	.750	1.35	1.569	0.715
7. I can sit at ease and feel relaxed		.78	.799	0.87	.320	0.796
8. I feel as if I have slowed down		1.19	.905	0.69	-.175	0.816
9. I get a sort of frightened feeling		.54	.751	1.40	1.656	0.707
10. I have lost interest in my appearance		.58	.829	1.30	.706	0.706
11. I feel restless as if I have to be on the move		.87	.888	0.74	-.287	0.816
12. I look forward with enjoyment to things		.68	.961	1.22	.296	0.709
13. I get sudden feelings of panic		.68	.793	1.04	.575	0.768
14. I can enjoy a good book or radio or TV program		.53	.901	1.70	1.825	0.625

Note N was 268 for all items, as were degrees of freedom for Shapiro Wilk test. All Shapiro Wilk statistics were significant $p < 0.001$.

Table 2. Fit indices of original and respecified versions of one factor, two factor and three factor models

Model	χ^2 (robust)	Df	Srmr	Cfi	TLI	Rmse (90% CI)	BIC
One factor	203.70	77	0.076	0.799	0.763	0.078 (0.067-0.089)	8831.78
Two factor original	135.90	76	0.066	0.905	0.886	0.054 (0.04-0.067)	8738.06
Two factor without HADD4	99.20	64	0.055	0.939	0.926	0.045 (0.030-0.060)	8047.85
Two factor without HADD4 or HADA4	70.10	53	0.042	0.967	0.959	0.035	7452.36
Three factor original	132.40	74	0.065	0.907	0.886	0.054 (0.041-0.067)	8746.52
Three factor without HADD4	96.40	62	0.054	0.941	0.925	0.045 (0.030-0.060)	8056.73
Three factor without HADD4 or HADA4	67.69	51	0.043	0.968	0.959	0.035 (0.009-0.053)	7460.95

Note: *SRMR* _ standardized root mean residual (<0.05 suggests good fit, <0.08 suggests adequate fit, >0.08 suggests poor fit); *CFI* _ comparative fit index (>0.95 suggests good fit, >0.9 suggests adequate fit, <0.9 suggests poor fit); *TLI* _tucker Lewis Index (> 0.95 indicates good fit, >0.9 suggests adequate fit, <0.9 suggests poor fit); *RMSEA* _ root mean square error of approximation (< 0.05 is good fit, <0.08 is adequate fit, >0.08 is poor fit) *CI*_confidence interval; *BIC*_ Bayesian information criterion; *HADD4*, Fourth item on the depression subscale; *HADA4*, fourth item on the anxiety subscale.

j

Table 3. Series of model comparisons to test measurement invariance of two and three factor models

Subgroup	Model	Invariance type	$\chi^2 (\Delta\chi^2)$	DF (Δ DF)	Δp	CFI(Δ CFI)	RMSEA
Gender	Two factor model without HADD4	Configural	210.77	128	N/A	0.910	0.069
		Metric	(13.25)	11	(0.277)	(0.002)	NA
		Strong	(17.71)	11	(0.088)	(<0.001)	NA
		Strict	(7.83)	2	(0.020)	(0.002)	NA
	Three factor model without HADD4	Configural	207.05	124	N/A	0.909	0.071
		Metric	(12.69)	10	(0.241)	(0.003)	NA
		Strong	(17.81)	10	(0.058)	(0.009)	NA
		Strict	(8.00)	3	(0.046)	(0.005)	NA
MMSE	Two factor model without HADD4	Configural	238.63	128	N/A	0.884	0.081
		Metric	(8.88)	11	(0.632)	(0.002)	NA
		Strong	(25.07)	11	(0.009)	(0.015)	NA
		Strict	(36.67)	2	(0.102)	0.003	NA
	Three factor model without HADD4	Configural	229.80	124	N/A	0.889	0.081
		Metric	(9.346)	10	(0.499)	(0.001)	NA
		Strong	(22.026)	10	(0.015)*	(0.013)*	NA
		Strict	(7.843)	3	(0.049)	(0.005)	NA

Note: For configural invariance fit indices CFI _ comparative fit index (>0.95 suggests good fit, >0.9 suggests adequate fit, <0.9 suggests poor fit), RMSEA _ root mean square error of approximation (< 0.05 is good fit, <0.08 is adequate fit, >0.08 is poor fit). For all other invariance types, Δ CFI < 0.01 implies that the invariance assumption still holds.

* indicates that invariance assumption is not met according to these criteria.