Word count 3498 (does not include footnote about copyright or instructions to insert tables)

The validity and usefulness of the Hospital Anxiety and Depression Scale in carers of people with dementia: evidence from confirmatory factor analysis, concurrent validity and measurement invariance in a large sample.

Authors: Joshua Stott, DClinPsy[1]* Martin Orrell, PhD[2], Georgina Charlesworth, PhD[1]

[1] Research Department of Clinical, Educational and Health Psychology

University College London WC1E 7HB

[2] Institute of Mental Health University of Nottingham

*Corresponding author – address contact details as above, telephone 02076795950, email:

j.stott@ucl.ac.uk

ABSTRACT
Objectives

The Hospital Anxiety and Depression Scale (HADS) is a self-report measure of anxiety and depression. It is recommended for clinical assessment and has been used as the primary outcome in large clinical trials with carers of people with dementia. Its validity and utility have never been examined in this population. The current study addresses this.

Design

Secondary data analysis of baseline data from a recent intervention trial (N = 284) with cross-validation in baseline data from a second trial (N=230).

Methods

We used confirmatory Factor Analysis to test whether a one, two or three factor structure best fit the data and used indices of model misspecification to re-specify. We assessed internal consistency, concurrent validity of obtained factors and measurement invariance across gender, age, kinship and cohabitation status.

Results

A three-factor structure best fit the data. Removal of one item improved model fit. The factors showed good internal consistency and high levels of concurrent validity. Measurement invariance was adequate across gender and kinship, but not age or cohabitation status. Results were replicated in the cross-validation sample, enhancing reliability.

Conclusions

In this group, the HADS measures three factors; depression, anxiety and negative affectivity. The depression scale can be used as originally intended, supporting results of large clinical trials. The HADS does not validly measure distress or anxiety. Consequently, clinical practice recommendations could be revisited and future research trials should not use HADS

anxiety or distress as outcomes. Researchers should pay attention to measurement invariance

when using HADS to compare carer subgroups

OBJECTIVES

Around 40% of carers of people with dementia have clinically significant anxiety, depression or other psychological symptoms (1). Such psychological morbidity is predictive of elder abuse(2) and breakdown in care with associated economic implications(3). Since there are clinically(4) and cost effective(5) interventions, it is important to be able to assess and detect anxiety and depression in this group.

The Hospital Anxiety and Depression Scale (HADS) (6) is a 14-item self-report measure that is brief, measures both anxiety and depression and can be used in those with comorbid physical health problems (6). The HADS is recommended in European consensus guidelines for carers of people with dementia (7) and has been used as the primary outcome measure in large clinical trials in this group (4, 8).

Despite widespread usage, there are concerns about the validity of the HADS. In particular, there is no consensus as to the underlying factors it measures (9). This is critical as without clear understanding of this 'structural validity', measurements cannot be adequately interpreted (10).

To our knowledge, no study has examined the factor structure of the HADS in a sample comprised solely of carers of people with dementia. Lack of consensus as to factor structure in the carer literature, with some trials interpreting the HADS as measuring a single factor of distress (4) and others (11) as well as clinical recommendations(7) , interpreting it as a two factor measure of anxiety and depression. Findings from the non-carer literature cannot be directly applied. Multiple studies have suggested that the HADS can differ in factor structure across even populations that appear superficially similar (e.g. patients with different health conditions) (9) measuring one factor (12) in one population and two (6) or even three factors, in others (9, 13). It is also possible that the validity and utility of HADS items may vary in dementia carers compared to other groups, by virtue of their experience (e.g. grief) (14) and demographic characteristics (e.g. a high proportion of older adults) (15). In the light of such

issues it is recommended (9) that HADS structure be evaluated in each population with which it is used, with reliability of conclusions enhanced through cross-validation in a separate sample (9) and validity confirmed through as expected correlations with related measures(16).

Importantly, examination of HADS structure in carers specifically also allows assessment of measurement invariance to determine whether the HADS can be accurately used to examine differences between subgroups of carers (17). This is of considerable import because carers are a heterogeneous group and researchers are often interested in whether particular characteristics of carers (e.g. gender, kinship, cohabitation with person with dementia, or age) impact on anxiety or depression (18, 19).

Hence, in the current study, we examine the usefulness of the HADS in a sample of carers of people with dementia, evaluating its factor structure to determine interpretation as an outcome measure or clinical tool, concurrent validity to reinforce conclusions and measurement invariance to inform comparisons across subgroups of carers.

METHODS
Design

This is a cross-sectional, secondary data analysis of baseline data from the 'Carer Support Programme/Remembering Yesterday Caring for Today (CSP-RYCT)' trial (11) with cross-validation in a sample comprised of baseline data from the 'Befriending and Cost of Caring (BECCA)' trial (8).

Participants

291 participants took part in the CSP RYCT trial (11) and 236 in the BECCA trial (8) All participants were carers to at least one person diagnosed with dementia of varying subtypes (Alzheimer's disease, vascular dementia and others) according to DSM-IV criteria. The recruitment procedures and sample for the trials have been described elsewhere (8, 11). Since seven individuals from the CSP RYCT trial and six from the BECCA trial did not complete

the HADS and this data was missing completely at random (Little's MCAR p<0.05), their data were removed from analyses(20). Thus 284 participants were included in the initial analysis and 230 in the cross-validation analysis. Demographic and clinical characteristics for both samples are presented in Table 1. Ethical approval for use of CSP-RYCT data was given by the Outer North East London NHS Research Ethics committee (reference number: 09/H0701/54) and for the BECCA data by the Eastern Multi Regional Ethics Committee (01/5/48). All participants gave written informed consent.

Instruments

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, a maximum score of 21 and cut offs of 11 for caseness (6). While structural validity and measurement invariance are not clear (9), reliability and other forms of validity are well established in non-carer populations (21).

The PANAS

We used the PANAS as a preliminary measure of concurrent validity as it was the best available measure in the dataset, having well-established patterns of association with the HADS in a large sample of healthy volunteers (22). It consists of two 10-item mood scales and was developed to provide brief measures of positive and negative affectivity(23). Items are rated from 1-5, score range on each scale is 5-50, with higher scores indicating greater positive/negative affectivity. Both PA and NA scales are internally consistent (α=0.93 and 0.91 respectively), are negatively correlated as expected and correlate with related measures(22).

Statistical Analyses

We used confirmatory factor analysis (CFA) to test the fit of the HADS data with the three

most commonly proposed factor structures: the original two-factor model (6), the one-factor

model(12), and Dunbar et al.'s three-factor non-hierarchical model (24). Diagrams

illustrating these models are shown in Figure 1. For all models, independence of error terms

was specified, and factors were allowed to correlate. The metric of latent variables was set by

fixing the loading of one of the indicators for each variable at 1(25). CFA was performed in

R(version 3.2.2) (26) statistical software using Lavaan (27) and Semtools (28) packages.

Where assumptions of univariate normality (assessed by Shapiro Wilks' test) and

multivariate normality (assessed by Mardia's test) were not met, Satorra Bentler corrected

(robust) indices were used to examine fit of models (29).
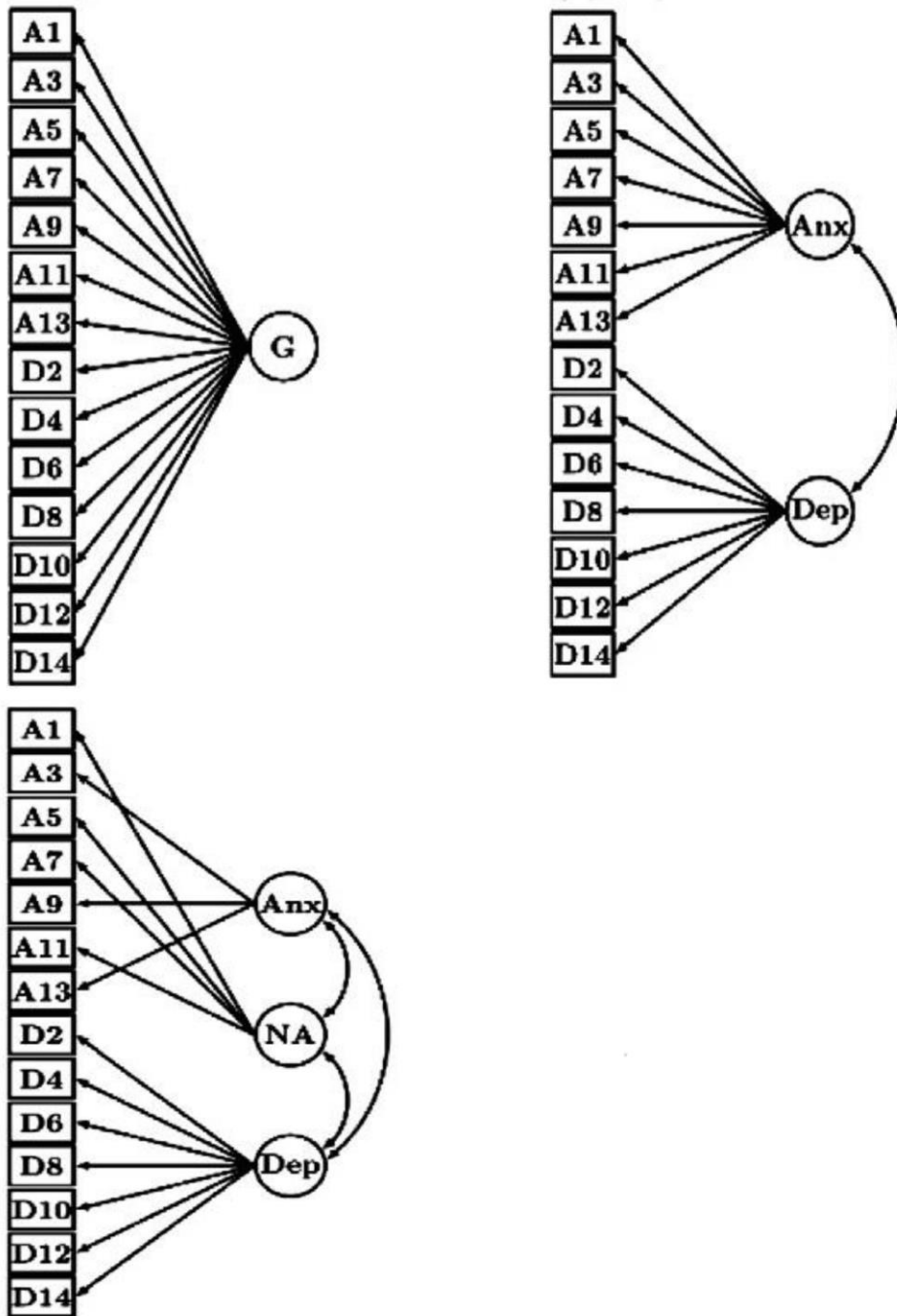
- Figure 1[1] -

In line with recommendations, we assessed model fit using the indices below:

The Standardized Route Mean Square Residual (SRMR) and the Root Mean Square Error of

---

Approximation (RMSEA), have cut-off scores of values <0.05 equating to good fit, <0.08 to adequate fit, and > 0.08 to poor fit. The Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) - have cut-off scores of >0.95 equating to good fit, >0.9, to adequate fit and <0.9 to poor fit. The Bayesian Information Criterion (BIC), is a comparative fit index, with smaller values indicating better fitting models, but no cut-off.

We used information about how well individual items fit within a CFA model to supplement fit indices data (25). Items with standardized residual values in excess of 2.58 and high modification indices were classified as misspecified (30). We used standardized residuals and modification indices to improve ('re-specify') models through specification searching(25) . Once a good fitting model was obtained, we examined parameters for interpretability, size and statistical significance, and the presence of out of range values (25).

We assessed obtained factors for concurrent validity with the PANAS using bivariate correlations. Where data were normally distributed, we used Pearson's r and, where not, Spearman's Rank (31). To evaluate hypotheses that the size of correlations between particular PANAS subscales and particular HADS subscales would significantly differ from one another, we used Steiger tests(32) .

Measurement invariance of the HADS is necessary if it is to be used to test for differences in anxiety and depression across subgroups of carers. 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(17) . To test for measurement invariance of the HADS, we split the data into subgroups in terms of four variables that are related to anxiety and/or depression in the literature(18, 19); gender, age (<65 vs $\geq$ 65), kinship (vertical or horizontal) and cohabitation status. Following this, we examined the final model derived from CFA 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(33) .

To enhance reliability of findings we re-ran (cross-validated) the CFA and measurement invariance analyses in the BECCA dataset.

RESULTS

Descriptive data for the HADS items are shown in Table 2. Carers in the CSP-RYCT and BECCA samples generally reported low levels of depression and anxiety, and consequently data was univariate non-normal with positive skew. Data for original and cross-validation samples were skewed with non-normal kurtosis, thus robust CFA indices were used to examine model fit in original and cross-validation analyses.

Table 1 showing sample characteristics

Characteristics/measure	Categories	N(%) Original sample	N(%) Cross-validation sample
Gender	Female	193(68)	149(64)
Ethnicity	White	264(93)	228(99.1)
	BME˜	20 (7)	2(0.9)
Kinship[+]	Horizontal	183 (63.4)	165 (71.7)
	Vertical	101 (35.6)	65 (28.2)
Education	School leaver:	198(69.7)	-
	Higher/further education	86(30.3)	85(37.3)

Cohabitation status	Cohabiting	224	201(87.5)
	Not cohabiting	59	29 (12.5)
Dementia subtype of relative	Alzheimer's disease	131 (46.1)	-
	Vascular dementia	49 (17.3)	-
	Other	104 (36.6)	-
		Median (IQR)	
Age (years)		68 (19)	69 (19)*
Months of caring		48(48)	36 (3)*
PANAS PA		31.46 (7.37)*	-
PANAS NA (range)		18 (10)	-

Note: PANAS-PA/NA_Positive and Negative Affect Scale Positive/Negative Affect subscales;

For all demographics and clinical measures N was 284 except for PANAS PA and NA (n=265).

˜BME; Black and Minority Ethnic (individuals of non-white descent)

[+]Kinship was defined as horizontal (from the same generation as the person with dementia) or vertical (from the generation below the person with dementia)

- Data was not recorded for the cross-validation sample

* Mean and SD reported as data normally distributed;

Table 2: HADS item descriptive statistics and factor loadings in the best fitting three factor model

HADS Items arranged by factor in final model	Descriptive statistics for HADS items		Factor loadings and parameters in the original analysis (CSP-RYCT data)		
	Mean(SD) original data	Mean (SD) cross-validation data	Estimate (SE)*	Residual Variance (SE)**	Communalities* **
Anxiety 3. I get a sort of frightened feeling as if something awful is about to happen	0.81(1)	1 (1)	1	0.29(0.05)	0.71
9. I get a sort of frightened feeling like butterflies in the stomach	0.57(0.76)	0.74(0.77)	0.78 (0.051)	0.15(0.02)	0.74
13. I get sudden feelings of panic	0.75(0.75)	0.79(0.84)	0.72(0.051)	0.19(0.02)	0.65
Depression 2. I still enjoy the things I used to enjoy	1.03(0.92)	1.37(0.94)	1	0.48(0.05)	0.43
4. I can laugh and see the funny side of things	0.61(0.76)	0.75(0.77)	0.91 (091)	0.29(0.04)	0.51
6. I feel cheerful	0.62(0.75)	0.63(0.71)	0.99 (0085)	0.2(0.02)	0.64
8. I feel as if I have slowed down	1.34(0.85)	1.65(0.94)	0.77 (.082)	0.5(0.05)	0.3
10. I have lost interest in my appearance	0.63(0.85)	0.67(0.82)	0.78 (0.11)	0.51 (0.06)	0.3

12. I look forward with enjoyment to things	0.83(0.9)	1.1(0.85)	0.89 (0.11)	0.41(0.05)	0.41
14. I can enjoy a good book or radio or TV program	0.84(0.84)	0.71(0.88)	1.18 (0.1)	0.31(0.04)	0.62
Negative Affectivity 1. I feel tense or 'wound up'	1.16(0.75)	1.17(1.1)	1	0.29(0.03)	0.49
7. I can sit at ease and feel relaxed	1.15(0.81)	1.39 (0.73)	1.25 (0.12	0.23(0.04)	0.66
11. I feel restless as if I have to be on the move	1.11(0.91)	1.32 (0.91)	1.16 (0.11)	0.46(0.05)	0.45
Excluded 5. Worrying thoughts go through my mind	1.05(0.95)	1.35(0.99)	-	-	

Note: All HADS items for both samples had significant positive skew and were significantly non- normally distributed (Shapiro Wilks, p<0.001)

*Non standardised parameter estimate - Equivalent to a regression coefficient of the individual prediction of the factor by the item, reflecting how much change in the latent variable is associated with one unit change in the item

** Variance in the item unaccounted for by the model;

***The proportion of the variance in the item that is not accounted for by the model.

CFA fit and specification searching

CFA indicated that the three-factor model of anxiety, negative affectivity and depression provided the best fit for the data with 'good fit' on one index (SRMR) and 'adequate fit' on the others. It performed better than the one factor model on all indices and better than the two-factor model on four indices (RMSEA, SRMR, TLI and BIC).

While it was best performing model, fit was not perfect (as three indices indicated 'adequate' not 'good' fit). Consequently, we conducted specification searching for sources of model misspecification in the three-factor model. Modification indices and standardised residuals indicated that item five (part of the negative affectivity factor) had the highest modification index (50.15) indicating significant cross-loadings onto the anxiety factor. Item five also had the highest single standardized residual (9.3) and highest number of standardized residuals (3) above 2.58. In addition to this evidence of substantial misspecification, item five has previously been empirically and conceptually associated with an 'anxiety' factor rather than a negative affectivity factor (25). Consequently we re-ran the model with item five removed. This significantly improved model fit with three fit indices (SRMR, CFI and TLI) now indicating a good fit to the data. Having identified this best fitting model, examination of non-standardized parameter estimates (detailed in table 2) revealed that all items had highly significant loadings onto their respective factors in the expected direction. Furthermore, the correlations between latent factors of anxiety and depression (0.32) anxiety and negative affectivity (0.29) and depression and negative affectivity (0.25) were relatively low (<0.5), providing further indication that a multifactorial solution is appropriate(30).

Cross-validation in the BECCA data confirmed initial results. The three-factor model was again the best fitting model (χ^2 (robust) = 137.86, 74 df; SRMR = 0.055; CFI = 94; TLI = 0.93; RMSEA = 0.065 (90%CI 0.048-0.082)) although fit was not quite as good as in the CSP-RYCT data with all indices of fit adequate. Also, similarly to the original data, the highest modification index (20.18) and most standardized residuals above 2.58 were associated with item five and removal of this item improved values on RMSEA and SRMR (χ^2 (robust) = 112.16, 78 df; SRMR = 0.052; CFI = 94; TLI = 0.93; RMSEA = 0.059 (90%CI 0.042-0.076)) supporting the lack of utility of this item in this group.

Table 3. Fit indices of original and respecified versions of one factor, two factor and three factor models

Model	χ^2 (robust)	Df	Srmr	Cfi	TLI	Rmsea (90% CI)	BIC
One factor	342.68	77	0.071	0.81	0.775	0.126(0.112-0.140)	8465.8
Two factor original	224.2	76	0.066	0.9	0.88	0.093(0.079-0.107)	8438.1
Three factor original	162.61	74	0.049	0.94	0.93	0.072(0.057-0.077)	8370.37
Three factor without item 5	112.64	62	0.041	0.96	0.95	0.06(0.042-0.077)	7686.89

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.*

Internal consistency and Concurrent validity

Internal consistency of the depression (α=0.85) and anxiety scales (α=0.87) was good, with internal consistency of the negative affectivity scale adequate (α=0.77). As expected, correlations between scores on all HADS-scales and those on PANAS-PA/NA were large, significant and in the expected direction. In line with previous literature[22], the HADS depression scale had a significantly larger negative correlation with PANAS-PA than either the HADS anxiety or HADS negative affectivity subscales (t (264)= 5.76 and 5.15 respectively, p<0.001). Also in line with expectations, the HADS negative affectivity scale had a significantly larger correlation with PANAS-NA than the HADS depression subscale (t

(264)=2.62, p<0.01). The one finding counter to expectations was that the PANAS-NA correlation with the HADS negative affectivity scale was not larger than that with the HADS-anxiety scale. It is notable, however, that unlike HADS-negative affectivity, HADS anxiety did not correlate significantly more with PANAS-NA than HADS depression.

Table 4: Correlations of HADS and PANAS subscales

	PANAS-Negative affect	PANAS-Positive Affect
HADS Anxiety	0.64	-0.37
HADS NA (with item 5 missing)	0.69	-0.45
HADS depression	0.57	-0.65

Note; All correlations significant at P<0.001(p values corrected for multiple comparisons using Holm's method(29). Degrees of freedom for all correlations = 263 (N=265).

Measurement invariance

We assessed measurement invariance of the best fitting structure (three factors, item five removed) across groups differing on important variables having first divided the CSP-RYCT and BECCA data divided into subgroups according to gender, age (<65 vs ≥65 ratio = 110:175 and 84:146 respectively), cohabitation status (cohabiting or not cohabiting) and kinship status (horizontal or vertical). With the exception of age, numbers and proportions in each subgroup in original and cross-validation samples are given in Table 1. The results of the analysis of the different types of invariance (configural, metric, strong and strict) in the CSP-RYCT data are shown in Table 3. Configural invariance findings indicate that on one index (CFI values above cut-off) but only just on another (RMSEA values at cut-off) the form of factor structure (three factors, item five removed) determined in the whole group applies across subgroups. Our findings suggest metric invariance across subgroups. This

implies that a unit change in the value of HADS item raw scores is related to the same amount of change in a latent factor score in all subgroups. The strong invariance assumption is met for gender, kinship and cohabitation status but not age. This indicates that for gender, kinship and cohabitation status but not age, the intercepts of items (their value when the latent variable is at 0) are similar across subgroups. The lack of strong invariance across groups in age implies that there may be a systematic response bias with adults in different age groups systematically endorsing different item scores on the HADS independent of their underlying levels of anxiety, depression or negative affectivity.

Table 5. Series of model comparisons to test measurement invariance of three factor model with item 5 removed in original (CSP-RYCT) data.

Subgrouping	Invariance type	$\chi^2 (\Delta\chi^2)$	DF (ΔDF)	Δp	CFI(ΔCFI)	RMSEA
Gender	Configural	241.68	124	N/A	0.93	0082
	Metric	(9.75)	(10)	0.46	(<0.0001)	NA
	Strong	(13.32)	(10)	0.21	(0.002)	NA
	Strict	(20.71)	(13)	0.078	(0.006)	NA
Age	Configural	230.11	124		0.94	0.078
	Metric	19.75	10	(0.032)	(0.006)	NA
	Strong	35.40	10	(0.0001)	(0.014)*	NA
	Strict	23.6	13	(0.035)	(0.001)	NA
Kinship	Configural	232.86	124		0.94	0.079
	Metric	16.44	10	(0.088)	(0.004)	NA
	Strong	27.62	10	(0.002)	(0.01)	NA
	Strict	18.07	13	(0.15)	(0.002)	NA

Cohabitation	Configural	224.21	124		0.94	0.076
status	Metric	13.61	10	(0.19)	(0.002)	NA
	Strong	25.15	10	(0.005)	(0.009)	NA
	Strict	0.82	13	(0.97)	(0.001)	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, ΔCFI < 0.01 implies that the invariance assumption still holds. * indicates that invariance assumption is not met according to these criteria.*

Associations between measurement invariance categories

To further explore why age was measurement variant and other variables were not when age is commonly related to cohabitation status, kinship and gender in dementia carers, we examined association of age with these variables in our sample. Results are given in Table 6. All variables were significantly related to age. The small and medium (rather than large) effect sizes of the associations with gender and cohabitation perhaps explain the difference between gender, cohabitation and age in invariance findings; the large effect size for association with kinship perhaps explains why the pattern of invariance results for kinship was very similar to age, but just failed to cross the significance threshold for strong invariance.

Table 6 showing relationship between measurement invariance categories

	Gender	Kinship	Cohabitation
χ^2 value in original sample	12.187	141.56	81.47
Effect size (Phi) in original sample	0.21	0.71	0.54
χ^2 value in cross validation sample	11.03	84.71	26.2
Effect size (Phi) in cross validation sample	0.22	0.61	0.34

Note: All χ^2 values significant at p<0.001; degrees of freedom for all tests were 1; the direction of the above results for both samples were that women are more likely to be older than men; cohabitees are likely to be older than nom-cohabitees, and horizontal-kinship carers are more likely to be older than vertical-Kinship carers; Effect size categories of 0.1 = small effect, 0.3 = medium effect and 0.5 = large effect were used for interpretation (34).

(34)-

Cross-validation of measurement invariance analysis
Measurement invariance analysis on the BECCA data supported the original analysis findings for gender and kinship, which again exhibited measurement invariance at all levels, (RMSEA always <0.08, CFI > 0.9, and ΔCFI never >0.01) and age, which exhibited measurement invariance at some but not other levels (although it exhibited strict rather than strong measurement variance in this second analysis). However, unlike the original results, configural invariance was not demonstrated for cohabitation status (CFI = 0.87 and RMSEA = 0.93) thus measurement invariance is not reliably demonstrated for this variable.

CONCLUSIONS
We are the first to evaluate the utility of the HADS in carets of people with dementia. We found that, in this group, the HADS is best interpreted in terms of three factors; depression, anxiety and negative affectivity. The depression scale can be interpreted as originally intended (6). This supports the results of recent large clinical trials, which have either used the HADS depression scale as a primary outcome(8) or shown improvement in HADS depression caseness (4). However, the lack of utility of the original anxiety scale suggests that current clinical practice recommendations (7) might be revisited. Future clinical trials

should use a three-factor interpretation of the HADS rather than the currently used one or two

factor models, with depression rather than anxiety or distress used as the primary outcome.

Cross-validation
Our finding of three factors is supported by cross-validation in a separate sample and

preliminary concurrent validity findings, which were as expected (22) as well as the good to

adequate internal consistency of scales, despite limited item numbers which can depress

alpha values (35). Our finding of three factors is also in line with research examining the

HADS in some other populations not defined by their mental health status (9)

Measurement of distress and anxiety

Our results do not support the use of the HADS as a single factor 'distress' measure in carers

of people with dementia. This structure fits the data least well. The use of the original HADS

anxiety scale, is also not supported, rather our findings support the tripartite interpretation of

Dunbar.et al. (30) which splits the anxiety subscale into 'anxiety' and 'negative affectivity'.

However, for pragmatic clinical use, this interpretation of anxiety and negative affectivity is

complex(16) with a lack of a simple way to understand range of symptomatology and

caseness(9, 36) . Furthermore, our finding that the best fit was obtained when one item (item

five) was removed from the negative affectivity scale, while in line with the literature(36) ,

and cross-validated in our study, reduces the size of the negative affectivity and anxiety

subscales to three items each. This is likely to impact on content validity (10) with potential

impact on screening sensitivity. Consequently, we suggest that use of the HADS anxiety

/negative affectivity scales in carers while empirically justified, is clinically limited.

Can the HADS be used to measure across carer subgroups?

There is significant heterogeneity in the carer population and variation in characteristics such

as age, gender, kinship and cohabitation status is related to anxiety or depression(18, 19). Our

measurement invariance data suggest that, in a dementia carer population, differences in

HADS subscale scores between older ($\geq$65) and younger (< 65) adults may be

uninterpretable. Specifically, there may be a systematic bias in the way in which different age groups respond to the HADS that is not to do with differences in anxiety, depression or negative affectivity(17). Consequently, research using the HADS to examine variations in anxiety and depression between those of different age groups will be difficult to understand. In clinical work, where normative reference groups differ in age to the carer, HADS scores may be difficult to meaningfully interpret. As age and kinship are closely associated, interpretation across kinship groups may also be problematic (although this was ambiguous in our data). This difficulty in interpretation may also be the case for cohabitation status as measurement invariance findings from the original analysis were not replicated in the cross-validation sample. As measurement invariance is clearly adequate across kinship and gender, comparisons in HADS scores across this variable can be meaningfully conducted.

Strengths and limitations

This is the first study to examine the utility and interpretability of the HADS in carers of people with dementia specifically. We assessed several international consensus defined areas of health outcome measurement quality(10) including structural validity, reliability, and construct validity through hypothesis testing. Particular strengths are use of a robust CFA approach, and assessment of measurement invariance, an important (37) and often overlooked element of validity. There are limitations. Lack of assessment of concurrent, discriminant and criterion validity are important issues given HADS usage and future research is needed. However, assessment of these validity aspects is in part dependent on understanding factor structure and when factor structure has been established in a population, these validity aspects may be less population variant (21). Consequently, given our support for the original HADS depression scale, future carer research and clinical practice could be preliminarily informed by the well-established depression scale criterion validity data from other populations. The mean score on HADS items was low in both samples. Future work

should examine this measure in samples where there is more variance and higher levels of depression and anxiety. Ethnicity may influence reporting of depression in this group and the limited numbers of Black and Minority Ethnic participants in our sample meant analysis of this was not possible. As our samples came from randomized controlled trials, our participants may differ from clinical populations, potentially limiting generalisability. Although we evaluated the most frequently proposed structures in the literature, not all potential HADS structures were considered. Future research should examine the bi-factor structure (38) and the impact of measurement artefacts (39). Finally, item Response Theory studies should be conducted in this population as these provide strong evidence of latent variable structure, tend to support one factor structures and are more generalizable from the sample to the population(9).

Implications
The HADS is a measure widely used and recommended in dementia carer research and practice. It is best interpreted as measuring three factors; depression, anxiety and negative affectivity. The use of the depression subscale is supported. However, clinical practice recommendations, currently based on a two-factor model should perhaps be revisited, with further research needed. Future trials should focus on HADS depression not anxiety or distress as an outcome measure. The HADS can be used to compare carers varying in gender and possibly kinship, but may not be accurate in comparing those varying in age or cohabitation status. These results run counter to the use of the HADS in carers of people with dementia thus far, and have significant implications for future research and practice with this widely used and recommended measure.

REFERENCES

1.   Mahoney R, Regan C, Katona C, et al: Anxiety and depression in family caregivers of people with Alzheimer disease: the LASER-AD study. The American Journal of Geriatric Psychiatry 2005; 13:795-801
2.   Cooper C, Selwood A, Blanchard M, et al: The determinants of family carers' abusive behaviour to people with dementia: Results of the CARD study. J. Affect. Disord. 2010; 121:136-142
3.   Brodaty H,Gresham M: Effect of a training programme to reduce stress in carers of patients with dementia. BMJ 1989; 299:1375-1379
4.   Livingston G, Barber J, Rapaport P, et al: Clinical effectiveness of a manual based coping strategy programme (START, STrAtegies for RelaTives) in promoting the mental health of carers of family members with dementia: pragmatic randomised controlled trial. 2013;
5.   Knapp M, King D, Romeo R, et al: Cost effectiveness of a manual based coping strategy programme in promoting the mental health of family carers of people with dementia (the START (STrAtegies for RelaTives) study): a pragmatic randomised controlled trial. 2013;
6.   Zigmond AS,Snaith RP: The hospital anxiety and depression scale. Acta Psychiatr. Scand. 1983; 67:361-370
7.   Moniz-Cook E, Vernooij-Dassen M, Woods R, et al: A European consensus on outcome measures for psychosocial intervention research in dementia care. Aging and Mental Health 2008; 12:14-29
8.   Charlesworth G, Shepstone L, Wilson E, et al: Befriending carers of people with dementia: randomised controlled trial. BMJ 2008; 336:1295-1297
9.   Cosco TD, Doyle F, Ward M, et al: Latent structure of the Hospital Anxiety And Depression Scale: a 10-year systematic review. J. Psychosom. Res. 2012; 72:180-184
10.   Mokkink LB, Terwee CB, Patrick DL, et al: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 2010; 63:737-745
11.   Charlesworth G, Burnell K, Crellin N, et al: Peer support and reminiscence therapy for people with dementia and their family carers: a factorial pragmatic randomised trial. J. Neurol. Neurosurg. Psychiatry 2016; jnnp-2016-313736
12.   Razavi D, Delvaux N, Farvacques C, et al: Screening for adjustment disorders and major depressive disorders in cancer in-patients. Br. J. Psychiatry 1990; 156:79-83
13.   Clark LA,Watson D: Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. J. Abnorm. Psychol. 1991; 100:316-336
14.   Chan D, Livingston G, Jones L, et al: Grief reactions in dementia carers: a systematic review. Int. J. Geriatr. Psychiatry 2013; 28:1-17
15.   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; 65:338-344
16.   Stott J, Spector A, Orrell M, et al: Limited validity of the Hospital Anxiety and Depression Scale (HADS) in dementia: evidence from a confirmatory factor analysis. Int. J. Geriatr. Psychiatry 2016;
17.   Drasgow F,Kanfer R: Equivalence of psychological measurement in heterogeneous populations. J. Appl. Psychol. 1985; 70:662-680

18.   Cooper C, Balamurali T,Livingston G: A systematic review of the prevalence and covariates of anxiety in caregivers of people with dementia. Int. Psychogeriatr. 2007; 19:175-195

19.   Cuijpers P: Depressive disorders in caregivers of dementia patients: a systematic review. Aging & mental health 2005; 9:325-330

20.   Graham JW: Missing data analysis: making it work in the real world. Annu. Rev. Psychol. 2009; 60:549-576

21.   Bjelland I, Dahl AA, Haug TT, et al: The validity of the Hospital Anxiety and Depression Scale: an updated literature review. J. Psychosom. Res. 2002; 52:69-77

22.   Crawford JR,Henry JD: The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non‐clinical sample. Br. J. Clin. Psychol. 2004; 43:245-265

23.   Watson D, Clark LA,Tellegen A: Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. 1988; 54:1063

24.   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. Br. J. Clin. Psychol. 2000; 39:79-94

25.   Byrne BM: Structural equation modeling with AMOS: Basic concepts, applications, and programming, Hove, UK, Routledge, 2013

26.   R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2013

27.   Rosseel Y: Lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software 2012; 48:1-36

28.   SemTools Contributors: semTools: Useful tools for structural equation modeling. R package version 0.4-9. , Retrieved from http://cran.r-project.org/package=semTools, 2015

29.   Hu L-t, Bentler PM,Kano Y: Can test statistics in covariance structure analysis be trusted? Psychol. Bull. 1992; 112:351-362

30.   Brown TA: Latent variable measurement models. Oxford library of psychology. The Oxford handbook of quantitative methods 2013; 2:257-280

31.   Field A: Discovering statistics using SPSS, Sage publications, 2009

32.   Steiger JH: Tests for comparing elements of a correlation matrix. Psychol. Bull. 1980; 87:245

33.   Hirschfeld G,von Brachel R: Multiple-group confirmatory factor analysis in RA tutorial in measurement invariance with continuous and ordinal indicators. Prac Assess Res Eval 2014; 19:1-12

34.   Cohen J: Statistical Power Analysis for the Behavioral SciencesNew JerseyLawrence Erlbaum Associates. Inc. Publishers 1988;

35.   Tavakol M,Dennick R: Making sense of Cronbach's alpha. International journal of medical education 2011; 2:53

36.   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 Res. 2003; 118:89-99

37.   Borsboom D: When does measurement invariance matter? Med. Care 2006; 44:S176-S181

38.   Norton S, Cosco T, Doyle F, et al: The Hospital Anxiety and Depression Scale: a meta confirmatory factor analysis. J. Psychosom. Res. 2013; 74:74-81

39.   Martin C,Newell RJ: Factor structure of the Hospital Anxiety and Depression Scale in individuals with facial disfigurement. Psychol. Health Med. 2004; 9:327-336

Figure legend:

'Figure 1 Showing schematics of 1 factor model of Razavi, 2 factor model of Zigmond and Snaith and 3 factor model of Dunbar.'